INSIGHT INTO COGNITIVE STRUCTURE

ASSESSMENT, ANALYSIS, AND INSTRUCTIONAL INNOVATIONS

KUMULATIVE HABILITATIONSSCHRIFT

Wirtschafts- und Verhaltenswissenschaftliche Fakultät
Albert-Ludwigs-Universität Freiburg im Breisgau

vorgelegt von
Dirk Ifenthaler
aus Müllheim / Baden

Wintersemester 2010 / 2011

To
Emma

Knowing is a process not a product
(Jerome S. Bruner)

ACKNOWLEDGEMENTS

This has been a thrilling scientific journey so far! During the last twelve years I had the special privilege to work with outstanding scientific researchers in the field of educational technology and cognitive psychology.

My journey began when I became a student teaching assistant for statistics at the Department of Educational Science at the Albert-Ludwigs-University of Freiburg. Working with *Norbert M. Seel*, *Klaus-Peter Wild*, and *Thomas Eckert* inspired me to dig deeper into the methodological understanding of education. Especially the application of statistical procedures for complex research designs kept me reading about and experimenting with various statistical software packages. Within this first stage of my journey I also developed my interest for the theoretical understanding of cognitive structures.

Using simulations for educational purposes marks the second stage of my scientific journey. Working with *Sara-Dunja Menzel* and *Volker Schweinbenz* on developing a simulation game for a better understanding of the complex processes of a school organization laid the foundation for a larger research project I recently initiated with my dear colleague and friend *Volker Schweinbenz*. Within this second stage I also got to know the scientific world outside of Freiburg through the ~monist project. Traveling to project meetings in Bielefeld and Frankfurt and discussing ideas of the project with *Dietrich Dörner*, *Sören Lorenz*, and *Wolfram Horstmann* set light into the various possibilities of scientific life.

The third stage of my scientific journey started when I got involved in a new project on model-based learning and teaching. Together with my innovative colleagues *Bettina Couné*, *Katharina Schenk*, and *Ulrike Hanke*, new approaches for the assessment and analysis of cognitive structures have been laid out.

My dissertation project marks the forth part of my scientific journey. Putting together my experience and ideas into a completely new project resulted in the development of a new technology for an automated assessment and analysis of cognitive structures – the SMD Technology. Defending my dissertation at the same day as my dear colleague and friend *Pablo Pirnay-Dummer* did, marked a very special day in this forth stage of my scientific journey.

Continuing working on my dissertation project and joining the ideas of *Pablo Pirnay-Dummer* with my ideas marks the highlight of the fifth stage of my scientific journey. Travelling the world and presenting our work together has always been a highly inspiring and joyful time. The number of my international collaborators has grown ever since. It is always great to discuss new ideas with wonderful people and great researchers such as *David H. Jonassen*, *Roy B. Clariana*, *Valerie J. Shute*, *Harold F. O'Neil*, *Tiffany A. Koszalka*, *James W. Pellegrino*, *Andrew S. Gibbons*, and many more. Furthermore, the continuous support of *J. Michael Spector* helped me to push towards new projects and implementing new ideas into powerful tools – HIMATT (Highly Integrated Model Assessment Technology and Tools). Closely related to my projects on assessment and analysis of cognitive structures is a great colleague and a wonderful friend, *Tristan E. Johnson*. All our projects turned out to be respected in the scientific community. Additionally, organizing various conferences at the Albert-Ludwigs-University of Freiburg introduced me to a new group of great researchers, namely *Pedro Isaías*, *Kinshuk*, and *Demetrios Sampson*. Together with *J. Michael Spector* I am honored to be part of the CELDA (Cognition and Exploratory Learning in the Digital Age) conference committee organizing an annual international conference. Furthermore, a strong international research group focusing on problem solving, serious games, and their assessment has grown constantly, including my great colleagues *Deniz Eseryel* and *Xun Ge.* As a result of this highly productive stage of my scientific journey, most of the papers of this cumulative work originate from this period. Additionally, several edited volumes and a monograph in collaboration with *Norbert M. Seel* are some of the products of this stage.

Moving from the Albert-Ludwigs-University of Freiburg to the University of Mannheim marks another important stage of my scientific journey. At this current stage I am happy to seek advice from many valued colleagues, especially from *Norbert M. Seel*, *Matthias Nückles*, *Oliver Dickhäuser*, *Olga Zlatkin-Troitschanskai*, *Klaus Breuer*, and *Peter Drewek*.

I want to thank all the above mentioned colleagues and friends and those I may have forgotten for their inspiration, motivation, and continuous support. I shall not attempt to thank my wife *Kathrin*, my son *Remo Max* and my family. Everything I am and will be is a complex combination of their unconditional love, patience and unique ways. I dedicate this effort to them and hope to be worthy of the lives they live. *I am looking forward to the next stages of this thrilling scientific journey!*

Dirk Ifenthaler

Freiburg, December 2010

Table of Contents

1

PROLOGUE

Strong theoretical foundations and precise methodology are always the one and only starting point for good research. Without sound foundations nothing follows, and thus a deep understanding of the theoretical assumptions of cognitive structure and methodology involved is mandatory for research on cognition and learning as well as for instructional design. Several research projects contribute to the overall scientific knowledge with regard to cognitive structure and its assessment, analysis, and instruction. Cognitive structure continued to be a key subject in different fields of research for more than a century. For good reason. Foundations from cognitive science, computer science, philosophy, and cognitive psychology describe the workings of the human mind in tasks of deductive and inductive reasoning, especially for reasoning in uncertainty. They lead to theories of problem solving and to theories of learning and instruction which are both highly interdependent. The development of useful systems has always been a goal for scientists and engineers serving professional communities in the fields of instructional design and instructional systems development. This cumulative work outlines a research project which enables an insight into cognitive structure highlighting ways of assessment, analysis, and instructional innovations.

Advances of technology

As instructional psychology is becoming more specialized and complex and technology is offering more and more possibilities for gathering data, instructional researchers are faced with the challenge of processing vast amounts of data. Yet the more complex our understanding of the field of learning and instruction becomes and the more our theories advance, the more pronounced is the need to apply the structures of the theories to sufficiently advanced methodology in order to keep pace with theory development and theory testing. In addition to obtaining a good fit between theory and diagnostics, this task entails making the methodology and tools feasible (easy to use and easy to interpret). Otherwise, the methodologies will only be used by their developers. The development of useful systems has always been a goal for scientists and engineers serving professional communities in the fields of instructional design and instructional systems development.

The progress of computer technology has enabled researchers to adopt methods from artificial intelligence, graph theory, feature analysis, feature tracking, and applied statistics and to use computers to implement computer-based instructional systems. Researchers have now also succeeded in developing more effective tools for the assessment of knowledge in order to enhance the learning performance of students.

The structure of this cumulative work

Several research projects contribute to the overall scientific knowledge with regard to cognitive structure. The following peer-reviewed publications build up this cumulative work highlighting ways of assessment, analysis, and instructional innovations. Table 1.1 illustrates the individual chapters and the corresponding publications.

Chapter 2 (based on Ifenthaler, 2010d) addresses information retrieval from human memory and how it will reflect in part the individual's cognitive structure within and between concepts or domains. Accordingly, this chapter critically reflects possibilities and limitations of a systematic assessment and analysis of cognitive structure and introduces important concepts (e.g., externalization, representation, re-representation).

In chapter 3 (based on Ifenthaler, 2010c) it is argued that a wide variety of empirical approaches for the analysis of external representations of cognitive structure exist, but they often lack a solid theoretical foundation and their analysis is considered to be very time consuming. On the other hand, new technologies such as concept mapping tools are being introduced into learning environments, but the analysis of data collected with such new technologies still places a huge demand on methodologies. The purpose of chapter 3 is to introduce the computer-based and automated SMD Technology for relational, structural, and semantic analysis of externalized representations.

Chapter 4 (based on Al-Diban & Ifenthaler, in press) determines the strength and limitations of new methodological approaches. Overall, it is worthwhile to compare analysis approaches for measuring externalized mental models systematically in order to test their advantages and disadvantages, strengths and limitations. A series of pair-wise comparative studies show strengths, unique characteristics, and collective viability of different assessment and analysis methods. However, the above mentioned study only focused on conceptual differences of the analysis approaches and did not use empirical data. Accordingly, chapter 4 reports an empirical case study and compares two analysis approaches - QFCA (Qualitative & Formal Concept Analysis) and SMD (Surface, Matching, Deep Structure) - using identical data. The aim of this comparative study is to determine conceptual and empirical strengths and limitations of two different approaches for analyzing externalized cognitive structure.

Chapter 5 (based on Pirnay-Dummer, Ifenthaler, & Spector, 2010) introduces an integrated set of assessment tools called HIMATT (Highly Integrated Model Assessment Technology and Tools) which addresses this deficiency. HIMATT is Web-based and has been shown to scale up for practical use in educational and workplace settings, unlike many of the research tools developed solely to study basic issues in human learning and performance. In this chapter, the functions of HIMATT are described and several applications for its use are demonstrated. Additionally, two studies on the quality and usability of HIMATT are presented.

The "mystery of cognitive structure" is questioned in chapter 6 (based on Ifenthaler, Masduki, & Seel, in press). Many research studies have clearly demonstrated the importance of cognitive structures as the building blocks of meaningful learning and retention of instructional materials. Identifying the learners' cognitive structures will help instructors to organize materials, identify knowledge gaps, and relate new materials to existing slots or anchors within the learners' cognitive structures. The purpose of this empirical investigation is to track the development of cognitive structures over time. Accordingly, it is demonstrated how various indicators derived from graph theory can be used for a precise description and analysis of cognitive structures. Results revealed several patterns that help to better understand the construction and development of cognitive structures over time.

Chapter 7 (based on Ifenthaler, accepted) investigates cross-domain distinguishing features of cognitive structures. In this experimental study, participants worked on the subject domains biology, history, and mathematics. Results clearly indicate different structural and semantic features of cognitive structures across the three subject domains. Additionally, we found that written texts and causal maps seem to represent different structure and content across the three subject domains when compared to an expert's representation.

Chapter 8 (based on Ifenthaler & Seel, in press) reports findings from an experimental study in which 73 participants in three experimental groups solved logical word problems at ten measurement points. Changes of cognitive structures are illuminated and significant differences between the treatments are reported. The results also indicate that supportive information is an important aid for developing cognitive structures while solving logical problems.

Chapter 9 (based on Pirnay-Dummer & Ifenthaler, in press) presents an experimental study which integrates automated natural language-oriented assessment and analysis methodologies into feasible reading comprehension tasks. With the newly developed toolset, prose text can be automatically converted into an association net which has similarities to a concept map. The study investigates the effects of association nets made available to learners prior to reading. The results reveal that the automatically created graphs are highly similar to classical expert graphs.

Chapter 10 (based on Ifenthaler, 2009) reports a final experimental study on automated individualized feedback. Here, feedback is considered an elementary component for supporting and regulating learning processes. Different types of model-based feedback are investigated. Seventy-four participants were assigned to three experimental groups in order to examine the effects of different forms of model-based feedback. With the help of seven automatically calculated measures, changes in the participants' understanding of the subject domain "climate change", represented by causal diagrams, are reported.

Finally, the epilogue highlights ongoing and future research projects for gaining a better insight into cognitive structure. These projects focus on new methodological developments as well on instructional applications.

TABLE 1.1
Peer-reviewed publications of the cumulative work

Chapter No.	*Publication*	*Impact factor from Journal Citation Reports®, Thomson Reuters (if available)*
Chapter 2	Ifenthaler, D. (2010). Scope of graphical indices in educational diagnostics. In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 213-234). New York: Springer.	*N/A*
Chapter 3	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	1.183
Chapter 4	Al-Diban, S., & Ifenthaler, D. (in press). Comparison of two analysis approaches for measuring externalized mental models: Implications for diagnostics and applications. *Journal of Educational Technology & Society.*	1.067
Chapter 5	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. doi: 10.1007/s11423-009-9119-8	1.183
Chapter 6	Ifenthaler, D., Masduki, I., & Seel, N. M. (in press). 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	1.341
Chapter 7	Ifenthaler, D. (accepted). Identifying cross-domain distinguishing features of cognitive structures. *Educational Technology Research and Development.*	1.183
Chapter 8	Ifenthaler, D., & Seel, N. M. (in press). A longitudinal perspective on inductive reasoning tasks. Illuminating the probability of change. *Learning and Instruction*. doi: 10.1016/j.learninstruc.2010.08.004	2.372
Chapter 9	Pirnay-Dummer, P., & Ifenthaler, D. (in press). Reading guided by automated graphical representations: How model-based text visualizations facilitate learning in reading comprehension tasks. *Instructional Science*. doi: 10.1007/s11251-010-9153-2	1.341
Chapter 10	Ifenthaler, D. (2009). Model-based feedback for improving expertise and expert performance. *Technology, Instruction, Cognition and Learning, 7*(2), 83-101.	*N/A*

2

SYSTEMATIC ASSESSMENT AND ANALYSIS OF COGNITIVE STRUCTURE 🕮

It is argued that the order in which information is retrieved from memory will reflect in part the individual's cognitive structure within and between concepts or domains. When compared to that of a novice, a domain expert's cognitive structure is considered to be more tightly integrated and to have a greater number of linkages between interrelated concepts. There is thus immense interest on the part of researchers and educators to diagnose a novice's cognitive structure and compare it with that of an expert in order to identify the most appropriate ways to bridge the gap. However, an assessment and analysis of cognitive structures is always biased as we do not know the direct functions of internalization and externalization. Additionally, the possibilities of externalization are limited to a few sets of sign and symbol systems – characterized as *graphical* and *language-based approaches*. This chapter critically reflects possibilities and limitations of a systematic assessment and analysis of cognitive structure and links them to theoretical and methodological foundations.

🕮 This chapter is based on: Ifenthaler, D. (2010). Scope of graphical indices in educational diagnostics. In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 213-234). New York: Springer.

Introduction

Knowledge representation is a key concept in psychological and educational diagnostics. Thus, numerous models for describing the fundamentals of knowledge representation have been applied so far. The distinction which has received the most attention is that between declarative ("knowing that") and procedural ("knowing how") forms of knowledge (see Anderson, 1983; Ryle, 1949). Declarative knowledge is defined as factual knowledge, whereas procedural knowledge is defined as the knowledge of specific functions and procedures for performing a complex process, task, or activity. Closely associated with these concepts is the term cognitive structure, also known as knowledge structure or structural knowledge (Jonassen, Beissner, & Yacci, 1993), which is conceived of as the manner in which an individual organizes the relationships between concepts in memory (Ifenthaler, et al., in press; Shavelson, 1972). Hence, an individual's cognitive structure is made up of the interrelationships between concepts or facts and procedural elements.

Further, it is argued that the order in which information is retrieved from memory will reflect in part the individual's cognitive structure within and between concepts or domains. When compared to that of a novice, a domain expert's cognitive structure is considered to be more tightly integrated and to have a greater number of linkages between interrelated concepts. There is thus immense interest on the part of researchers and educators to diagnose a novice's cognitive structure and compare it with that of an expert in order to identify the most appropriate ways to bridge the gap (Ifenthaler, et al., in press; Ifenthaler & Seel, 2005). By diagnosing these structures precisely, even partially, the educator comes closer to influencing them through instructional settings and materials.

Functions of representation and re-representation

However, it is not possible to measure these internal representations of knowledge directly. Additionally, it is argued that different types of knowledge require different types of representations (Minsky, 1981). Therefore, we argue that it is necessary to identify economic, fast, reliable, and valid techniques to elicit and analyze cognitive structures (Ifenthaler, 2008). In order to identify such techniques, one must be aware of the complex processes and interrelationships between internal and external representations of knowledge. Seel (1991, p. 17) describes the function of internal representation of knowledge by distinguishing three zones – the object zone **W** as part of the world, the knowledge zone **K**, and the zone of internal knowledge representation **R**. As shown in Figure 2.1, there are two classes of functions: (1) f_{in} as the function for the internal representation of the objects of the world (internalization), and (2) f_{out} as the function for the external re-representation back to the world (externalization).

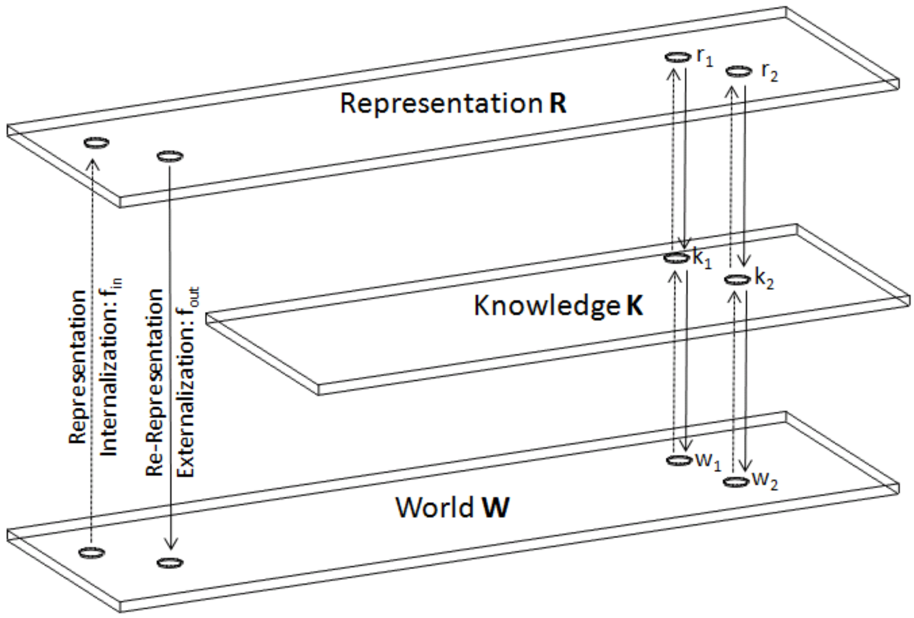

FIGURE 2.1. *Functions of representation and re-representation*

Neither class of functions is directly observable. Hence, a measurement of cognitive structures is always biased as we are not able to more precisely define the above described functions of internalization and externalization (Ifenthaler, 2008). Additionally, the possibilities of externalization are limited to a few sets of sign and symbol systems (Seel, 1999b) – characterized as *graphical* and *language-based approaches*.

Lee and Nelson (2004) report various graphical forms of external representations for instructional uses and provide a conceptual framework for external representations of knowledge. Graphical forms of externalization include (1) knowledge maps, (2) diagrams, (3) pictures, (4) graphs, (5) charts, (6) matrices, (7) flowcharts, (8) organizers, and (9) trees. However, not all of these forms of externalization have been utilized for instruction and educational diagnosis (Ifenthaler, 2008; Scaife & Rogers, 1996; Seel, 1999a). Other forms of graphical approaches are the structure formation technique (Scheele & Groeben, 1984), pathfinder networks (Schvaneveldt, 1990), mind tools (Jonassen, 2009; Jonassen & Cho, 2008), and causal diagrams (Al-Diban & Ifenthaler, in press). Language-based approaches include thinking-aloud protocols (Ericsson & Simon, 1993), teach-back procedures (Mandl, Gruber, & Renkl, 1995), cognitive task analysis (Kirwan & Ainsworth, 1992), and computer linguistic techniques (Pirnay-Dummer, et al., 2010; Seel, Ifenthaler, & Pirnay-Dummer, 2009).

As discussed above, there are numerous approaches for eliciting knowledge for various diagnostic purposes. However, most approaches have not been tested for reliability and validity (Ifenthaler, 2008; Seel, 1999a). Additionally, they are almost only applicable to single or small sets of data (Al-Diban & Ifenthaler, in press; Ifenthaler, 2010c). Hence, new approaches are required which have not only been tested for reliability and validity but also provide a fast and economic way of analyzing larger sets of data. Additionally, approaches for educational diagnostics also need to move beyond the perspective of correct and incorrect solutions. As we move into the 21st century, we argue that the application of alternative assessment and analysis strategies is inevitable for current educational diagnostics.

Alternative assessment and analysis strategies

Externalizations are the only available artefacts for empirical investigations. An externalization is always made by means of interpretation. But the externalization also needs interpretation for its analysis. These are two different kinds of interpretation. All kinds of features may be clustered for a description and aggregation of the artefact. Some of the interpretation is done by the learner and some of it is carried out by humans and technology. In most cases a mixture of all three interpreters will be part of the assessment. This mixture and the complexity of the construct both make it specifically difficult to trace the steps and bits of knowledge.

Not all types of externalizations have the same types of properties and strengths, e.g., written language is always sequenced and has multiple dimensions at the same time (it is still impossible to trace them all), concept maps are not semantic webs most of the time due to underspecification problems and a lack of homogeneity, association networks do not have directions and propositions, causality networks can not deal with dynamics, and representations of dynamic systems are almost impossible to aggregate – nor are they supposed to be aggregable in the first place. The list is not even complete (see Ifenthaler & Pirnay-Dummer, 2010a).

There is no easy and no complete way to integrate any of them, and the strength of good research therefore lies, maybe more than in other research domains, in a fitting integration: Multiple perspectives on the same construct are usually needed. Only if the research questions are very specific may a single approach suffice. But this is rarely the case. Researchers and practitioners will have to carefully justify their selection alongside their research questions and goals, especially if important long-term decisions are based upon the assessments. The same care should be taken for decisions in the field. The only way to make better decisions about the kind of externalization as well as the type of instrument to be used on it is to know the strengths and weaknesses of the instruments (Ifenthaler, 2008; Ifenthaler & Pirnay-Dummer, 2010a). It is worth the effort to acquaint oneself with at least a representative selection of the available tools.

Once the external re-representations have been assessed and aggregated, two competing demands are at hand: First, we need to keep as much information from the external re-representations as possible. Secondly, especially in large datasets the information needs to be condensed in such a way that we are still able to selectively decide on or test our theories and practical goals. Combining both demands is not always easy and the measures need to be chosen carefully with an eye to the research question, evaluation, analysis, or designed plans in order to provide the proper answers.

In the field of computer-based diagnostics knowledge artefacts (objects of investigation) are very often graphs. If they are not graphs from the start, they are usually transferred into graphs after assessment. The purpose is aggregation (Ifenthaler, 2010d; Ifenthaler & Pirnay-Dummer, 2010a). Purely qualitative methods are the exception. However, their opposition to any kind of aggregation lies in their nature, and they can be aided by computer programs but not carried out automatically. Any aggregation of qualitative research results is at least to be considered a mixed method: Aggregation is quantitative by nature. This does not, on the other hand, mean that all aggregation serves the same purpose or that it can not differ in quality and the amount of information it preserves. As always, the choice of the right measures and comparisons is determined by the research question or practical goal. The main reason for comparison is the further processablity of the artefacts, which is especially interesting for computer based analysis because it can be automated. The measures allow questions about whether one group of experts structures things differently than another or whether a group of learners makes progress over time, e.g., as compared to experts.

With computer-based analysis, large data sets are attainable even if resources are limited. When the objects under investigation are graphs, graph theory provides the only logical choice for analysis and a stable basis for several further developments (Harary, 1974; Tittmann, 2003, 2010). Surprisingly, the application of graph theory can only rarely be found in research on learning and instruction (Ifenthaler, 2010d). Usually very simple measures are used as single indicators which do not carry much of the initially rich information and are usually not validated at all (Ifenthaler, 2008). And even in the case that graph theory is applied, the measures used sometimes lack a connection to the theories of learning and instruction, and the scope of the measures is sometimes misinterpreted.

Good theories and sound research have a great chance of leading to practical improvements. The process may take time, but eventually when things are explained properly, the process succeeds; slower but usually more stable than by the use of intuitive approaches. But sometimes the odds are even more optimistic. These are the cases where the investigation itself is *part* of the improvement. The need for assessment strategies which support the process under assessment at the same time is not new (Ifenthaler & Pirnay-Dummer, 2010b).

However, with new technologies at hand, at least parts of this demand can be better fulfilled. This cumulative work will start with knowledge constructs, representations, and assessment methods and moves on to decisions on specific measures and reasoning. Then, the impact the assessment, the interpretation, the aggregation, and methodological decisions have on knowing and the learning process itself is presented. As diverse as they may be, the methods and technologies which will be described have one common advantage: They use the cognitive facilities and assess them at the same time. Moreover, they all use them in the way in which they are used in everyday situations. Even when used for assessment only, these methods do not create an artificial assessment situation which leads too far away from the usual reflection. Thus, this leads back to the beginning, where it is stated that the investigation of knowledge is recursive – and that the recursion may very well be infinite in theory (Ifenthaler & Pirnay-Dummer, 2010b).

3

TOWARDS A NEW METHODOLOGY 🕮

A wide variety of empirical approaches for the analysis of external representations of cognitive structure exist, but they often lack a solid theoretical foundation and their analysis is considered to be very time consuming. On the other hand, new technologies such as concept mapping tools are being introduced into learning environments, but the analysis of data collected with such new technologies still places a huge demand on methodologies. The purpose of this chapter is to introduce the computer-based and automated *SMD Technology* for relational, structural, and semantic analysis of externalized representations. First, the theoretical foundation fort he proposed methodology is introduced. Second, the complex processes of externalizing internal knowledge representations (re-representation) will be discussed. Third, the *SMD Technology*, which enables a measurement of graphical representations and concept maps with three different quantitative indices, is presented. Then, the empirical reliability and validity testing of the *SMD Technology* is highlighted. Finally, a broad field of applications for the *SMD Technology* within the field of research, learning, and instruction is discussed.

🕮 This chapter is based on: 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

Introduction

The demand for good instructional environments presupposes valid and reliable tools, instruments, and methodologies for educational research. However, many of them are developed with little or no theoretical justification, which leads to doubtful findings and no contribution to the improvement of learning environments (Novak, 1998). Accordingly, the development of new tools, instruments and methodologies to capture key latent variables associated with human learning and cognition requires a solid theoretical foundation.

One central interest of psychological and educational research is internal cognitive processes and systems, which are described by theoretical constructs such as mental models and schemata (Seel, 1991). However, mental models and schemata are theoretical scientific constructs which are not directly observable. Accordingly, researchers can only learn about mental models or schemata if (1) individuals communicate their internal systems (Seel, 1991) and if (2) valid and reliable instruments and methodologies are used to analyze them (Seel, 1999a). A wide variety of empirical approaches for the analysis of external representations of mental models and schemata exist (Al-Diban, 2002), but they often lack a solid theoretical foundation and their analysis is considered to be very time consuming (Ifenthaler, 2008). On the other hand, new technologies such as concept mapping tools are being introduced into learning environments, but the analysis of data collected with such new technologies still places a huge demand on methodologies.

The purpose of this chapter is to introduce the computer-based and automated *SMD Technology* for relational, structural, and semantic analysis of graphical representations and concept maps. First, the theoretical constructs of mental models and schemata as a key concept for understanding human learning and problem solving processes are introduced. Second, the complex processes of externalizing internal knowledge representations (re-representation) will be discussed. Third, the *SMD Technology*, which enables a measurement of graphical representations and concept maps with three different quantitative indices, is presented. Then, the empirical reliability and validity testing of the *SMD Technology* is highlighted. Finally, a broad field of applications for the *SMD Technology* within the field of research, learning, and instruction is discussed. The chapter ends with a conclusion and future perspectives.

Background

Mental models and schemata are theoretical constructs for understanding human learning and problem solving processes. Following the verdict of Piaget (1950, 1976), it is argued that new information is processed by the complimentary processes of assimilation and accommodation. According to Seel (1991), a person can assimilate new information as long as an adequate schema can be activated. If the activated schema does not match exactly, it can be adjusted by means of *accretion, tuning,* or *reorganization*. The *accretion process* is defined as an accumulation of new information to the existing schema. *Tuning* can be described as a change of single components within the activated schema. The result of a successful adjustment of a schema is a subjective plausible solution of a problem or the understanding of new information. However, if the processes of *accretion* and *tuning* are not successful or if no schema is available at all, new information can only be accommodated by the process of *reorganization*. According to Seel (1991), the process of *reorganization* is realized by constructing a mental model (see Figure 3.1).

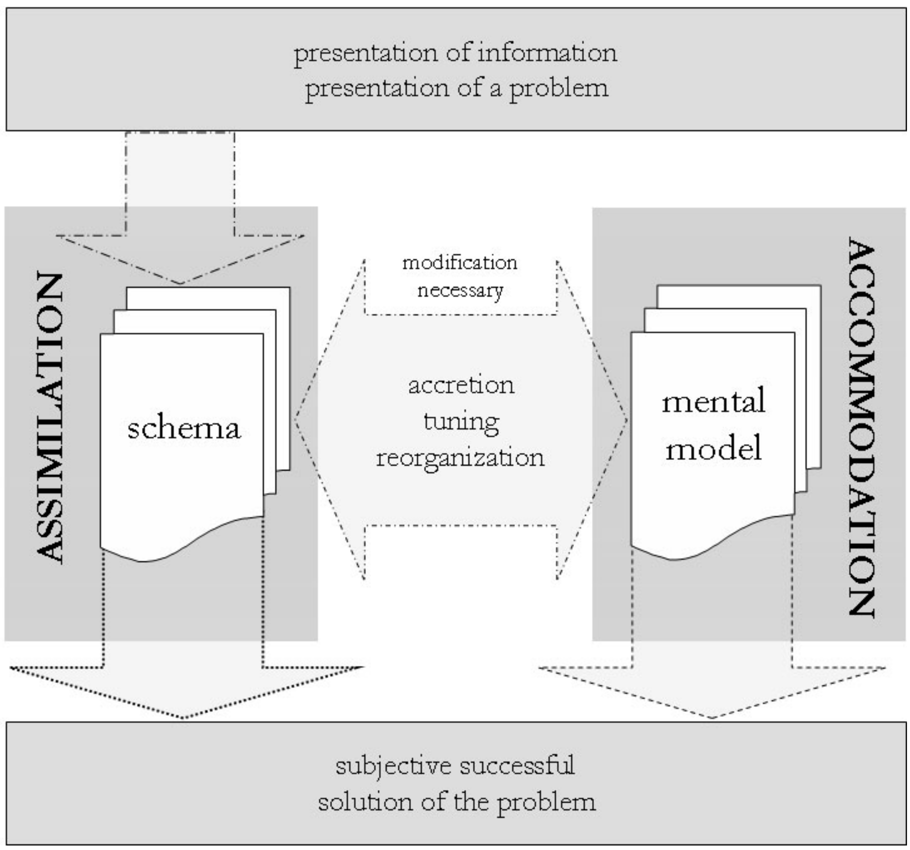

FIGURE 3.1. *The process of assimilation and accommodation*

Mental models are dynamic ad hoc constructions of individuals that provide subjective plausible explanations on the basis of restricted domain-specific information. Johnson-Laird (1983) describes the model building process as a step-by-step reconstruction of an initial mental model (*fleshing out*). Additionally, the *reduction to absurdity* (Seel, 1991) is used to test whether the activated mental model can be replaced by another mental model. However, as long as an activated mental model provides enough subjective plausibility to meet the requirements of a phenomenon to be explained, there is no need for the construction of a new mental model. Seel (1991) assigns mental models four general functions, (1) simplification, (2) envisioning, (3) analogical reasoning, and (4) mental simulation. Depending on the objective of the model-building person, one of the four functions is used for the mental model building process. In comparison to the activation of an available schema, the mental effort for the construction of a mental model is higher and more time consuming (Seel, 2008).

Accordingly, learning, reasoning, and problem solving involve the construction of mental models and schemata. In order to support successful learning, reasoning, and problem solving, it is necessary to investigate the mental model building process precisely. However, as it is not possible to measure internal representations of knowledge directly (e.g., schemata, mental models), the following paragraph will focus on the complex processes of externalizing internal knowledge representations.

Externalization of internal knowledge structures

Theoretical constructs such as the mental models and schemata discussed above are used by cognitive and educational researchers to explain the complex phenomenon of human learning, reasoning, and problem solving. As long as these internal knowledge structures are not directly observable, researchers require adequate tools, instruments, and methodologies to allow people to externalize them. According to Scandura (2007), there exist various possibilities how to construct such knowledge representations. We consider the process of externalization as a conscious process of communicating mental models or schemata using adequate sign and symbol systems (see Le Ny, 1993). Hence, externalization can be realized through speaking out aloud, writing a text, drawing a picture, or constructing a diagram, graphic, or concept map (Ifenthaler, 2008).

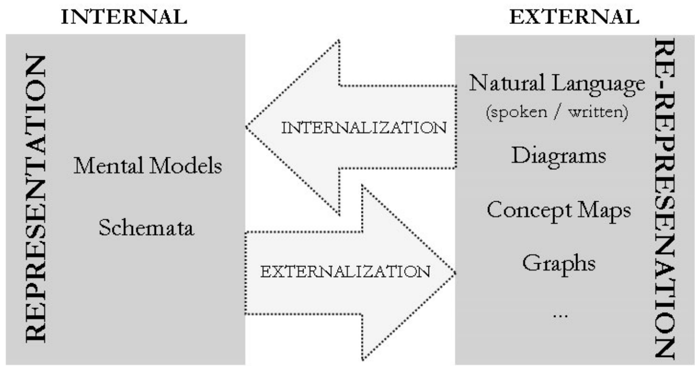

FIGURE 3.2. *Interrelation of internal and external representations*

As shown in Figure 3.2, we are able to distinguish between internal representations (e.g., mental models, schemata) and external re-representations (communicated using adequate sign and symbol systems). Furthermore, we argue that these two types of model representations are interrelated. First, through the process of *internalization*, a person is able to construct a mental model or activate an available schema. From the point of view of instructional design, the process of *internalization* is where we can systematically influence the construction of mental models by providing well-designed external re-representations (e.g., learning materials, feedback, etc.) of phenomena to be explained (e.g., Norman, 1983).

Second, the process of *externalization* enables a person to communicate his or her understanding of phenomena in the world. This perspective is the only way in which researches can learn more about a person's internal representations. Accordingly, adequate tools, instruments, and methodologies for the analysis of mental models or schemata can only be developed with a clear understanding of the complex processes of *internalization* and *externalization*. Although it appears to be possible to assess internal representations through their externalized re-representations, we need to keep in mind that the re-representations might be biased through the lack of communication skills, the use of inadequate sign and symbol systems or the use of insufficient research instruments.

Therefore we argue that instruments used for the analysis of such constructs must have a strong theoretical foundation and be tested for reliability and validity (Ifenthaler & Seel, 2005; Seel, 1999a). A detailed review of methodologies for the assessment of graphical representations revealed a huge demand for an automated and computer-based tool (Ifenthaler, 2006). As a result, the *SMD Technology* was developed.

SMD technology

Based on the theory of mental models (Seel, 1991) and graph theory (Bonato, 1990; Chartrand, 1977; Harary, 1974; Tittmann, 2003), the computer-based and automated *SMD Technology* (Surface, Matching, Deep Structure) uses (a) graphical representations such as concept maps or (b) natural language expressions to analyze individual processes in persons solving complex problems at single time points or multiple intervals over time. In the following, we define the externalized knowledge structures as a model M.

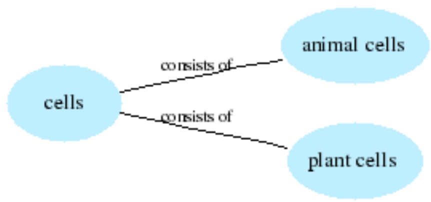

FIGURE 3.3. *Model M_3 composed of two propositions P_i*

Depending on the elicitation process (e.g., using the *Structure Formation Technique* [paper and pencil]; *concept mapping tools* [computer-based]; *natural language statements* [computer-based or paper and pencil]), the raw data should be stored pairwise (as propositions P_i) including (a) the *model number* as an indicator of which model a proposition belongs to, (b) *node1* as the first node of the proposition, (c) *node2*, which is connected to the first node, and (d) a *link* which describes the link between the two nodes (see Figure 3.3 and Table 3.1).

TABLE 3.1
Raw data of a model stored pairwise (as propositions)

Model number	*Node1*	*Node2*	*Link*
003	cells	animal cells	consists of
003	cells	plant cells	consists of
…			

After the raw data has been transformed into the standardized format (see Table 3.1), it is stored on a SQL (structured query language) database. However, the transformation process of paper and pencil models (e.g., *Structure Formation Technique*) is very time consuming. Therefore, we recommend the use of computer-based elicitation techniques which already support the standardized format (e.g., C-Map, DEEP, MITOCAR) in order to guarantee a more economical analysis and additionally a highly reliable transformation process (Ifenthaler, 2006).

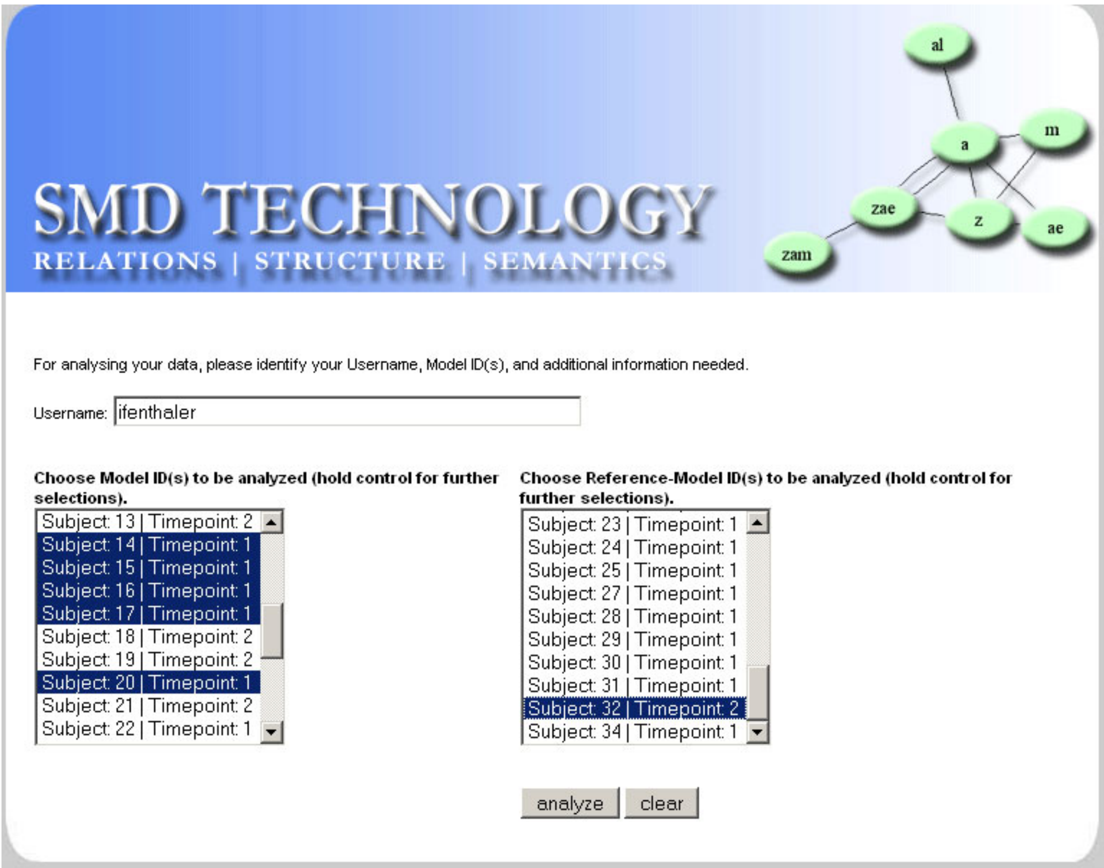

FIGURE 3.4. *User interface of the SMD technology*

The automated analysis process of the *SMD Technology* will be started by the researcher through the *User Interface*, where all stored models in the SQL database can be selected (see Figure 3.4). After selecting the models M_i for the analysis process, the system will automatically calculate three numerical indicators out of all nodes and links - *Surface*, *Matching*, and *Deep Structure* - and generate standardized graphical re-representations for each individual model M_i (Ifenthaler, 2006).

Surface structure

The relational structure of each individual model M_i is represented on the *Surface Structure*. This simple and easily calculable indicator is computed as the sum of all propositions P_i in a model M_i.

$$\theta = \sum_{i=0}^{n} P_i \qquad [1.1]$$

θ is defined as a value between 0 (no proposition = no model) and *n* (n propositions P_i of a model M_i). The *Surface Structure* of model M_3, represented in Figure 3.3, would result in θ = 2. According to the theory of mental models (Seel, 1991), the number of nodes and links or propositions a person uses is a key indicator for the investigation of the progression of knowledge over time in the course of problem solving processes (Scandura, 1988). However, although this first indicator enables a rapid and economical analysis of the relational structure of a model M_i, additional indicators are required for a more detailed analysis.

Matching structure

The structural property of a model M_i is displayed on the *Matching Structure*. The second level of the *SMD Technology* indicates the range and complexity of a model M_i.

$$\mu = \max_{i,j}\{d(i,j)\} \qquad [1.2]$$

μ is computed as the diameter of the spanning tree of a model M_i and can lie between 0 (no links) and n. In accordance with graph theory, every model M_i contains a spanning tree. Spanning trees include all nodes of a model M_i and are acyclic (Tittmann, 2003). Figure 3.5 illustrates model M_5 and its corresponding spanning tree.

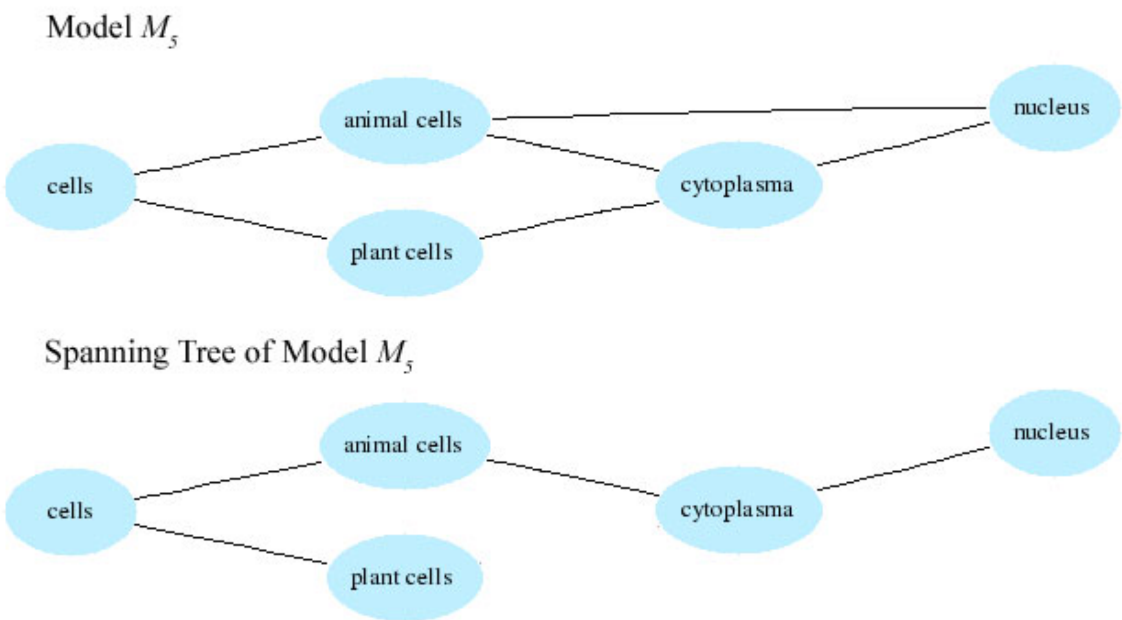

FIGURE 3.5. *Model M5 and its corresponding spanning tree*

A diameter is defined as the quantity of links of the shortest path between the most distant nodes. For the calculation of the *Matching Structure* index, the spanning tree is transformed into a distance matrix *D*.

$$D = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 1 \\ 4 & 3 & 2 & 1 & 0 \end{pmatrix} \qquad [1.3]$$

The *Matching Structure* index is calculated as the maximum value of all entries in the distance matrix D. The diameter or *Matching Structure* of the spanning tree in Figure 3.5 is calculated as follows:

$$\mu = \max_{i,j}\{d(i,j)\} = 4 \qquad [1.4]$$

The change in range or complexity of a person's model M_i is our second key indicator for the analysis of learning and problem solving processes (Seel, et al., 2009). Further graph theoretical such as *maximum circumference* (all possible relations), *ruggedness* (quantity of sub models which are independent or not linked), *linking density* (quotient of actual amount of relations and the total amount of possible relations), or *node centrality* (weight of a single node within a model) can be used to describe and analyze the structure of a model M_i in more detail.

Deep structure

The semantic composition of a model M_i is measured on the *Deep Structure*. The *Deep Structure* is calculated with the help of the similarity measure (Tversky, 1977) as the semantic similarity between an individual model M_i and a reference model M_r. A reference model M_r is defined as a subject domain-specific model (e.g. expert solution; another subject's model; the same subject's model constructed at a different time point).

In contrast to the graph theory-based calculation of the *Surface* and *Matching Structure*, model analysis on the *Deep Structure* is realized through a similarity calculation between a model M_i and a domain-dependent reference model M_r. Hence, a reference model M_r of high quality is a necessary precondition for a comprehensive analysis of the *Deep Structure*.

A similarity measure describes the degree of similarity between two objects, represented by a number between 0 and 1. Decisive for a similarity measure are objects with similar and different features. Tversky (1977) considered an object as an amount of features. The identification of a similarity between two objects is realized through a comparison of their features. The similarity formula takes not only the amount of similar features into account, but also the amount of different features. Lin (1998) defines similarity with the following three statements:

1. The similarity between *A* and *B* is related to their commonality. The more commonality they share, the more similar they are.
2. The similarity between *A* and *B* is related to the differences between them. The more differences they have, the less similar they are.
3. The maximum similarity between *A* and *B* is reached when *A* and *B* are identical, no matter how much commonality they share.

Accordingly, the smallest similarity between two objects A and B is given if no common features exist. In this case, the two objects are completely different and the similarity measure is 0. The similarity measure increases with a rise in the number of common features. A complete similarity of all features results in a similarity measure of 1.

The similarity of models on the *Deep Structure* is identified through the feature „proposition“ – the semantic characteristic of the proposition. The *Deep Structure* index δ is defined as the Tversky (1977) similarity between a model M_i and a reference model M_r. In general, we calculate:

$$\delta = \frac{f(A \cap B)}{f(A \cap B) + \alpha \cdot f(A - B) + \beta \cdot f(B - A)} \qquad [1.5]$$

A and *B* are the amount of propositions of a model comparison. The function *f(M)* corresponds to the number of elements in the amount *M*. The parameters α and β control the weighting of similar and different features. Both similar and different features are considered in the calculation if the weighting of α and β is equal ($\alpha = \beta = 0.5$). The value of the *Deep Structure* index δ is defined between 0 (no semantic similarity between the models) and 1 (absolute similarity between the models).

The *Deep Structure* or semantic similarity between model M_6 and reference model M_r is calculated in an automated iterative process. Every proposition in model M_6 is analysed for similarity with every proposition in the reference model M_r. The *Deep Structure* index is calculated as follows:

$$\delta = 0.57 \qquad [1.6]$$

Thus, the semantic similarity between model M_6 and reference model M_r is $\delta = 0.57$ or 57%. The quantitative measures of the *Surface*, *Matching*, and *Deep Structure* can be used for further statistical analysis. A qualitative analysis is made possible with the standardized re-representations of the *SMD Technology*.

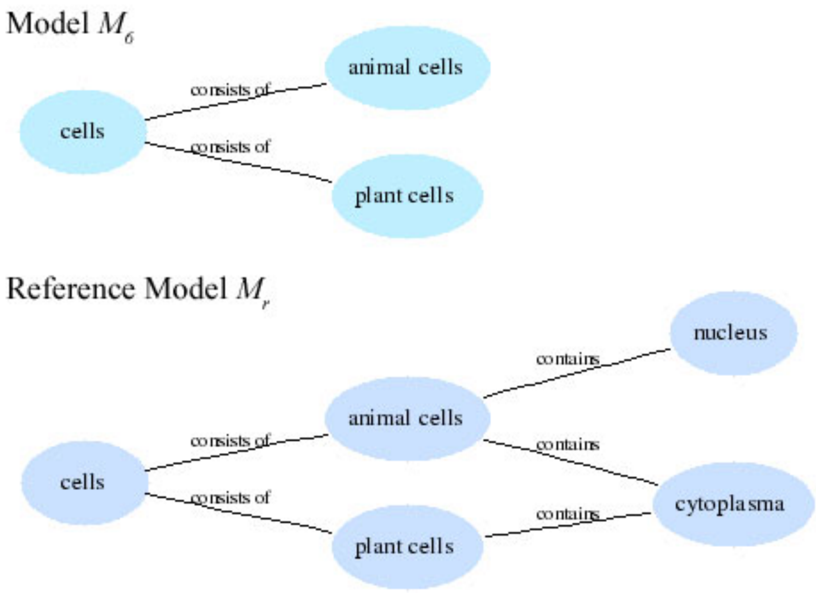

FIGURE 3.6. *Model M_6 and reference model M_r*

Standardized re-representations

The standardized graphical re-representation of the subject's data is constructed as an undirected or directed graph with named nodes and links. This automated feature of the *SMD Technology* is realized with the help of the open source graph visualization software *GraphViz* (Ellson, Gansner, Koutsofios, North, & Woodhull, 2003). For every single analysis, four standardized *PNG* (Portable Network Graphics) images are generated. Images (1) and (2) are the re-representations of model M_i and reference model M_r (for an example see Figure 3.6). Image (3) represents the *similarity model*, including only the nodes and links which are semantically similar between model M_i and reference model M_r (see Figure 3.7).

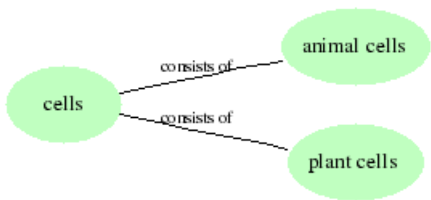

FIGURE 3.7. *Similarity re-representation of model M_6 and reference model M_r*

Image (4) is defined as the *contrast model*. It includes only nodes and links which have no semantic similarity within model M_i and reference model M_r (see Figure 3.8).

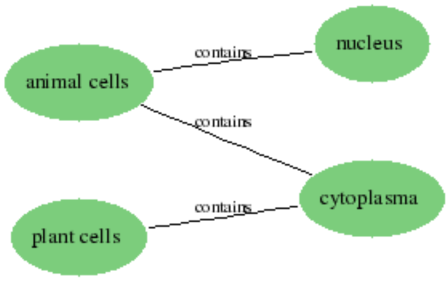

FIGURE 3.8. *Contrast re-representation of model M6 and reference model Mr*

Validation study

To investigate the objectivity, reliability, and validity of the computer-based and automated *SMD Technology*, we conducted three quasi-experimental studies. The objectivity of the *SMD Technology* was guaranteed by the computer-based and automated realization of the instrument. In the following section we report our results for reliability and validity of the *SMD Technology*.

Subjects

Three quasi-experimental studies (Studies 1, 2, and 3) were conducted with 106 subjects (70 female and 36 male) at the University of Freiburg. Their mean age was 18.3 years (SD = 4.6). The subject domain of Study 1 was geology and that of Studies 2 and 3 was geophysics. The subjects spent five hours on successive days working on complex problems with a multimedia discovery-learning environment.

Learning environment

The multimedia discovery-learning environment consisted of four modules. The modules could be divided into declarative and heuristic modules. The declarative modules contained all information needed to solve the phenomenon in question, while the heuristic modules primarily supported the model building process (Dummer & Ifenthaler, 2005).

Starting from the *problem & learning task* area, the subjects solve complex tasks from specific subject domains (Study 1: geology; Studies 2 and 3: geophysics). The subjects can navigate through different topics of the subject domain within the *curriculum module*. Additional information about the subject domain is provided in the form of various text documents, pictures, and audio recordings in the *knowledge archive*. The *Model Building Kit (MoBuKi)* provides the subjects with information about models, model building, and analogical reasoning. It contains three levels of abstraction of the material provided: (1) knowledge level; (2) procedural level; and (3) examples level. The *toolbox* is used to elicit the subjects' understanding of the phenomenon in question constructing open concept maps.

Procedure

The three quasi-experiments took place in the computer laboratory at the University of Freiburg. Subjects had to solve a complex problem while working with a multimedia discovery-learning environment. The problem solution had to be elicited on six subsequent measurement points as an open concept map. Every subject was given an introduction to the use and construction of open concept maps.

All subjects were randomly assigned to three types of treatments. The groups were distributed as (a) *scaffolding-based learning*, (b) *self-guided learning*, and (c) *control group*. The subjects in group (a) received detailed feedback concerning their concept map during the model building process, subjects in group (b) received no feedback, and subjects in group (c) received no feedback and worked within a multimedia discovery-learning environment whose content was not linked to the complex problem to be solved. The quasi-experimental procedure consisted of three main parts:

1. *Pretest*: Before the subjects were able to access the multimedia discovery-learning environment, a pretest was conducted which included: (a) the domain specific knowledge test; (b) elicitation of the preconception of the complex problem to be solved as an open concept map; (c) a test on cognitive learning strategies (LIST-Test); (d) a test on intellectual abilities (BIS-Test).
2. *Model building process*: During the quasi-experimental session, the subjects were asked to solve a complex problem while working within the multimedia discovery-learning environment. At five measurement points, the subjects had to elicit their understanding of the complex problem in question as an open concept map.
3. *Posttest*: The individual learning outputs were captured with: (a) a domain specific declarative knowledge test; (b) elicitation of the final solution to the complex problem as an open concept map.

The primary interest of the empirical investigation in this article is the experimental validation of the *SMD Technology*. Therefore, we focus in the following section on reliability and validity tests. However, details on the learning-dependent progression of externalized models and treatment effects during the three quasi-experiments are reported in detail by Ifenthaler (2006) and Ifenthaler, Pirnay-Dummer, and Seel (2007).

Reliability test

For the computation of the test-retest reliability (Spearman's rank correlation), the *Surface*, *Matching,* and *Deep Structure* indices of measurement points three and four (control group) were used.

TABLE 3.2
Test-Retest Reliability of the *SMD Technology*

	Test-retest reliability
Surface Structure	.824**
Matching Structure	.815**
Deep Structure	.901**

** p < .01 (two-sided significance)

The results in Table 3.2 show a high significant correlation between the indices (*Surface*, *Matching*, and *Deep Structure*). Accordingly, this result is a broad hint for the reliability of the quasi-experimental study. On the other hand, we want to point out that mental models are individual ad hoc constructions (Seel, 1991), and therefore standard reliability tests, e.g., *Test-Retest-*, *Split-Half-* or *Odd-Even-Method* (Rost, 2005), have only limited validity as they consider the latent variable to be stable. However, the detailed research design of the three quasi-experimental studies and the applied learning environment guarantee at least an exact repeatability of the experiments.

Validity test

Especially with newly designed and developed instruments (*e.g., SMD Technology*), it is necessary to map theory based characteristics to measurable criteria. The goal of the construct validation is to determine from a theoretical point of view what the instrument really measures. For this purpose, several methodological best practices[1] are available (see Lienert & Raatz, 1994). A comprehensive analysis of the theory of mental models (Johnson-Laird, 1983) and available instruments for the assessment of

[1] Correlation of a test with several outside criteria; Correlation with tests with similar validation requirements; correlation with tests that assess other criteria; analysis of inter- and intraindividual differences in test results; factorial analysis (see Lienert & Raatz, 1994).

models constitutes the basis for the theory-based development of the *SMD Technology*. From an empirical point of view, the validity of the *SMD Technology* is identified with the outside criterion (1) MITOCAR, and (2) domain specific knowledge.

Pirnay-Dummer (2006) developed the instrument MITOCAR (**M**odel **I**nspection **T**race **O**f **C**oncepts **A**nd **R**elations), which enables a structural and conceptual analysis of natural language expressions. The raw data of the third quasi-experimental study (N = 47) was analyzed with the MITOCAR software, which was tested for reliability and validity (Pirnay-Dummer, 2006). In the following, we use the results of the MITOCAR analysis for validity tests of the *SMD Technology*.

TABLE 3.3
Correlation between the SMD Technology and MITOCAR (N = 47)

	MITOCAR (concept and structure)	Surface Structure	Matching Structure
MITOCAR (concept and structure)	-	.610**[1]	.527**[1]
Surface Structure		-	.766**[1]
Matching Structure			-

** p < .01; * p < .05 (two-sided significance)
[1] Pearson's Correlation

The results in Table 3.3 show significant correlations between the outside criterion MITOCAR and the *Surface* and *Matching Structure* of the *SMD Technology*[2]. After verifying convergent validity of the *SMD Technology*, we want to test the *SMD Technology* with another outside criterion. This second validity test is for divergent validity on the basis of a valid and reliable *domain specific knowledge test* consisting out of 19 multiple-choice questions (Couné, Hanke, Ifenthaler, & Seel, 2004). We assume that there is no correlation between the *Surface* and *Matching Structure* of the *SMD Technology* and the *declarative knowledge* measure. Further, we assume a correlation between the *Deep Structure* and the *declarative knowledge*.

The results in Table 3.4 show no correlations between the *declarative knowledge* and the *Surface* and *Matching Structure*. This is consistent with the theoretical and methodological assumptions of the *SMD Technology* - the indices of the *Surface* and *Matching Structure* have no direct connection to the subject domain. The significant correlation between the declarative knowledge and the *Deep Structure* confirms the assumptions of the *SMD Technology* – we assume that persons with high declarative knowledge in a specific subject domain will also have

[2] The *Deep Structure* index δ of the *SMD Technology* compares the semantic similarity between a model and a reference model. This feature is not available with MITOCAR. Accordingly, the calculation of correlations between the *Deep Structure* and the MITOCAR indices is not necessary.

a high *Deep Structure* index δ. To sum up, the empirical analysis revealed convergent and divergent validity with regard to the outside criterion. Additionally, the *SMD Technology* was part of a series of comparative studies of different quantitative and qualitative methodologies conducted in order to determine the methodologies' strength and unique characteristics and to report collective validity (see T. E. Johnson, O'Connor, Spector, Ifenthaler, & Pirnay-Dummer, 2006).

TABLE 3.4
Correlation between the SMD Technology and the declarative knowledge test (N = 47)

	declarative knowledge	Surface Structure	Matching Structure	Deep Structure
declarative knowledge	-	.273[1]	.112[1]	.355*[2]
Surface Structure		-	.766**[1]	.089[2]
Matching Structure			-	.166[2]
Deep Structure				-

** p < .01; * p < .05 (two-sided significance)
[1] Pearson's Correlation; [2] Spearman's Correlation

Applications for research, learning, and instruction

The use of different computer-based tools for re-representing knowledge structures (e.g. concept mapping software) has become increasingly accepted for research, learning, and instruction (Jonassen, Reeves, Hong, Harvey, & Peters, 1997). In various research projects, concept maps have been used for analyzing learning outcomes, learners' knowledge structures, and for self-assessment (Eckert, 2000; Mansfield & Happs, 1991; Stracke, 2004). In the field of learning and instruction, concept maps have been used for providing feedback and advance organizers and for facilitating problem solving tasks (Al-Diban, 2002; Jonassen, et al., 1997; Stoyanova & Kommers, 2002). However, a large number of the available tools do not support automated feedback and analysis features. Accordingly, the development of the computer-based and automated *SMD Technology* opens up a broad field of applications for research, learning, and instruction.

SMD & research

Re-representations of knowledge structures are often analyzed by raters using diverse scoring approaches (see Hilbert & Renkl, 2008; Jonassen, et al., 1997; Taricani & Clariana, 2006). Depending on the research question, the raters focus on the quantity and quality of nodes and links, causal relationships, semantic content, direction and strength of links, hierarchy, or other visual arrangements. However, measuring the diverse information of individual concept maps by hand is very time consuming, and almost impossible for larger sets of data. Additionally, to guarantee high reliability and validity, every human rater must be an expert in the subject domain in question and in the application of quantitative and qualitative assessment strategies (Taricani & Clariana, 2006). Therefore, the automated analysis procedure of the *SMD Technology* calculates quantitative indicators of concept maps, which then can be used for further statistical computations.

So far, the *SMD Technology* has been applied in different fields of mental model research. Ifenthaler (2006) investigated the trajectory of mental models constructed by subjects working on complex problem solving tasks. An HLM analysis of three quasi-experimental studies (N = 106) showed a significant increase of propositions when subjects worked for five hours in a multimedia learning environment (*Surface Structure*). Accordingly, as long as new information is subjective plausible it will be added to a person's knowledge structure. Further results indicate a significant increase in the diameter of the externalized knowledge structures (*Matching Structure*). Consequently, we found not only a significant learning-dependent increase in the number of propositions, but also a significant learning-dependent increase in structural complexity.

In order to investigate the learning-dependent progression of novices' mental models to more expert-like models, Ifenthaler (2006) compared the semantic similarity of externalized knowledge structures of novices with expert knowledge structures in different subject domains. The results of the *Deep Structure* indicator of the *SMD Technology* revealed a significant increase in similarity between novice and expert models. However, further HLM analysis indicated that the learning time of five hours was not long enough to integrate all information provided and consequently to gain higher similarity to an expert's solution of a problem. Predictions about novice's problem solving skills to become more expert like are also possible (e.g., Ifenthaler, et al., 2007). Additionally, the provided learning materials and feedback could be improved for further experiments.

Ifenthaler et al. (2007) investigated the role of cognitive learning strategies and intellectual abilities in mental model building processes using the *Deep Structure* indicator of the *SMD Technology*. The results indicate that the training of mental model building skills is a complex problem which should be investigated further with regard to the roles of conditions based on the theory of mental models (Seel, 1991).

Additionally, the *SMD Technology* has been used to investigate sharedness among team members (T. E. Johnson, Ifenthaler, Pirnay-Dummer, & Spector, 2009). The focus on individually constructed concept maps and team re-representations can help to identify problems of team performance and lead to a better understanding of the complex performance processes within teams. Thanks to the flexibility of the *SMD Technology*, other indicators can be easily implemented in order to produce specific measures for a large number of research questions.

SMD & learning and instruction

In the following, we will focus on the application of the *SMD Technology* for *knowledge diagnosis*, *self-assessment*, and *knowledge management*. Other applications in the field of learning and instruction, such as *analysis of navigation paths* in learning environments (Dummer & Ifenthaler, 2005), could be discussed on another occasion.

In order to provide learners with the best possible learning materials, the instructor or an Intelligent Tutoring System (ITS) must be aware of their state of knowledge. In general, *knowledge diagnosis* is applied by collecting necessary information about the learner with the help of various tests. By integrating the *SMD Technology* or parts of it (graphical re-representation; quantitative indicators) either into a computer-based learning environment or other instructional settings, it can easily be applied for individual knowledge diagnosis. The *SMD Technology* has been implemented as a cross-platform application which enables an easy integration into a computer-based learning environment. Therefore, the instructional designer may choose which components of the *SMD Technology* should be applied for an adequate knowledge diagnosis. The quantitative indicators could provide instant longitudinal information about the individual learning process. The indicators (Surface, Matching, and Deep) provide multiple information about changes in the knowledge structure and domain-specific knowledge acquisition. Depending on the results of the *SMD Technology*, the learning environments will provide specific feedback or other instructional materials to foster future learning processes. On the other hand, the graphical re-representation of the *SMD Technology* can be easily applied for individual feedback on specific tasks. The instructor could use the re-representation at a specific point during the learning phase to discuss the strength and weaknesses of a learner's learning process. Additionally, the similarity and contrast model provide further feedback materials.

Another use of the *SMD Technology* in the field of learning and instruction could be various fields of *self assessment*. As self assessment has the ambitious goal of making judgments about a learner's own learning process, the feedback of an automated system should be very sensible to changes in the learner's knowledge structure. As discussed above, the quantitative indicators and/or graphical re-representations of the *SMD Technology* could be applied for self assessment. A learner could receive quantitative information about his or her learning progress after working for a defined period with a computer-based learning environment.

Additionally, the graphical re-representation could provide descriptive information about the learner's knowledge structure. Furthermore, the similarity and contrast representation could elicit differences between previous points during the learning process or other learners or experts. This feature could therefore easily help to avoid the construction of misconceptions during self assessment phases. The major advantage of the *SMD Technology* for self assessment is the automated and instant generation of desired results. When learners receive the results of self assessment directly, their motivation to continue with the learning environment may be obtained longer than with other options of self assessment.

Finally, the *SMD Technology* could be applied for analysis of *knowledge management* processes. Individuals may use the quantitative indicators and or the graphical re-representations to compare it with other team members while working on a project. Also, the affordances of a task could be compared with the individual understanding of the task and gaps could be identified to solve it effectively. Another application of the *SMD Technology* for knowledge management could be the communication of individual or group knowledge for better cooperation and understanding with other members or groups of a project team. Further applications could include knowledge identification, knowledge use, and knowledge generation (Tergan, 2003).

Conclusion and future perspectives

The new developed *SMD Technology* is based on the theory of mental models (Seel, 1991) and graph theory (Tittmann, 2003) and captures key latent variables associated with human learning and cognition. Graphical representations such as concept maps or natural language expression can be analyzed on three different levels. These levels help to describe individual knowledge structures from a relational, structural, and semantic point of view. Additionally, graphical re-representations of the *SMD Technology* provide further information regarding the externalized knowledge structures of a person.

The objectivity, reliability, and validity of the computer-based and automated *SMD Technology* were investigated in three quasi-experimental studies. The results show a high reliability and validity in all indicators. Based on our findings, we developed further ideas for developing new features for the *SMD* Technology. These developments will include a tool for constructing concept maps, new techniques for describing the constructed models, and automated statistical reports.

Nevertheless, the *SMD* Technology or parts of it (graphical re-representation; quantitative indicators) can be easily integrated into various applications. The tool can be used not only in mental model research, but also in various fields of learning and instruction. Beyond this, such computer-based and automated instruments could also prove to be beneficial in a wide span of other fields of research on technology and instructional development.

4

DETERMINING STRENGTHS AND LIMITATIONS OF METHODOLOGICAL APPROACHES 🕮

Over the past years, several possible solutions to the analysis problems of mental models have been discussed. Therefore, it is worthwhile to compare analysis approaches for measuring externalized mental models systematically in order to test their advantages and disadvantages, strengths and limitations. A series of pair-wise comparative studies show strengths, unique characteristics, and collective viability of different assessment and analysis methods. However, the above mentioned study only focused on conceptual differences of the analysis approaches and did not use empirical data. This chapter reports an empirical case study and compares two analysis approaches - *QFCA* (Qualitative & Formal Concept Analysis) and *SMD* (Surface, Matching, Deep Structure) - using identical data. Accordingly, the aim of this comparative study is to determine *conceptual* and *empirical* strengths and limitations of two different approaches for analyzing externalized cognitive structure.

🕮 This chapter is based on: Al-Diban, S., & Ifenthaler, D. (in press). Comparison of two analysis approaches for measuring externalized mental models: Implications for diagnostics and applications. *Journal of Educational Technology & Society*.

Introduction

Mental models are a basic cognitive construct which describes complex learning and problem solving processes. Generally speaking, a person constructs a mental model in order to explain or simulate specific phenomena of objects or events if no sufficient schema is available. Thus, mental models organize domain specific knowledge in such a way that phenomena of the world become plausible for the individual. Compared to that of a novice, a domain expert's mental model is considered to be more elaborated and complex. Therefore, we argue that mental models mediate between an initial state and a desired final state in the learning process. Accordingly, there is an immense interest on the part of researchers to analyze a novice's mental model and compare it with an expert's in order to identify the most appropriate ways to bridge the gap.

Over the past years, several possible solutions to the analysis problems of mental models have been discussed (e.g., Clariana & Wallace, 2007; Ifenthaler, 2008; T. E. Johnson, et al., 2009). Therefore, it is worthwhile to compare analysis approaches for measuring externalized mental models systematically in order to test their advantages and disadvantages, strengths and limitations. Johnson et al. (2006) set up a series of pair-wise comparative studies in order to determine the strength, unique characteristics, and collective viability of different assessment and analysis methods. A total of six studies compare the methods ACSMM (Analysis Constructed Shared Mental Models; T. E. Johnson, et al., 2009), SMD (Surface, Matching, Deep Structure; Ifenthaler, 2010c), MITOCAR (Model Inspection Trace of Concepts and Relations; Pirnay-Dummer & Ifenthaler, 2010), and DEEP (Dynamic Evaluation of Enhanced Problem Solving; Spector & Koszalka, 2004). Through study of their methodologies, the authors hope to better quantitatively and qualitatively represent individual and team mental models and better understand mental model development by comparing individuals and experts (T. E. Johnson, et al., 2006). However, the above mentioned study only focused on conceptual differences of the analysis approaches and did not use empirical data.

In addition to the above described comparative study by Johnson et al. (2006), our current study compares two analysis approaches - *QFCA* (Qualitative & Formal Concept Analysis) and *SMD* (Surface, Matching, Deep Structure) - using identical data. Accordingly, the aim of our comparative study is to determine *conceptual* and *empirical* strengths and limitations of two different approaches for analyzing externalized mental models. Our comparison framework is laid out as follows: First, both analysis approaches are introduced. Second, we present the empirical study. Third, we report the results analyzed with both approaches, QFCA and SMD. Forth, on the basis of our results, we compare both analysis approaches. Finally, we conclude by determining how the two approaches could be used in conjunction for further mental model research.

Analysis approaches

A mental model is always content related and the assessment (elicitation) and analysis (measurement of elicitation) should allow a psychological and content based interpretation. However, the yet unsolved question is how to accurately diagnose mental models. Some issues that have yet to be resolved include identifying reliable and valid ways to elicit mental models and the actual analysis of the externalized models themselves (Ifenthaler & Seel, 2005; Kalyuga, 2006a). However, the possibilities of assessment (elicitation) of mental models are limited to a few sets of sign and symbol systems (Seel, 1999b) – characterized as *graphical* and *language-based approaches*. Graphical approaches include the structure formation technique (Scheele & Groeben, 1984), pathfinder networks (Schvaneveldt, 1990), mind tools (Jonassen & Cho, 2008), and test for causal models (Al-Diban, 2008). Language-based approaches include thinking-aloud protocols (Ericsson & Simon, 1993), cognitive task analysis (Kirwan & Ainsworth, 1992), and computer linguistic techniques (Seel, et al., 2009). However, not all of these elicitation methods interact with available analysis approaches. Therefore, we identified two analysis approaches (QFCA and SMD) which interact well with the graphical assessment method *test for causal models* (TCM).

Analysis I: Qualitative & formal concept analysis (QFCA)

As a first step of the QFCA, the amount of assessed data (graphical or natural language-based) will be reduced semi-automatically with help of coders, which look for semantic similarities, synonyms, and metaphors and build hierarchies of concepts and propositions. Second, the data is imported into Cernato (Navicon, 2000). This program is based on lattice theory (Birkhoff, 1973) and allows content based comparisons of individual mental model representations. Figure 4.1 shows an example of the results of an analysis. The figure presents a comparison of the preconceptions of 12 participants on the level of generic concepts. In the third step of the analysis the problem of structure isomorphism occurs, which usually prevents content based comparisons of simple concept mapping methods (see Nägler & Stopp, 1996). This problem consists of the possibility that any number of identical concepts can be connected in the factorial number of arrays. This makes it nearly impossible to make content based comparisons of entire model representations. With the help of formal concept analysis (Ganter & Wille, 1996) all objects (here participants) can be systematically structured according to the entirety of all true attributes (here concepts or propositions).

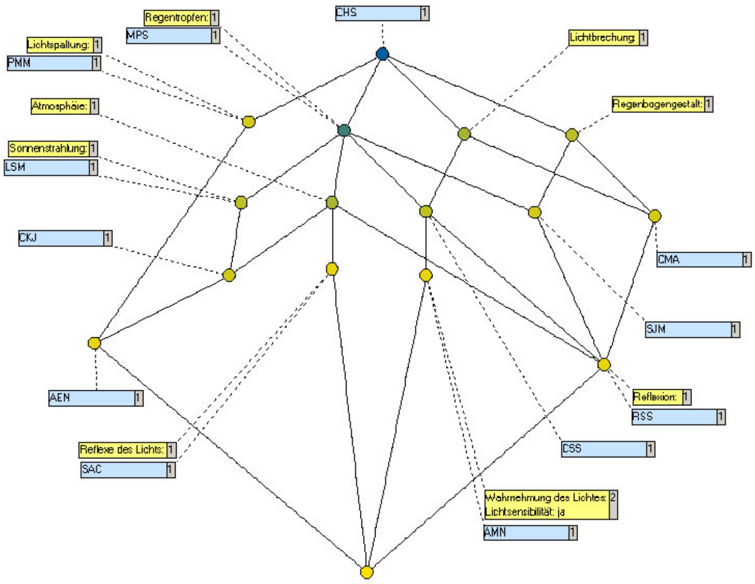

FIGURE 4.1. *QFCA analysis of the "rainbow phenomenon"*

Accordingly, the formal concept analysis follows the following procedure: (a) Since the data is preserved for the most part in natural language, it is possible to reconstruct incorrect or missing concepts in the preconceptions of the participants (e.g., decomposition of light instead of color dispersion; a biological reflex instead of a physical reflex) and then discover any exceptional concepts participants used. (b) The whole of semantic surface features are preserved and can be compared. This allows us to, e.g., distinguish between participants with a low and high amount of prior knowledge. (c) Since concept "volume" is defined by all objects which can be reached by downward lines (see Figure 4.1), we are able to reconstruct which participants used, e.g., the concept "raindrop" (only 9 of the 12 participants). (d) We are able to analyze special questions (sections) in detail, e.g. what characterized the preconceptions of the participants who used the concept "rainbow figure" – two used "refraction" (RSS, CMA) and one also used "reflexion" (RSS). However, no one used "dispersion," "perception," "sensibility for light," or "solar radiation." Research designs with more than one point of measurement would allow very interesting content-based comparisons of changes.

Analysis II: Surface, matching, deep structure (SMD)

The advent of powerful and flexible computer technology enabled us to develop and implement a computer-based analysis approach which is based on the theory of mental models and graph theory (Chartrand, 1977). SMD uses three core measures for describing and analyzing externalized mental models (Ifenthaler, 2010c). Additional measures are applied for an in-depth analysis (Ifenthaler, et al., in press). SMD requires for the assessed data to be stored pairwise (vertex-edge-vertex) for further analysis procedures. If the required data format is available (see Table 4.1), the raw data can be stored on an SQL (structured query language) database and the automated analysis procedure can be initiated by the researcher.

TABLE 4.1
Example of pair-wise raw data

ID	vertex 1	vertex 2	edge	subject number
001	Licht	Ausbreitung	!	912abz3
001	Licht	Spalt	-	912abz3
…	…	…	…	…

As a result, SMD generates three core measures, additional measures, and standardized graphical re-representations of the previously externalized mental models. These re-representations are concept map-like images with named nodes and named links (e.g., Figure 4.2).

FIGURE 4.2. *SMD re-representation of data shown in Table 1*

The core measures are composed of three levels – surface, matching, and deep structure. The *surface structure* measures the size of the externalized model, computed as the sum of all propositions (vertex-edge-vertex). It is defined between 0 (no propositions) and *n*. The computed *surface structure* of the re-represented model in Figure 4.2 would result in $\theta = 3$. The pedagogical purpose is to identify additions or removals of vertices (growth or decline of the graph) as compared to previous knowledge representations and track change over time.

In order to analyze the complexity of an externalized model, Ifenthaler (2010c) introduced the *matching structure* μ. It is computed as the diameter of the spanning tree of an externalized model and can lie between 0 (no links) and n. The complexity indicator of the re-represented model in Figure 4.2 would result in $\mu = 2$. The pedagogical purpose is to identify how broad (complex) the learner's understanding of the underlying subject matter is.

Whereas the two above described measures focus on analyzing the *organization* or *structure* of an externalized model, the *deep structure* measures its *semantic content*. It is computed with the help of the similarity measure (Tversky, 1977) as the semantic similarity between an externalized model and a reference model (e.g., expert solution, conceptual model, etc.). The measure is defined between 0 (no similarity) and 1 (full similarity). The pedagogical purpose is to identify the correct use of specific propositions (concept-link-concept), i.e. concepts correctly related to each other. Additionally, misconceptions can be identified for a specific subject domain by comparing known misconceptions (as propositions) to individual knowledge representations.

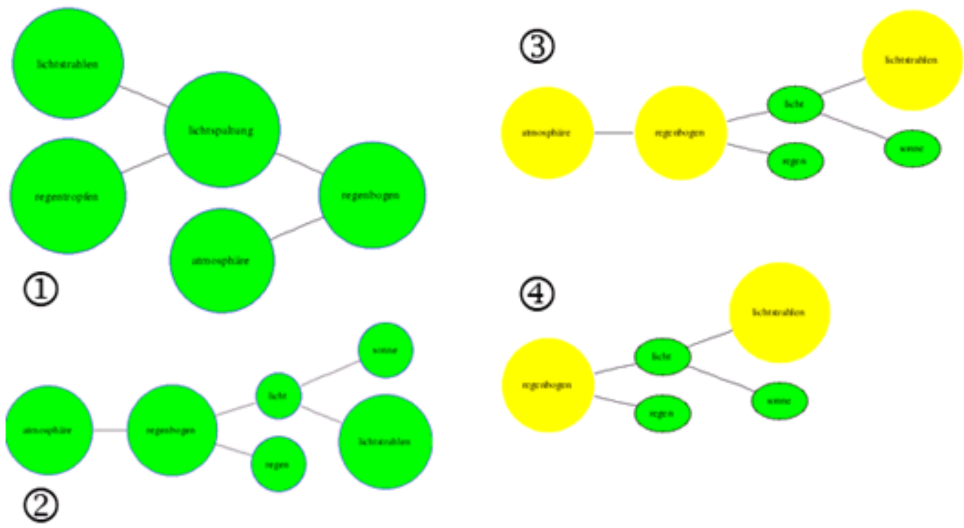

FIGURE 4.3. *SMD reference (1), learner (2), cutaway (3), and discrepancy (4) re-representations*

In addition to the core measures, further graph theory based indicators are applied to more precisely describe the externalized mental models. With regard to analyzing the *organization* of the externalized models, Ifenthaler and colleagues (in press) introduced the measures *connectedness*, *ruggedness*, *cyclic*, *average degree of vertices*, *density of vertices* and *structural matching*.

1. The indicator *connectedness* analyses how closely the nodes and links of the externalized model are related to each other. The *connectedness* measure of the re-represented model in Figure 2 would result in $\phi = 1$ (it is possible to reach every node from every other node). From educational point of view, a strongly connected knowledge representation could indicate a subjective deeper understanding of the underlying subject matter.
2. *Ruggedness* indicates whether non-linked vertices of an externalized model exist, and if so it computes the sum of all submodels (a submodel is part of the externalization but has no link to the "main" model). The pedagogical purpose is to identify possible non-linked concepts, subgraphs or missing links within the knowledge representation which could point to a lesser subjective understanding of the phenomenon in question.
3. The measure *cyclic* is an indicator for the closeness of associations of the vertices and edges used. A *cycle* is defined as a path returning back to the start vertex of the starting edge of an externalized model. A cycle in the re-represented model in Figure 4.2 would be: Licht – Ausbreitung – Spalt – Licht.
4. The *average degree of vertices* measure is computed as the average degree of all incoming and outgoing edges.
5. The *density of vertices* indicator describes the quotient of concepts per vertex within a graph. Graphs which only connect pairs of concepts can be considered weak models; a medium density is expected for most good working models.
6. The *structural matching* measure compares the complete structures of two graphs without regard to their content. This measure is necessary for all hypotheses which make assumptions about general features of structure (e.g., assumptions which state that expert knowledge is structured differently from novice knowledge).

The pedagogical purpose of these measures is to identify the strength of closeness of associations of the knowledge representation. Knowledge representations which only connect pairs of concepts can be considered weak; a medium density is expected for most good working knowledge representations. The additional *semantic* indicator *vertex matching* analyzes the use of semantically correct single concepts compared to a reference model. This measures is also used in the classic MITOCAR analysis procedure (see Pirnay-Dummer & Ifenthaler, 2010). The pedagogical purpose is to identify the correct use of specific concepts (e.g., technical concepts). The absence of a great number of concepts with regard to a reference representation indicates a less elaborated domain specific knowledge representation.

For an in-depth qualitative analysis, SMD automatically generates standardized re-representations. Figure 4.3 shows an example of a reference (1), learner (2), cutaway (3), and discrepancy (4) re-representation which also function as feedback within learning environments (Ifenthaler, 2009). These re-representations highlight semantically correct vertices (compared to a reference representation) as circles (ellipses for dissimilar vertices).

Various experimental studies on different subject domains have confirmed the high reliability and validity of the SMD (see T. E. Johnson, et al., 2006). Ifenthaler (2010c) reports test-retest reliability for SMD measures as follows: surface structure, r = .824, matching structure, r = .815, and deep structure, r = .901. Also convergent and divergent validity has been successfully tested (see Ifenthaler, 2010c).

Comparative study

This initial comparative study determines *conceptual* and *empirical* strengths and limitations of the above described approaches for analyzing externalized mental models – QFCA and SMD. In order to have identical data available, we conducted a study (pre-post design) in physics and theology with high school students. This section introduces briefly the study's methodology.

Subjects

The 12 participants (9 female, 3 male) of the reported pilot study were students in the 10^{th} grade from a traditional high school in Europe. Their mean age was 15.25 years (*SD* = .45), mean score CFT 20-R intelligence test = 106.92 (*SD* = 9.89). There were nine members of religious communities among the participants. Eight are active in their communities and eleven have religious interests. The participants volunteered in response to an advertisement posted at their school. After finishing the study each participant was given a reward of 20 Euros.

Materials

The overall design (see Figure 4.4) included an assessment of the preconceptions of the participants in physics and theology, which began with a free association test with scenic pictures of rainbows (physics) and tsunami (religion) which served as an "ice-breaker-function" for the topic. This was followed by word problems with written text protocols and a dependant measure of the same problems from the test of causal models (TCM, Al-Diban, 2008). The participants were assessed according to relevant traits like intelligence with the standardized test of intelligence CFT 20-R (Weiß, 2006). The *culture fair test* measures the fluid intelligence factor with figural material, which is a substantial indicator for inductive reasoning and flexibility of thinking. Relevant learning strategies were assessed with LIST (Wild, 2000). Additionally, we used the standardized Neo-FFI test (Borkenau & Ostendorf, 2006) to examine general self-concept, self-perceived self-efficiency (Schwarzer & Jerusalem, 1999), and personality. Furthermore, the assessment contained a test on domain specific declarative knowledge in physics and religion. Demographic data of the participants were documented in an informal questionnaire.

Assessment: Test for causal models (TCM)

This assessment instrument was developed in order to realize the postulated theoretical functions of mental models, such as high individuality, phenomenon relatedness, situational permanence, reduction of complexity, and knowledge gain (Al-Diban, 2008). The standardized TCM (Test for Causal Models) is a combination of the Structure Formation Technique (Scheele & Groeben, 1984) and Causal Diagrams (Funke, 1990) and is a practicable method for discovering structure which is in line with the theory of mental models. The participants have to transform their answers into subjectively relevant causal sequences of if-then relations or cause-consequence relations of the problem and its preconditions. The connections between single concepts represent the subjective causal thinking in a broad sense (van der Meer & Schmidt, 1992). A guided practice session in which the participants construct an example is provided in order to improve their competence in using the TCM. For the data assessment phase we used the computer based software MaNET (Mannheim Network Elaboration Technique, Reh, 2007) to enhance the usability for the participants and to allow a standardized data processing for the subsequent analysis process. Additionally, we used the purpose-built *graph to context* interface (GTC, Al-Diban & Stark, 2007) to export the assessed data and make them available to both analysis approaches, QFCA and SMD.

Procedure

All participants visited a learning lab at a European university on two subsequent days. The assessment procedure took three hours per day. The first part of the assessment consisted of a free association test, a demonstration of some slides with photographs of rainbows and life-threatening diseases. The participants had to write down all concepts they were spontaneously able to remember. All concrete problems, three in physics and three in religion, were measured twice: first as an open problem with transcribed text protocols from the teach back interview and second as a dependant measure which was constructed around these answers with the TCM. This test was conducted on laptops using the software MaNET. The working time was limited to 20 minutes. The participants had the task of depicting their answers with the help of a test of causal models (TCM) comprised of concepts and causal relations. The other traits measured in this test are shown in Figure 4.4.

Demonstration Free Association	**Teach Back Interview**	**Test for Causal Models**
Physics:		
I Rainbow		
II Crack experiment III Light electrical effect	General Explanations	Concrete Explanations : Why can we see a (rainbow ,..,..) ?
Biology & Religion:		
IV Disease situation	General Explanations	Concrete Explanations : Why do people fall ill?
Traits: intelligence , learning strategies , emotions , self concept , interests , attitudes		
Parallel Tasks & AquisitionMethods		

FIGURE 4.4. *Research design*

On the one hand the two different topics – light models in physics and disease models in biology in combination with religion – were oriented toward the curriculum and the courses of instruction. On the other hand, these topics should represent two very different knowledge domains. This allows us to compare the mental model representations of the same persons in very different knowledge domains. It should be emphasized that the results of this initial study are descriptive single cases only and not valid for a greater population group and general educational implications.

Results

The data collected in our study were analyzed with *QFCA* and *SMD* separately. Therefore, we describe our results in two separate sections and then compare the results of both analysis approaches. The “expert models” and “correct model concepts” applied to evaluate the semantic criteria of objective plausibility were developed with the help of specialists in physics education and theology. The expert models resulted in a rainbow (11 propositions), crack experiment (12 propositions), light electrical effect, (10 propositions) and disease situation model (18 propositions). The “correct model concepts” represent key concepts and are a precondition for understanding each phenomenon correctly. In all cases, the criteria of objective plausibility are dependent on the semantic correspondence of the student model to the propositions of the expert model.

As far as the measured traits are concerned, there was a negative correlation $r = -.625^{*}$ between the trait “agreeableness” (Neo-FFI) and knowledge on the level of concepts in physics but no significant correlation with concepts concerning the disease problem. The objective plausibility of all three model representations to physical problems together (sum of all the physic problems) and the learning strategy “critical thinking” shows a high and significant correlation $r = .869^{**}$, such as with “openness for new experiences” $r = .707^{*}$. This result might indicate that the objective plausibility of the investigated physical problems is associated with intensive “critical thinking” learning strategies and a high personal “openness for new experiences”.

Qualitative & formal concept analysis (QFCA)

The QFCA analysis approach includes five quantitative structural measures (count of concepts, count of propositions, depth of connectivity, intensity of connections, ruggedness) and an in-depth content-based investigation. Table 4.2 shows the results of the five quantitative structural measures. On a descriptive level, there are remarkable differences between the four problems for the measures count of concepts and count of propositions. The other structural measures, intensity of connections and ruggedness, show almost equal values with comparable standard deviations. The majority of the mental model representations of all problems have a low depth of connectivity, a low intensity of connections, and are not rugged. Additionally, the standard deviations show high interindividual differences in the "crack experiment" (II) and the "disease problem" (IV) for the measures count of concepts and count of propositions.

TABLE 4.2
QFCA structural measures

	DOMAIN	M	SD	Min	Max
count of concepts	I	7.08	2.64	4	13
	II	5.91	3.05	3	14
	III	5.67	1.12	4	7
	IV	9.09	3.02	6	15
count of propositions	I	6.75	3.31	3	14
	II	5.45	4.61	1	18
	III	5.3	1.50	3	8
	IV	12.36	5.68	5	22
depth of connectivity	I	1.08	0.16	0.83	1.33
	II	1.0	0.24	0.60	1.36
	III	1.12	0.18	1.00	1.50
	IV	1.39	0.27	1.00	1.89
intensity of connections	I	0.34	0.11	0.18	0.5
	II	0.39	0.16	0.19	0.67
	III	0.43	0.16	0.33	0.83
	IV	0.35	0.10	0.18	0.53
ruggedness	I	1.25	0.45	1	2
	II	1.27	0.65	1	3
	III	1.00	0.16	1	1
	IV	1.00	0.00	1	1

Note: *DOMAIN*: I = rainbow experiment (*N=12*), II = crack experiment (*N=10*), III = electrical effect experiment (*N=9*), IV = disease situation (*N=12*)

In the next step, we analyzed the results for generic conceptss and propositions and determined to what extent they corresponded to the expert models (see Table 4.3).

TABLE 4.3
Content based similarity measures between participant and expert solutions

	DOMAIN	M	SD	Min	Max
relative objective plausibility [propositions in %]	I	51.09	19.65	22.2	80
	II	33.70	38.22	0	100
	III	28.94	23.58	0	66.7
	IV	45.8	26.70	5.2	100
abs. objective plausibility [prop., max.11/12/10/18]	I	3.08	1.24	2	6
	II	1.20	1.03	0	3
	III	1.44	1.24	0	4
	IV	4.50	1.45	1	6
correct model concepts [6/7/8/20]	I	1.17	0.94	0	3
	II	1.10	0.74	0	2
	III	0.88	0.78	0	2
	IV	3.50	1.17	2	5

Note: *DOMAIN*: I = rainbow experiment (*N=12*), II = crack experiment (*N=10*), III = electrical effect experiment (*N=9*), IV = disease situation (*N=12*)

Focusing the averages of the match with the expert models - relative and absolute objective plausibility - can be called small in general. The minimum of most semantic criteria represents the mental models to the physic problem (III) “light electrical effect”. This problem seems to be most difficult for the participants. The solutions to the biology & theology problem “disease situation” were slightly more competent. The use of correct model concepts is very low for all problem solutions, too. This indicates that the participants did not possess sufficient concept knowledge, which is a precondition for mental models with high objective plausibility.

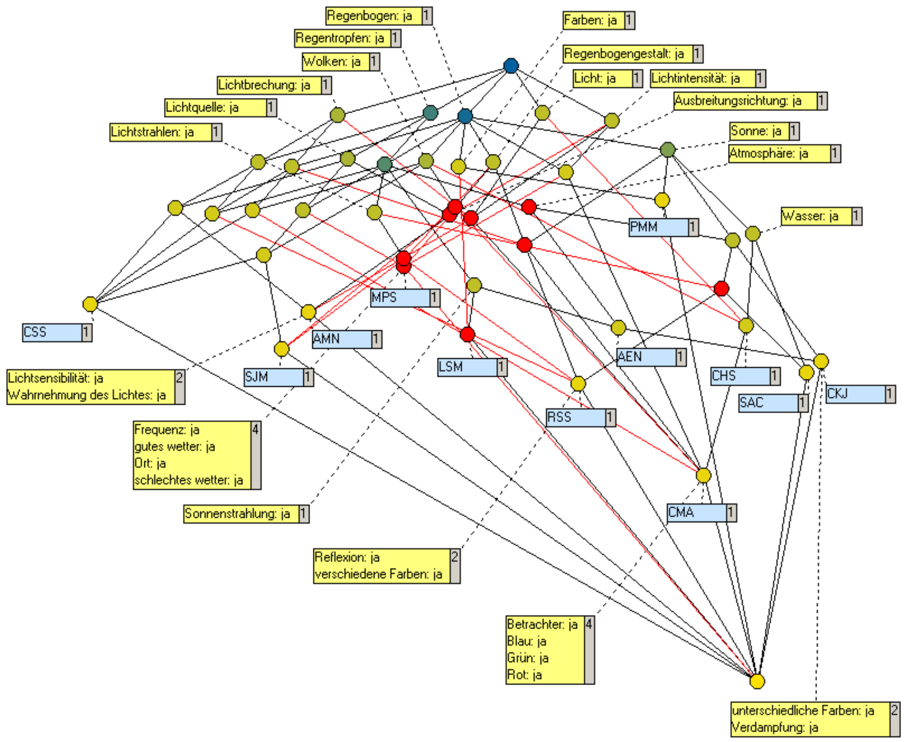

FIGURE 4.5. *Comparison of participants for domain specific problem (I)*

It is easy to see which of the correct model concepts from the expert model are present and which are absent. Basically, the preconceptions are based solely on the radiation model. The absent correct concepts are “diffraction,” “dispersion,” “light rays”, and a “constant color spectrum” in contrast to the simple concept “colors.” These mental model representations contain no elements to explain its color spectrum. Instead, some participants worked with the “figure of rainbow” and tried to find explanations for this.

In addition, QFCA allows content based comparisons of the single cases with small groups (see Figure 4.6). Clearly, the participants CKJ and CMA show more knowledge then the participants LSM and CHS. Moreover, this method displays the data in such a way that the content becomes obvious. In a comparison of participants

CHS and CMA – Figure 4.6 – there is empirical evidence, that they share all five concepts used by CHS. But CMA was able to supplement his preconceptions with adequate concepts like "intensity of light" and "refraction" and also spent time thinking about "figure of rainbow," "observer," and the colors "blue," "green," and "red."

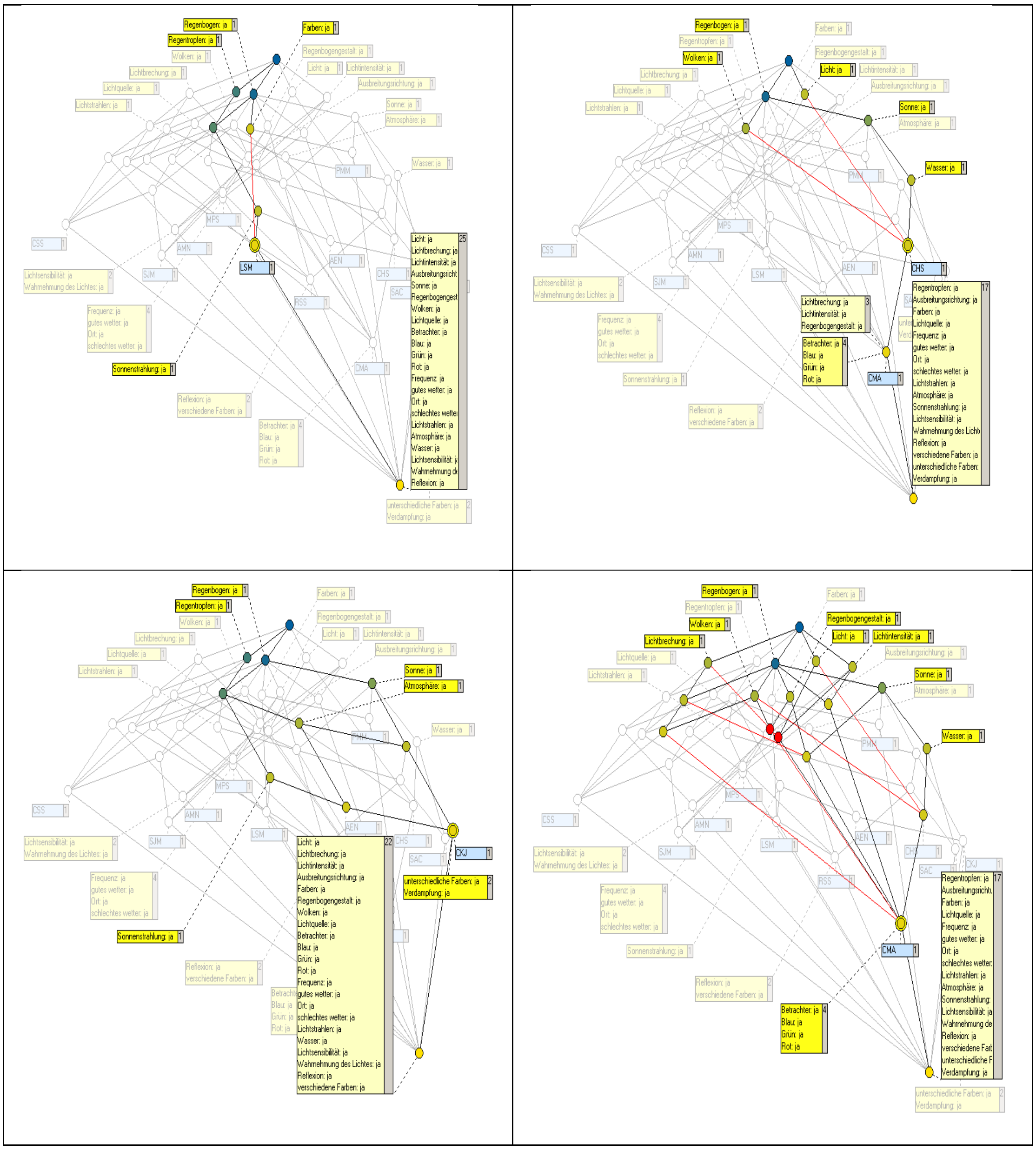

FIGURE 4.6. *Four single cases domain specific problem (I)*

In summary, QFCA can be a useful tool for making empirically based conclusions about mental model representations for single cases and small groups. It makes the content-based quality of preconceptions and special areas of interest easy to evaluate. With the help of data from more than one measurement point, conceptual changes become better and more accurately observable too.

Surface, matching, deep structure (SMD)

The automated analysis procedure of SMD generates the above described quantitative measures. The results for the three physics domains and biology & religion domain are presented in Table 4.4 and 4.5. As can be seen by the frequencies and the Kolmogorov-Smirnov one-sample tests, we found no interindividual differences between the subjects, except for the measures connectedness and ruggedness in the first physics domain (rainbow experiment), and for the measure cyclic in the biology & religion domain (disease situation).

TABLE 4.4
Structural SMD measures

	DOMAIN	M	SD	Min	Max	KS-Z	p
surface structure	I	14.25	7.26	1.00	26.00	.39	.998
	II	16.50	13.29	3.00	42.00	.53	.942
	III	5.56	1.42	3.00	8.00	.71	.692
	IV	12.42	6.36	5.00	27.00	.59	.872
matching structure	I	4.92	1.93	1.00	7.00	.67	.761
	II	3.90	1.52	2.00	7.00	.55	.923
	III	3.67	.71	3.00	5.00	.82	.520
	IV	5.00	1.95	3.00	10.00	.77	.601
connectedness	I	0.92	.29	0	1	1.84	.002*
	II	1	0	1	1	-	-
	III	1	0	1	1	-	-
	IV	1	0	1	1	-	-
ruggedness	I	1.08	.29	1	2	1.84	.002*
	II	1	0	1	1	-	-
	III	1	0	1	1	-	-
	IV	1	0	1	1	-	-
cyclic	I	.58	.51	0	1	1.29	.070
	II	.4	.52	0	1	1.20	.110
	III	.44	.53	0	1	1.07	.204
	IV	.75	.45	0	1	1.59	.013*
average degree of vertices	I	1.89	.27	1.5	2.29	.80	.542
	II	1.73	.46	1	2.43	.38	.999
	III	1.83	.26	1.5	2.29	.69	.723
	IV	2.29	.44	1.67	3.14	.44	.991
density of vertices	I	.51	.19	.22	1.00	.55	.925
	II	.40	.21	.19	.78	.79	.546
	III	.39	.13	.10	.50	.71	.699
	IV	.31	.14	.10	.50	.95	.328
structural matching	I	14.67	6.53	2.00	27.00	.57	.897
	II	11.80	6.34	5.00	26.00	.67	.761
	III	5.78	1.20	4.00	7.00	.72	.678
	IV	9.92	3.20	6.00	14.00	.78	.577

Note: *DOMAIN*: I = rainbow experiment (*N=12*), II = crack experiment (*N=10*), III = electrical effect experiment (*N=9*); IV = disease situation (*N=12*);*KS-Z* = Kolmogorov-Smirnov one-sample test; * $p < .05$; ** $p < .01$

In order to locate differences between the four domains, we computed conservative Kruskal-Wallis H-Tests. The frequencies of the surface structure between the domains were significantly different, $\chi2$ (3, N = 43) = 11.40, p > .05. We also found significant differences for the measures structural matching, $\chi2$ (3, N = 43) = 14.80, p > .05, vertex matching, $\chi2$ (3, N = 43) = 19.42, p > .001, and propositional matching, $\chi2$ (3, N = 43) = 11.36, p > .01. However, we found no significant differences for the remaining measures.

TABLE 4.5
Semantic SMD measures

	DOMAIN	M	SD	Min	Max	KS-Z	p
vertex matching	I	12.50	5.50	1.00	21.00	.95	.330
	II	10.70	6.17	3.00	24.00	.66	.777
	III	3.00	1.32	1.00	5.00	.66	.778
	IV	6.50	3.12	3.00	11.00	.71	.693
deep structure (propositional matching)	I	14.00	7.09	1.00	25.00	.48	.974
	II	15.80	12.84	3.00	40.00	.64	.811
	III	5.11	1.62	3.00	8.00	.54	.932
	IV	10.83	4.78	4.00	18.00	.78	.579

Note: *DOMAIN*: I = rainbow experiment (*N=12*), II = crack experiment (*N=10*), III = electrical effect experiment (*N=9*); IV = disease situation (*N=12*);*KS-Z* = Kolmogorov-Smirnov one-sample test; * $p < .05$; ** $p < .01$

Besides the descriptive measures (see Table 4.4 and 4.5), SMD compares the individual representations with an expert representation (see Table 4.6 and 4.7).

TABLE 4.6
SMD similarity measures (structure) between participant and expert solutions

	DOMAIN	M	SD	Min	Max	KS-Z	p
surface structure	I	.682	.260	.06	1.00	.550	.923
	II	.546	.244	.21	.93	.758	.614
	III	.427	.109	.23	.62	.711	.692
	IV	.388	.199	.16	.84	.594	.872
matching structure	I	.729	.239	.25	1.00	.706	.701
	II	.711	.213	.40	1.00	.510	.958
	III	.844	.155	.60	1.00	.860	.450
	IV	.654	.166	.43	.86	.670	.760
density of vertices	I	.778	.160	.41	.93	.797	.548
	II	.687	.204	.36	.99	.698	.714
	III	.622	.209	.16	.79	.708	.699
	IV	.715	.214	.36	1.00	.551	.922
structural matching	I	.564	.142	.29	.86	.556	.917
	II	.731	.143	.50	1.00	.547	.926
	III	.871	.113	.67	1.00	.645	.799
	IV	.592	.099	.40	.80	1.039	.230

Note: *DOMAIN*: I = rainbow experiment, II = crack experiment, III = electrical effect experiment; IV = disease situation; *KS-Z* = Kolmogorov-Smirnov one-sample test; * $p < .05$; ** $p < .01$

The comparisons are described with the help of the Tversky similarity (0 = no similarity; 1 = total similarity). Our analysis revealed interindividual differences in the three physics domains for the measure *propositional matching*. For all other measures, we found no interindividual differences between our subjects (see Table 4.6 and 4.7). Regarding the differences between the subject domains, the Kruskal-Wallis H-Test revealed significant differences between the measures *surface structure*, χ^2 (3, N = 43) = 10.26, p > .05, *structural matching*, χ^2 (3, N = 43) = 20.53, p > .001, and *vertex matching*, χ^2 (3, N = 43) = 19.37, p > .001.

TABLE 4.7
SMD similarity measures (semantics) between participant and expert solutions

	DOMAIN	M	SD	Min	Max	KS-Z	p
vertex matching	I	.096	.076	.00	.27	.781	.575
	II	.104	.077	.00	.27	.837	.486
	III	.243	.080	.17	.42	.570	.901
	IV	.159	.050	.05	.23	.629	.824
deep structure (propositional matching)	I	.010	.024	.00	.07	1.720	.005*
	II	.011	.035	.00	.11	1.657	.008*
	III	.024	.049	.00	.12	1.409	.038*
	IV	.035	.042	.00	.11	1.029	.240

Note: *DOMAIN*: I = rainbow experiment, II = crack experiment, III = electrical effect experiment; IV = disease situation; *KS-Z* = Kolmogorov-Smirnov one-sample test; * $p < .05$; ** $p < .01$

In addition to the above reported quantitative measures, SMD enables us to automatically create cutaway and discrepancy re-representations for qualitative analysis. These standardized re-representations could be used for an in-depth analysis of the individual re-representations (see Figure 4.3).

The quite elaborated cutaway re-representation in Figure 4.7 includes all vertices and edges of the subject. Compared to the reference re-representation (expert solution of the crack experiment question) seven vertices are semantically correct (vertices as circles). However, there are also seven vertices which are incorrect compared to the expert solution. Additionally, the cutaway re-representation reveals that the student's understanding of the phenomenon in question is not fully connected (2 submodels). Furthermore, the re-representation includes three circles. However, these circles include incorrect vertices (e.g. farben-rot-regenbogen-grün-farben).

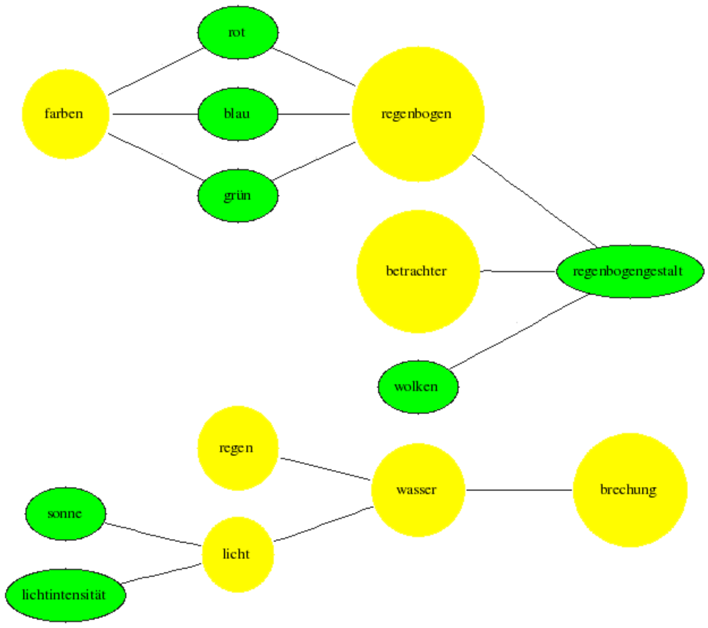

FIGURE 4.7. *SMD cutaway re-representation, domain II (crack experiment)*

Pedagogical implications

The primary purpose of this initial study was to compare the methodological range of QFCA and SMD. However, we briefly discuss the results from an educational point of view. Results from both analysis approaches show that the structural and semantic measures highlight important changes of the assessed knowledge representations. The *structural measures* of QFCA (e.g., count of concepts) and SMD (e.g., surface structure) show remarkable differences between the four subject domains. For the electrical effect experiment, we found significant less concepts in the subjects' representations. The *semantic measures* (QFCA: correct model concepts; SMD: vertex matching, deep structure) show that the learners are far from using correct concepts compared to experts. Hence, the subjects of this initial study are still in their initial stage of the learning process. An instructional intervention would now focus on missing concepts or misconceptions found in the individual re-representation (e.g., Figure 4.7) and/or structural conspicuities (e.g., many submodels).

Comparison of QFCA and SMD analysis approaches

Using the same set of data, we were able to conduct an in-depth investigation of both analysis approaches. Minor differences in the results are caused by the transformation of the participant's data into a raw data file. Hence, further studies should also focus on various assessment techniques and available interfaces to analysis approaches to identify their strength and weaknesses as well. Although both analysis methods work quite well and produce a lot of indicators, there are several difficulties and differences to report.

The first point concerns the placement (classification) of the indicators in relation to the mental model results. This is essential not only to compare the empirical results of different indicators but also to compare results of different studies. A precondition for this point is to find arithmetic similarities between the analysis indicators (see Table 4.8). Although the quantitative measures should be equal, the values differ. After intensive checking we found that the export function of the assessment technique was not accurately exporting the raw data. Therefore, the quantitative measures differ minimally. The QFCA method uses the assessed data directly; for SMD we used the imprecise exported data.

TABLE 4.8
Comparison of indicators, scientific quality, and exploratory power of both analysis approaches

	QFCA	SMD
Quantitative measures	count of concepts & propositions ruggedness	structural measures semantic measures various graph theory measures (e.g., ruggedness, cyclic)
Qualitative measures	relative objective plausibility absolute objective plausibility correct model concepts	standardized re-representations cutaway- and discrepancy re-representations
Objectivity	semi-automated analysis raw data based algorithms	automated analysis of predefined raw data structure
Reliability	partly tested (see Al-Diban, 2002)	tested (see Ifenthaler, 2010c)
Validity	not tested	tested (see Ifenthaler, 2010c)
Areas of application	limited comparisons single case analysis small group analysis	unlimited comparisons single case analysis large group analysis stochastic analysis
Advantages and limitations	semi-automated analysis structural decomposition into 5 formal categories recomposition into 3 content-based criteria	automated analysis structural decomposition into 3 key categories recomposition into “re-representations”

Second, the scientific quality criteria objectivity, reliability, and validity should be checked and reported. The analysis step of qualitative restructuring of data in QFCA to find generic concepts and propositions is not wholly objective and characterized by degrees of freedom.

A third point is concerned with the areas of application for research and practice. These areas are limited in QFCA and almost unlimited in SMD. This great advantage of SMD is bought at the price of limitations in precision and the pedagogical information value of the highly aggregated criteria. Due to its automated analysis, SMD is especially at an advantage for applications in pedagogical practice, where results are needed as quickly as possible. The QFCA results were analyzed with the help of coders, which is time consuming.

Conclusions and future developments

Basic questions of a reliable and valid diagnosis of mental models are not solved completely (see Ifenthaler, 2008). This article focuses on the quality of two analysis approaches, a matter in which there is a major lack of systematic research, and in which one seldom finds scientific criteria like objectivity, reliability, and validity (T. E. Johnson, et al., 2006). Actually, there is a lack of stochastic modelling concerning the analysis methods of the mental models approach, especially for content-based data.

Future research with bigger samples should focus on (a) the comparison of available assessment and analysis approaches, and (b) on the observation of processes of learning-dependent change (e.g., Ifenthaler, et al., in press). In this way, different types of subjective mental models could be identified and classified. When more is known about the modes by which mental model representations change, it will become possible to increase the individual specificity and efficiency of instructional designs (see Ifenthaler, 2008). Both described analysis approaches, QFCA and SMD, are applicable to different knowledge domains. Disadvantages of QFCA might be its capacity for no more than about small groups, or its inability to analyze complex knowledge representation contents. Hence, the approach is labor intensive and there is a need for further service interfaces. In contrast, SMD proved to be highly economical due to its automated process. The integration of the SMD analysis features into a new web-based research platform, HIMATT (Highly Integrated Model Assessment Technology and Tools) with graphical and text-based assessment and analysis techniques is a consequent and forward-looking approach (see Pirnay-Dummer, et al., 2010). A further development of HIMATT could also include the QFCA approach. These future developments will open up new opportunities for continuing research on mental models and lead to new instructional implications.

5

HIGHLY INTEGRATED MODEL ASSESSMENT TECHNOLOGY AND TOOLS 🕮

There has been little progress in the area of practical measurement and assessment, due in part to the lack of automated tools that are appropriate for assessing the acquisition and development of complex cognitive skills and structures. In the last two years, an international team of researchers has developed and validated an integrated set of assessment tools called HIMATT (Highly Integrated Model Assessment Technology and Tools) which addresses this deficiency. HIMATT is Web-based and has been shown to scale up for practical use in educational and workplace settings, unlike many of the research tools developed solely to study basic issues in human learning and performance. In this chapter, the functions of HIMATT are described and several applications for its use are demonstrated. Additionally, two studies on the quality and usability of HIMATT are presented. The chapter concludes with research suggestions for the use of HIMATT and for its further development.

🕮 This chapter is based on: 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. doi: 10.1007/s11423-009-9119-8

Introduction

Knowledge is at the center of all cognition. Knowledge is constructed by internal representation processes (e.g., mental models, schemata). Knowledge is activated and deployed through the use of external re-representation processes (e.g., concept maps, diagrams, verbal discourse). This means that models used for representation and re-representation are critical in nearly all decision making and problem solving activities. Moreover, representation and re-presentation processes are critical for learning and instruction. However, how models can be developed and deployed effectively and efficiently to support learning, performance, and instruction is not well understood. One impediment to progress has been the lack of appropriate assessment tools that establish meaningful inferential links between external re-presentations and internal representations.

Previously, tools to support research into mental model development and the acquisition of skilled performance required a great deal of time and effort on the part of highly trained researchers (e.g., think-aloud protocol analysis). As a result, such assessment tools have been limited to basic research and have not had an impact on practical issues such as the design of effective instructional systems and learning environments. The desire to have practical assessment tools that are useful for improving learning, performance, and instruction has motivated significant developments in the last several years (Ifenthaler, 2008). Techniques such as the *structure formation technique* (Bonato, 1990; Scheele & Groeben, 1984), *concept mapping* (Cañas, et al., 2004; Novak, 1998; Nückles, Gurlitt, Pabst, & Renkl, 2004; Spector, 2006; Spector, Dennen, & Koszalka, 2006), or the *test for causal diagrams* (Al-Diban, 2008) use graphical representations for assessment purposes. For language-oriented assessment, the *thinking aloud protocol* (Ericsson & Simon, 1993, 1998) and *MITOCAR* (Pirnay-Dummer & Ifenthaler, 2010) have been developed for quantifying verbal data. Other assessment tools have been automated, such as *Pathfinder* (Schvaneveldt, 1990), but only a few of these tools are fully automated, including automation of both the elicitation and the analysis processes involved in assessing learning and performance. One tool that is fully automated is the *SMD Technology*: Structure, Matching, Deep Structure (Ifenthaler, 2010c), which is included together with several compatible tools in HIMATT (Highly Integrated Model Assessment Technology and Tools). The HIMATT tools have been developed, implemented, studied, and systematically validated within numerous international research collaboration studies (T. E. Johnson, et al., 2006). After the cross-validation of the different individual tools that are now integrated in HIMATT, the researchers involved noted that these various tools were based on different but compatible methodologies; furthermore, they were implemented differently on diverse platforms. However, the underlying approach was quite similar and the notion of using external representations to determine how well internal representations were being developed ran through all these tools. The idea was then born to create a comprehensive toolset which combines and further automates these state-of-the-art model-based assessment methodologies.

Automation is particularly important when we think about applications in the field. As long as the tools are not automated and accessible to practitioners (e.g., teachers, instructional designers, trainers), they will only be used in prototype and research settings but not in the real-world applications. Feasible instruments that can help track the development of knowledge and skill of many individuals without excessive cost and effort are especially important when we apply the methodologies to time series experiments to systematically track changes over time (Ifenthaler, 2008; Ifenthaler & Seel, 2005) or if we use them to show effects within a series of interventions (Ifenthaler, 2010d; T. E. Johnson, et al., 2009).

HIMATT (Highly Integrated Model Assessment Technology and Tools) is a new combined toolset which accounts for all of these constraints. It was developed to convey the benefits of each methodological approach into one environment which can be used by researchers with only little prior training. It is implemented to run on the Web, thus presenting all content on a standard Web browser to both the researchers and the subjects.

Theoretical foundation

Every implemented technology in HIMATT has its own theoretical background. This was one of the most important criteria in the decision as to which methodology should be used for HIMATT.

DEEP (Dynamic Enhanced Evaluation of Problem Solving) was developed specifically to assess progress of learning towards expert-like performance in domains involving complex and ill-structured problems, such as engineering design, environmental planning, and medical diagnosis (Spector & Koszalka, 2004). DEEP was inspired by causal influence diagrams – a knowledge elicitation technique used by system dynamicists when developing simulation models for complex systems. In DEEP, a variation of causal influence diagrams called annotated causal concept maps is used to elicit a conceptualization of how an individual (or small group of persons) is thinking about a problem situation. The method involves identifying representative problems and then presenting them to respondents who are first asked to identify and describe the five or ten key factors influencing the problem situation. Problem respondents are then asked to identify and describe the relationships that exist among these key factors. These external representations can be compared with those of experts in a number of ways to see if learners are improving their representations over time and through instruction and beginning to think more like domain experts. DEEP only automated the process of eliciting the representation; in its first incarnation it did not automate the analysis, although the analytical methods used by Spector and Koszalka (2004) are completely compatible with those of the next two tools we describe (one of the motivations for integrating these tools).

MITOCAR (Model Inspection Trace of Concepts and Relations) and T-MITOCAR (Text-MITOCAR) have a background in mental model theory (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Seel, 1991), association psychology (Davis, 1990; Lewin, 1922; McCoon & Ratcliff, 1992; McNamara, 1992, 1994; Stachowiak, 1979), and linguistics (Frazier, 1999; Pollio, 1966; Russel & Jenkins, 1954). Both MITOCAR and T-MITOCAR rely on the dependence of syntax and semantics within natural language and use the associative features of text as a methodological heuristic to represent knowledge from text sources. Unlike tools from Web ontologies and the semantic Web (Ding, 2001), MITOCAR and T-MITOCAR can work on a comparably small amount of text (350 words +).

SMD and MITOCAR both combine analysis and comparison functions based on graph theory (Bollobàs, 1998; Tutte, 2001), set theory (Jech, 2007), model theory (Rothmaler, 2000), and similarity distribution measures (Kruskal, 1964; Tversky, 1977). SMD also contains foundations for the measurement of change (e.g., Collins & Sayer, 2001; Harris, 1963; Ifenthaler, 2008; Ifenthaler & Seel, 2005).

Methodologically, the tools integrated into HIMATT touch the boundaries of qualitative and quantitative research methods and provide bridges between them. On the one hand, text can be analyzed very quickly without loosening the associative strength of natural language (MITOCAR and T-MITOCAR). Furthermore, conceptual graphs can be annotated by experts (DEEP). All of the data, regardless of how it is assessed, can be analyzed quantitatively with the same comparison functions for all built-in tools without further manual effort or recoding. Additionally, HIMATT generates standardized images of text and graphical representations.

HIMATT architecture

The HIMATT architecture consists of two major platforms: a) *HIMATT Research Engine* and b) *HIMATT Subject Environment*. Functions for conducting and analyzing experiments are implemented within the HIMATT Research Engine. These functions include 1) Experiment Management, 2) Researcher Management, 3) Subjects Management, 4) View Function, and 5) Analysis and Compare Function. The HIMATT Subject Environment dynamically provides assigned experiments to individual subjects.

HIMATT has been implemented and runs on a Web server using Apache, MySQL (**MY** **S**equential **Q**uery **L**anguage), and PERL (**P**ractical **E**xtraction and **R**eport **L**anguage), plus additional packages such as GraphViz (Ellson, et al., 2003).

Experiment management

The core unit in HIMATT is the experiment, which can be laid out flexibly by the researcher. Experiments in HIMATT consist of three assessment modules: (1) DEEP, (2) T-MITOCAR, and (3) MITOCAR as well as an INSTRUCTION module which is used to give the subject instructions and explanations (see Figure 5.1). The instructions are texts which may contain HTML code (e.g., to link pictures, videos, or other objects). Thus, they may also be used to present simple interventions to the subjects between the assessments, although this option is not very well developed.

Besides mandatory labels and names for experiments, the researcher can add meta-information about them. This helps to identify the purpose of the experiment and quickly select from a large number of experiments with the help of a search function. Figure 5.1 shows an experiment in which three modules have been laid out. The sequence of this sample experiment is as follows: 1) introduction to the subject, where the purpose of the experiment and additional information is presented; 2) the T-MITOCAR module, where the subject is asked to write a statement of at least 350 words; 3) an “outro,” where the subject gets further information after completing the experiment. The number and sequence of modules and the content of all subject information can be changed any time. Once an experiment is laid out completely, subjects may be assigned to the experiments with the subject management function.

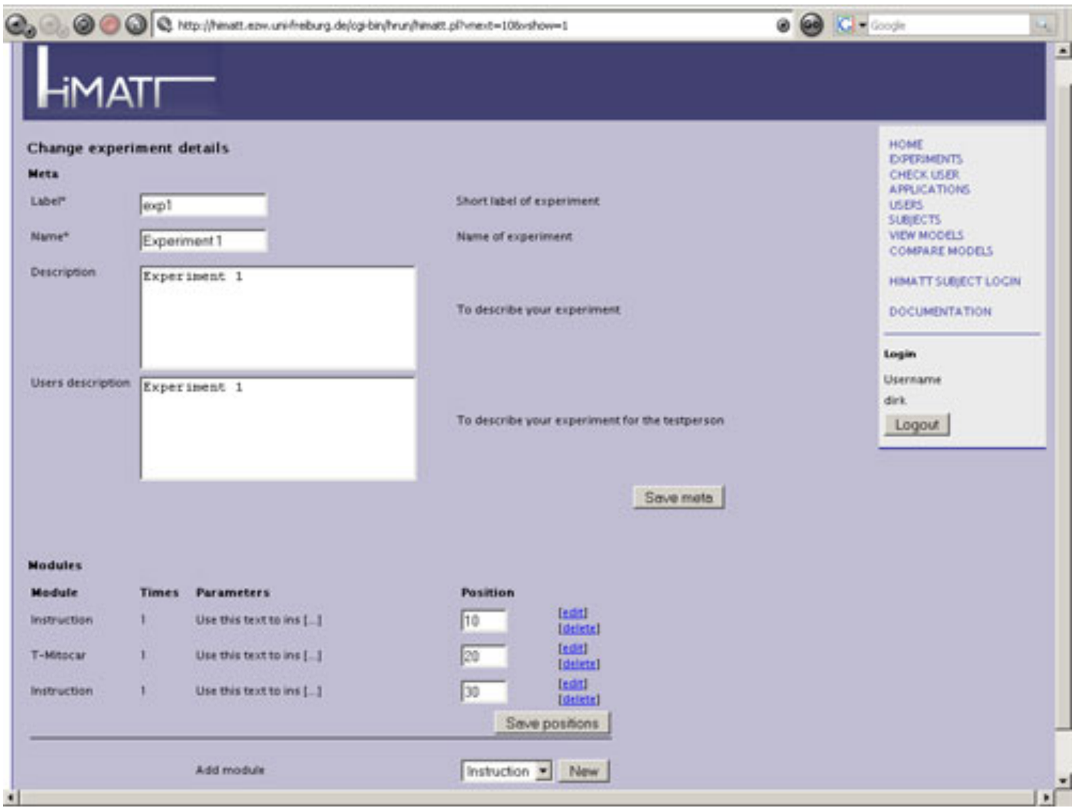

FIGURE 5.1. *HIMATT Experiment Management*

Subject management

The subject management function includes multiple options. First, a researcher can add subjects to the HIMATT database. Subject information includes at least a username and a password. If a researcher wants to add a large number of subjects, HIMATT can automatically generate a specified number of subjects with individual usernames and passwords. Additionally, the user can include a prefix to all usernames or passwords in order to more easily identify them later on during experimentation and analysis procedures.

Another important option within the subject management is the assignment of subjects to experiments. Once an experiment has been laid out completely and subjects have been added to the database, researchers can assign subjects to experiments. HIMATT also contains an export function which enables the researcher to export all assigned subjects from an experiment onto a spreadsheet. Naturally, all subject information can be deleted and changed whenever the researcher wishes.

Researcher management

As scientific experiments are very rarely conducted only by a single researcher, HIMATT supports research teams with members assigned to these roles: a) HIMATT Administrator, b) HIMATT Researcher, and c) HIMATT Research Assistant. Each role comes with permission to use different functions (see Table 5.1).

TABLE 5.1
HIMATT roles (X indicates permission)

Role	Sponsor Researchers	Experiment Management	Subjects Management	View Function	Analysis and Compare Function
HIMATT Administrator	X	X	X	X	X
HIMATT Researcher		X	X	X	X
HIMATT Research Assistant			X	X	X

Only the HIMATT Administrator can sponsor other researchers and give them access to the HIMATT Research Engine and HIMATT Subject Environment. So far, the three HIMATT Administrators are the authors of this article. A sponsored HIMATT Researcher has permission to create new experiments, add subjects, and view and analyze the results of the experiments.

View function

The view function presents the knowledge graph as a picture to the researcher. This function allows the researcher to choose from specific experiments and knowledge graphs, which are then available as *PNG* (Portable Network Graphics) images for download.

Depending on the underlying module (DEEP, T-MITOCAR, or MITOCAR) the graphs will have different features: annotations for DEEP concept maps, associative strengths at the links for T-MITOCAR, and pairwise rated strengths for MITOCAR.

Essentially, the standardized re-representation is done in the same way for all three modules using the pairwise stored information from the database and GraphViz (Ellson, et al., 2003).

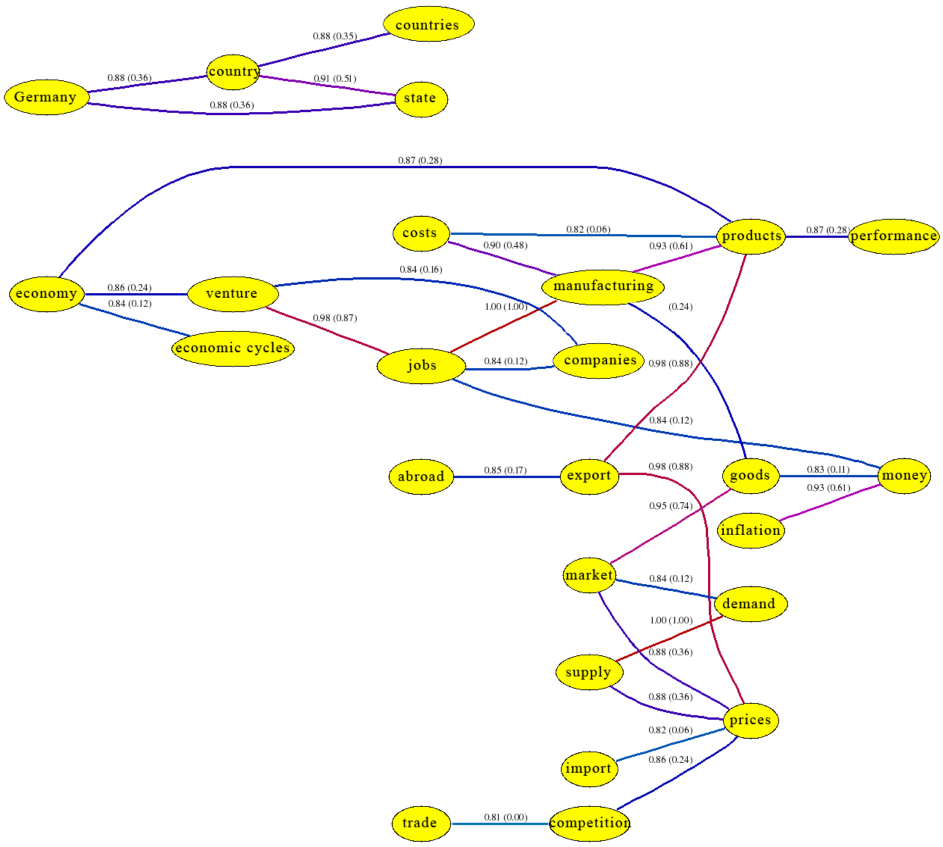

***FIGURE 5.2**. HIMATT sample graph*

Analysis and compare function

The analysis function mainly calculates descriptive measures for the stored knowledge representations. These descriptive measures include various structural indicators derived from graph theory (Harary, 1974; Hietaniemi, 2008).

- Connectedness (SMD). Computed as the possibility to reach every node from every other node in the knowledge representation (Ifenthaler, et al., in press).
- Ruggedness (SMD). Computed as the sum of subgraphs which are independent or not linked (Ifenthaler, et al., in press).
- Average degree of vertices (SMD). Computed as the average degree of all incoming and outgoing edges of the knowledge representation (Ifenthaler, et al., in press).
- Number of Cycles (SMD). Computed as the sum of all cycles (a path returning back to the start node of the starting link) within a knowledge representation (Ifenthaler, et al., in press).
- Vertices, Nodes (SMD). Computed as the sum of all nodes within a knowledge representation (Ifenthaler, et al., in press).

- Edges, Links (SMD). Computed as the sum of all links within a knowledge representation (Ifenthaler, et al., in press).

The measures for comparison can be applied to any undirected graph, not only to re-representations from MITOCAR and T-MITOCAR. There are six core measures for the comparison of conceptual graphs from the SMD Technology (Ifenthaler, 2006, 2010c) and from MITOCAR (Pirnay-Dummer, 2006). Some of the measures count specific features of a given graph. For a given pair of frequencies f_1 *and* f_2, the similarity is generally derived by this function:

$$s = 1 - \frac{|f_1 - f_2|}{\max(f_1, f_2)}$$

Which results in a measure of $0 \leq s \leq 1$, where $s=0$ is complete exclusion and $s=1$ is identity. The other measures collect sets of properties from the graph (e.g., the vertices = concepts or the edges = relations). In this case, the Tversky similarity (Tversky, 1977) applies for the given sets A and B:

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

α and β are weights for the difference quantities which separate A and B. They are usually equal ($\alpha = \beta = 0.5$) when the sources of data are equal. However, they can be used to balance different sources systematically (e.g., comparing a learner model which was constructed within five minutes to an expert model, which may be an illustration of the result of a conference or of a whole book).

- Surface (SMD). The surface measure (Ifenthaler, 2006, 2010c) compares the number of vertices within two graphs. It is a simple and easy way to calculate values for surface complexity.
- Graphical Matching (SMD). The graphical matching (Ifenthaler, 2006, 2010c) compares the diameters of the spanning trees of the graphs, which is an indicator for the range of conceptual knowledge. It corresponds with structural matching as it is also a measure for complexity only.
- Concept Matching (MITOCAR). Concept matching (Pirnay-Dummer & Ifenthaler, 2010) compares the sets of concepts (vertices) within a graph to determine the use of terms. This measure is especially important for different groups which operate in the same domain (e.g., using the same textbook). It determines differences in language use between the models.

- Density of Vertices (MITOCAR). The density of vertices (Pirnay-Dummer & Ifenthaler, 2010) describes the quotient of terms per vertex within a graph. Since both graphs which connect every term with each other term (everything with everything) and graphs which only connect pairs of terms can be considered weak models, a medium density is expected for most good working models.
- Structural Matching (MITOCAR). The structural matching (Pirnay-Dummer & Ifenthaler, 2010) compares the complete structures of two graphs without regard to their content. This measure is necessary for all hypotheses which make assumptions about general features of structure (e.g., assumptions which state that expert knowledge is structured differently from novice knowledge).
- Propositional Matching (SMD). The propositional matching (Ifenthaler, 2006, 2010c) value compares only fully identical propositions between two graphs. It is a good measure for quantifying semantic similarity between two graphs.
- Balanced Semantic Matching. The balanced semantic matching uses both concepts and propositions to match the semantic potential between two model representations.

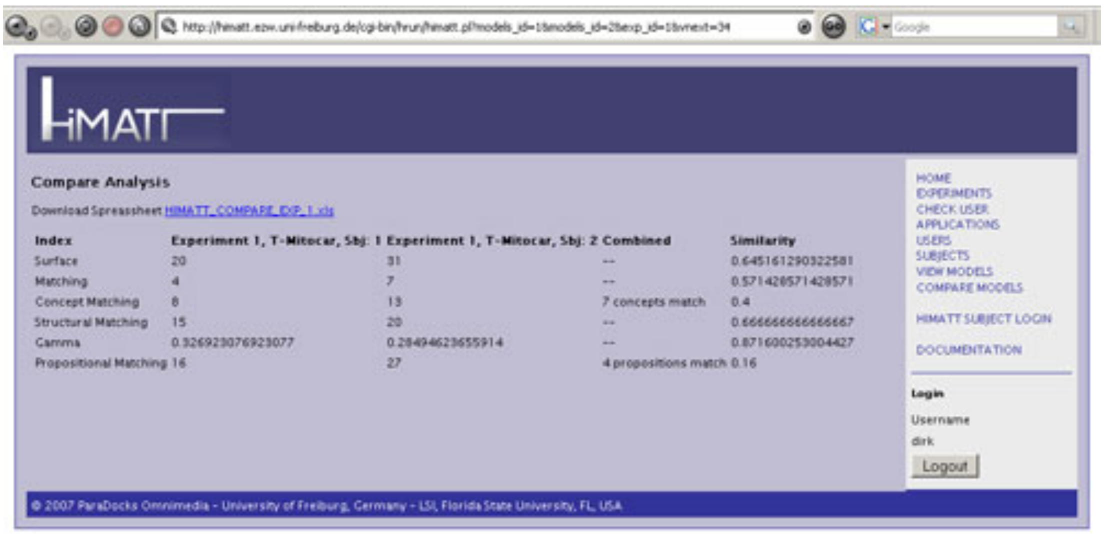

FIGURE 5.3. *Compare function including all six HIMATT core measures*

The measures are calculated automatically within seconds and are then displayed as pairwise sets including the six core measures described above (see Figure 5.3).

Additionally, the researcher can download a spreadsheet containing all measures for further statistical analysis.

Subject environment

Subjects login to another part of the software – the HIMATT Subject Environment. If they are only assigned to one experiment, they will be led directly to that experiment. If they are assigned to more than one experiment, they choose from a list of assigned experiments. In the experiment all instructions and modules are presented as laid out by the researcher in the HIMATT Experiment Management function. A final screen with a thank you statement marks the end of an experiment for the subject. Re-Login is of course possible if further experiments are available for the subject.

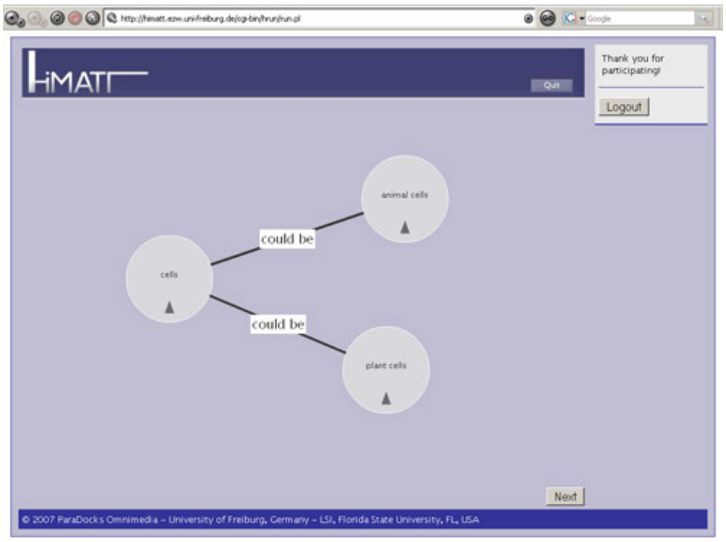

FIGURE 5.4. *Subject environment with DEEP module*

Figure 5.4 illustrates the HIMATT Subject Environment, where the subjects create a concept map within the DEEP module. Within this module, the subjects can add nodes and links to the concept map and annotate them with additional information.

HIMATT test quality

Objectivity

As with all reactive instruments, all assessment parts of HIMATT measure behavior previously induced by an intervention, such as instructions to help the subject create a concept map or write a text on a given topic. In HIMATT all parts of an experiment are standardized for all subjects. The same holds true for all parts of the analysis and comparison. Therefore, HIMATT is completely objective as long as experiments are carried out in the designed way.

Reliability

HIMATT supports an approximate representation of semantic and symbolic cognitive structures, such as schemata and mental models. Reliability will always depend on the theoretical construct under investigation. For schemata high reliability measures should always be expected because the construct is meant to be stable. With mental models, it is a different story. Mental models are on-the-fly constructions used to explain unexpected or complex phenomena in the outside world; they are believed to be discarded by the system after usage and may be involved in the construction of a schema if applied frequently and successfully.

However, promising reliability indices exist for most of the instruments integrated into HIMATT. For the SMD indices the reliability is reported as $r = .82$ for surface structure, $r = .82$ for graphical matching, and $r = .90$ for propositional matching (Ifenthaler, 2006). For MITOCAR indices the retest reliability is reported to be between $r = .94$ (strength of connectedness measures) and $r = .79$ (contrast measures) for the proximity vector leading to the output graph (Pirnay-Dummer, 2006). As already mentioned for objectivity, the measurements used to construct the graph from a text are not dependent on any interpretation. Therefore, reliability comes down to the question as to whether one is able to write the same text twice in response to the same task. From an experimental point of view, it is as easy to test this as it is to test classic items. Finding the right trade-off between memory effects, expressivity of language, and uncertainty of outputs which rely on the same constructs (e.g., for mental models) is not an easy task and should be handled with outmost methodological care. Critics would certainly address the memory problem with natural language and issues with learning during assessment while supporters would argue in the direction of expressivity and the problem of construct shifts if the reasoning processes are too far away from one other.

Validity

The comparison indices built into HIMATT using the SMD-MITOCAR methodologies address either the structure or the semantics of an assessed construct. They can be equally applied to natural language analysis and concept mapping. All of the indices make measurements of the graphs. Convergence is expected to be different between the structural and the semantic indices. The correlation matrix shows the convergent validity within each area and the divergent validity between them. Validity was tested on N = 1,849,926 individual pairwise model comparisons. Each pair of models belongs to the same subject domain.

TABLE 5.2
HIMATT validity measures

	BSMatch	CMatch	PMatch	Surface	GMatch	SMatch	Gamma
BSMatch	1.00						
CMatch	0.71	1.00					
PMatch	0.91	0.68	1.00				
Surface	0.20	0.26	0.18	1.00			
GMatch	0.17	0.21	0.16	0.79	1.00		
SMatch	-0.24	0.36	0.53	0.63	0.48	1.00	
Gamma	0.18	0.24	0.15	0.37	0.38	0.08	1.00

Balanced semantic matching (BSMatch), concept matching (CMatch), and propositional matching (PMatch) are the semantic indices of HIMATT. Surface matching (Surface), graphical matching (GMatch), structural matching (SMatch), and gamma are structural indices. All convergent validity measures are reported in *italics*; the others are divergent validity measures (see Table 5.2). High validity measures can be reported throughout all of the semantic indices. The three structural indices aiming at the complexity (Surface, GMatch) or the full structure (SMatch) of the models are also aligned quite well. Gamma, however, is different. It accounts for the density of the model rather than for its complexity, which may be a reason why it does not correlate very well with the other structural indices. This may be a hint that gamma should be treated differently in the future. The surprisingly high correlation between propositional matching and structural matching is another interesting point to discuss and investigate further. At the moment we do not have a complete theoretical explanation for this effect throughout all of the models and investigated domains; but since both are more complex indices for addressing either structure or semantics, this may point to an interconnectedness between structure and semantics which might not be visible on a more cursory level of comparison (Jackendoff, 1983).

HIMATT usability

We applied a usability test which included 26 items (see Appendix A, Table 5.4, for a translation of the items) which had to be answered on a Likert scale ranging from 1 (highly disagree) to 5 (highly agree). Seventy-four students (66 female and 8 male) from the University of Freiburg, Germany, participated in the usability study. Their average age was 21.9 years (*SD* = 2.3).

First, an explorative factorial analysis (varimax rotation) was carried out by means of selected variables (see Appendix A, Table 5.4). The eight extracted factors represent 72.8 % of the variance. The first factor is determined by six items (Nr. 4, 14, 15, 17, 18, 21). Consequently, the first factor represents colors and screen design (Cronbach's α = .843). The second factor is determined by five items (Nr. 3, 19, 20, 23, 24) and represents the coherence of the HIMATT software (Cronbach's α = .794). Factor three represents the learnability of HIMATT functions (Cronbach's α = .725) and is determined by four items (Nr. 1, 2, 6, 8). The fourth factor is determined by four items (Nr. 7, 9, 10, 22). They represent the reliability and handling of HIMATT (Cronbach's α = .733). The fifth factor is determined by three items (Nr. 5, 11, 12) and represents the complexity of HIMATT functions (Cronbach's α = .594). Factor six represents the character set of HIMATT (Cronbach's α = .687), determined by two items (Nr. 25, 26). The seventh factor is determined by one item (Nr. 16) and represents use of colors for instructions. The eighth and last factor is also determined by one item (Nr. 13). It represents directions at the start of HIMATT.

Secondly, the eight factors were used to investigate the usability of HIMATT. Table 5.3 shows the descriptive statistics of the eight factors.

TABLE 5.3
Usability test results

Factor Nr.	M	SD	Min	Max
I	3.42	.64	1	5
II	4.16	.45	3	5
III	4.31	.48	3	5
IV	3.86	.51	2	5
V	4.23	.39	3	5
VI	3.99	.56	2	5
VII	3.51	.57	1	5
VIII	4.15	.66	2	5

The results of our usability test show that HIMATT and its features are widely accepted among the users. Particularly well accepted is the easy learnability of HIMATT functions (factor 3). This is also expressed by the high acceptance of factors five (complexity of HIMATT functions) and two (coherence of HIMATT). The usability test also revealed a high level of acceptance of the instructions at the start of HIMATT (factor 8).

HIMATT applications

Basically, with HIMATT it is possible to investigate anything which addresses states and changes, analysis and comparison within the methodological boundaries of concept mapping, and the annotation of association networks on the basis of different kinds of text sources. Both groups and individuals can be assessed within classical experimental settings and field applications, for example, in learning and instruction or schooling and education. So far, individual tools from HIMATT have been used successfully in navigation tracking (Dummer & Ifenthaler, 2005), measurement of learning-dependent progression (Ifenthaler, et al., in press; Ifenthaler & Seel, 2005), cognitive learning strategies and intellectual abilities (Ifenthaler, et al., 2007), research on the quantitative comparison of expertise, reading comprehension (Pirnay-Dummer & Ifenthaler, in press), needs assessment, ontology oriented data mining, and organizational knowledge management. The comprehensive toolset will enable researchers to continue working on all of these research interests. It will also be possible to address additional fields due to the combination of the assessment and analysis tools. Not only will this make things easier and more integrated but also faster since the data will not have to be transferred from one tool to another anymore.

Future development and directions

While the current version of HIMATT represents a state-of-the-art assessment tool suite. HIMATT features such as arrows that reflect relative weights through thick and thin lines, nested diagrams that allow layers of a complex problem to be developed, elicited, and explored could be added. A significant direction for future development would be to take HIMATT and other sophisticated assessment tools and transform them into teaching tools. Since the earliest development of DEEP, users have commented that such assessment tools would make excellent teaching tools as well. Progress in the design of instruction for complex tasks requires tools such as HIMATT. Progress in developing personalized learning systems requires an extended version of HIMATT and other tools that can support formative feedback and self-regulatory behaviors. Just as science is cumulative, the tools used by scientists are cumulative. In this case, perhaps HIMATT represents a contribution to the development of cumulative knowledge and tools for both scientists (i.e., educational researchers) as well as for practitioners (i.e., teachers and instructional designers).

Appendix A

TABLE 5.4
Original items of the usability questionnaire and corresponding translations

Item Nr.	Factor Nr.	Item loading	Original item	Item translation
1	III	.795	Die Bedienung der Software ist leicht erlernbar.	It is easy to learn how to work with the software.
2	III	.449	Ohne Unterstützung sind alle Funktionen zu bedienen.	All functions can be used without support.
3	II	.611	Die Navigation innerhalb der Software ist mir leicht gefallen.	I found it easy to navigate through the software.
4	I	.512	Optisch ist die Software ansprechend gestaltet.	The design of the software is optically appealing.
5	V	.529	Alle Buchstaben und Sonderzeichen erscheinen in üblicher Form auf dem Bildschirm.	All letters and special characters appear as they should on the screen.
6	III	.403	Die Mausbedienung ist einfach.	It is easy to use the mouse with the software.
7	IV	.645	Die Tastaturbedienung ist einfach, z.B. bei der Steuerung des Cursors.	It is easy to use the keyboard, e.g., to move the cursor.
8	III	.842	Tippfehler können vor Ausführen einer Eingabe korrigiert werden.	Typos can be corrected before making an entry.
9	IV	.848	Die Software reagiert robust und informierend auf Bedienungsfehler.	The software provides reliable and informative support in the case of operating errors.
10	IV	.459	Die Software arbeitet fehlerfrei, zuverlässig und kontrollierbar, auch bei falschen Befehls- oder Antworteingaben.	The software is error-free, reliable, and controllable, even when incorrect commands or answers are entered.
11	V	.556	Der Befehlsumfang für die Benutzung ist einfach.	It is easy to learn the commands necessary to operate the software.
12	V	.805	Befehle, Begriffe und Symbole für gleiche Sachverhalte und Bedienungsfunktionen werden einheitlich verwendet.	Commands, terms, and symbols for the same item or operating function are uniform.
13	VIII	.729	Die Benutzungshinweise, die am Anfang gegeben werden, sind klar und verständlich.	The instructions provided at the beginning are clear and understandable.
14	I	.820	Die Qualität der Farben ist gut, z.B. durch klare Kontraste.	The quality of the colors is good, e.g., clear contrast.
15	I	.671	Durch farbliche Hinweise wird die Bedienung der Software erleichtert und erklärt.	The color codes serve to simplify and explain the operation of the software.
16	VII	.810	Die Farben zur Verdeutlichung der Bedienung werden einheitlich eingesetzt.	The colors used to simplify the operation of the software are applied uniformly.
17	I	.616	Die Farbgestaltung trägt sinnvoll zur Erleichterung und Erklärung der Bedienung der Software bei.	The colors are a useful aid for explaining how to operate the software.
18	I	.914	Insgesamt sind die Farben effektiv, sinnvoll und motivierend eingesetzt.	In general, the use of color is effective, sensible, and motivating.

TABLE 5.4 continued
Original items of the usability questionnaire and corresponding translations

Item Nr.	Factor Nr.	Item load-ing	Original item	Item translation
19	II	.793	Der Bildschirmaufbau ist übersichtlich und verständlich.	The screen layout is clear and comprehensible.
20	II	.776	Die Textgestaltung ist sinnvoll, übersichtlich und gut lesbar.	The text layout is sensible, clear, and easy to read.
21	I	.844	Die Farben sind effektiv, sinnvoll und motivierend eingesetzt.	The use of color is effective, sensible, and motivating.
22	IV	.731	Die Anpassungsmöglichkeiten der Software sind umfangreich.	There are many options for customizing the software.
23	II	.732	Die Navigation der Software ist benutzerfreundlich.	The navigation of the software is user-friendly.
24	II	.444	Die Qualität der Grafiken ist gut, d. h. klare Linien, Formen, Kontraste und verständliche Darstellungen.	The quality of the graphics is good, i.e. they have clear lines, forms, and contrast and are well designed.
25	VI	.641	Insgesamt ist die Textgestaltung sinnvoll, übersichtlich und gut lesbar.	In general, the text layout is well designed and organized and is easy to read.
26	VI	.865	Der Zeichensatz ist in seiner Form und Größe geeignet und gut lesbar, vor allem unter Berücksichtigung der Darstellung am Bildschirm.	The font is suitable in form and size and is easy to read, particularly with regard to its appearance on the screen.

6

MYSTERY OF COGNITIVE STRUCTURE? 🕮

Many research studies have clearly demonstrated the importance of cognitive structures as the building blocks of meaningful learning and retention of instructional materials. Identifying the learners' cognitive structures will help instructors to organize materials, identify knowledge gaps, and relate new materials to existing slots or anchors within the learners' cognitive structures. The purpose of this empirical investigation is to track the development of cognitive structures over time. Accordingly, it is demonstrated how various indicators derived from graph theory can be used for a precise description and analysis of cognitive structures. Results revealed several patterns that help to better understand the construction and development of cognitive structures over time. The chapter concludes by identifying applications for learning and instruction and proposing possibilities for the further development of the research approach.

🕮 This chapter is based on: Ifenthaler, D., Masduki, I., & Seel, N. M. (in press). 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

Introduction

Many research studies have clearly demonstrated the importance of cognitive structures, which refer to how concepts within a domain are organized and interrelated within a person's mind as the building blocks of meaningful learning and retention of instructional materials (Shavelson, 1974; Snow & Lohman, 1989). Ausubel (1963) highlighted the importance of this hypothetical construct as the principal factor in the accumulation of knowledge: "If existing cognitive structure is clear, stable, and suitably organized, it facilitates the learning and retention of new subject matter. If it is unstable, ambiguous, disorganized, or chaotically organized; it inhibits learning and retention" (p. 217).

As pointed out by Jonassen (1987), identifying the learners' cognitive structures will help instructors to organize materials, identify knowledge gaps, and relate new materials to existing slots or anchors within the learners' cognitive structures. In the process, misconceptions and preconceptions can also be identified and rectified (Seel, 1999a). The diagnosis of cognitive structures can act as a "topographic map" to identify key areas of learning difficulties and facilitate instructional interventions (Snow, 1989).

This approach can lead to the most suitable methods of instruction being utilized since different instructional strategies can lead to different cognitive structures and therefore to different learning outcomes (Mayer & Greeno, 1972). It can also be used to assess the effectiveness of learning by comparing the students' cognitive structures to those of instructors, domain experts, and even to the knowledge structures of other outstanding students (Acton, Johnson, & Goldsmith, 1994; Herl, Baker, & Niemi, 1996; Jonassen, 1987).

Numerous researchers have explored techniques for assessing and analyzing cognitive structures (Clariana & Wallace, 2007; Ifenthaler, 2006; Jonassen, 1987; Kalyuga, 2006a, 2006b; Koubek & Mountjoy, 1991; Pirnay-Dummer, 2006; Preece, 1976; Young, 1998). Some of these methods, however, can be too time consuming and unsuitable as an assessment tool within instructional environments such as a classroom or work setting (Kalyuga, 2006b; Spector, et al., 2006). Additionally, some of the techniques may have questionable reliability and validity in terms of assessment outcomes (Seel, 1999a).

The purpose of this empirical investigation is to track the development of cognitive structures over time. Accordingly, it is demonstrated how various indicators derived from graph theory can be used for a precise description and analysis of cognitive structures. The following section focuses on various definitions of cognitive structures. In the next section the perennial question of how to accurately diagnose cognitive structures is discussed. Then, the experimental study and the results are presented; followed by a discussion of how the research approach can be used to assess and analyze cognitive structures in various instructional settings. Finally, suggestions for further development of research approach are presented.

Cognitive structure

The advent of adaptive learning environments with its emphasis on learners' variable proficiency levels and cognitive preferences places greater urgency on the need for reliable and valid methods of diagnosing learners' cognitive structures (Kalyuga, 2006a; Snow, 1990). The term "cognitive structures," however, has many interpretations and since the definition of "cognitive structures" as a construct has strong implications on how it will be measured (Shavelson & Stanton, 1975), it is imperative that various definitions by researchers be examined for a better understanding of the term.

Many researchers conceive of cognitive structures, also known as knowledge structures or structural knowledge (Jonassen, et al., 1993), as the manner in which an individual arranges facts, concepts, propositions, theories, and raw data at any point in time (Taber, 2000), or more specifically as "a hypothetical construct referring to the organization of the relationships of concepts in memory" (Shavelson, 1972, p. 226). It is assumed that the order in which information is retrieved from long-term memory will reflect in part the individual's cognitive structure within and between concepts. By assessing the structure, even partially, the educator comes closer to influencing it in the student's memory so that it corresponds with the structure of instructional materials. In other words, learning requires students to reorganize their cognitive structures, which are made up of a collection of ideas in semantic memory (Jonassen, 1988). These ideas are also known as "schema" and can be an object, event, or proposition with a set of attributes that the individual perceives as being associated with the idea. For example, the schema for a pencil can include attributes such as its shape and also its function as a writing tool that occasionally needs sharpening.

According to Seel (1991) new information can be assimilated by a learner through the activation of an existing schema. In other words, an individual utilizes an existing schema in order to makes sense of the new information. In instances where the new information does not exactly fit into the schema, the schema undergoes adjustments by means of *accretion*, *tuning*, or *reorganization* (see Rumelhart & Norman, 1978). *Accretion* is the process of fitting in the new information into the existing areas within a schema. *Tuning* is defined as the process of changing certain parts of a schema to accommodate the new information. The outcome of the accretion and tuning process is the comprehension of the new information or as subjective plausible solutions to a problem. However, if *accretion* and *tuning* are unsuccessful, or in situations where no schema existed in the first place; new information is accommodated by means of the *reorganization* process. In other words, the individual uses the new information to create a new schema.

The accommodation process often leads to the development of mental models, which are dynamic ad hoc representations of reality to help the individual understand or simplify a phenomenon (see Gentner & Stevens, 1983; Johnson-Laird, 1983; Seel, 1991, 2001).

Hence, an individual's cognitive structure is made up of various schemata and mental models that can be embedded within one another within a hierarchy. A schema provides a framework that is used to interrelate various components of information about a topic into one conceptual unit. A schema is also made up of statements about important attributes of the conceptual unit, its purpose, and rules for selecting as well as using it (Norman, Gentner, & Stevens, 1976). These concepts are all organized within an interrelated network known as a semantic network which represents our cognitive structures. Since the schemata in our semantic network are interrelated based on various associations, an accepted method for representing such networks is through active structural networks (see Quillian, 1968). These structural networks are represented by nodes (schemata) and labeled links that connect nodes to one another – making it possible to represent what a learner knows through these networks. Learning thus takes place when we create new nodes that are then linked to the existing ones and to each other. In other words, new cognitive structures are built upon pre-existing structures (Norman, et al., 1976).

Koubek and colleagues (1994; 1991) expanded upon the attributes of knowledge structures as “the structure of interrelationships between elements, concepts and procedures in a particular domain, organized into a unified body of knowledge.” Within a given domain, elements refer to unique units of information which can be declarative elements such as concepts or facts; or procedural elements pertaining to how to do things within the domain. An individual’s knowledge structure is made up of the interrelationships between these elements. In this regard, cognitive structures can also be viewed as conceptual knowledge which transcends the mere storage of declarative knowledge. It is “an understanding of a concept's operational structure within itself and between associated concepts.” Through knowledge of the interrelationships between concepts, conceptual knowledge can be used to develop procedural knowledge for problem solving purposes within a specific domain (Tennyson & Cocchiarella, 1986).

Therefore, cognitive structure has major implications for comprehension, integration of new information, and the ability to solve domain-specific problems (Jonassen, et al., 1993; Shavelson, 1974). When compared to that of a novice, a domain expert’s cognitive structure is considered to be more tightly integrated and has a greater number of linkages among interrelated concepts. There is thus immense interest on the part of researchers to assess a novice’s cognitive structure and compare it with an expert’s in order to identify the most appropriate ways to bridge the gap.

Diagnosis of cognitive structures

Given the relevance of cognitive structures as a construct for assessing knowledge organization, assimilation, and accommodation, the perennial question is how to accurately diagnose them. Some issues that have yet to be resolved include identifying reliable and valid tools to elicit the external representation of such internal structures and the actual analysis of the structures themselves (Ifenthaler, 2008; Jonassen, et al., 1993; Kalyuga, 2006a). However, as it is not possible to measure cognitive structures directly, individuals have to elicit or externalize them before researchers can analyze and interpret them (see Ifenthaler, 2008).

Elicitation of cognitive structure

A variety of techniques have been developed which can be classified as (a) *natural language* and as (b) *graphical approaches*. Prominent *natural language approaches* are (1) *Thinking Aloud Protocols* (e.g., Ericsson & Simon, 1993, 1998), (2) *Word Association* (e.g., Gunstone, 1980; Shavelson, 1972), (3) *Structure Formation Technique* (Scheele & Groeben, 1984), and (4) *MITOCAR*, which stands for Model Inspection Trace of Concepts and Relations (Pirnay-Dummer, 2006). These and other natural language-based approaches utilize the most automated and natural means by which humans externalize their cognitive structures. They enable the verbalization of individual cognitive processes. However, Nisbett and Wilson (1977) question the quantification of the collected data and the explicit relation to cognitive processes as well validity and reliability of such techniques. On the other hand, it is argued that *natural language approaches* are less biased than *graphical approaches*, because natural language is more trained and highly automated (Pirnay-Dummer, 2006). However, *graphical approaches* such as (1) *Concept Mapping Tools* (Cañas, et al., 2004; Nückles, et al., 2004), (2) *Test for Causal Diagrams* (Al-Diban, 2002), (3) *DEEP*, which stands for Dynamic Evaluation of Enhanced Problem-solving (Pirnay-Dummer, et al., 2010; Spector & Koszalka, 2004), and (4) *Pathfinder* (Schvaneveldt, 1990) also provide a sound basis for the elicitation of cognitive structures. Undeniably, the application of graphical approaches must always include extensive training on how to use these tools. Nevertheless and regardless of the type of approach, we claim that tools which are used for the elicitation and analysis of cognitive structure must have a strong theoretical foundation and need to be tested for reliability and validity accordingly (Ifenthaler, 2010c).

Tracking changes in cognitive structure

Equally important are the issues of tracking the progression of cognitive structures, which captures the transition of learners from the initial state to the desired state (Snow, 1989, 1990); and for repetitive measurements of change over an extended period of time for a more accurate diagnosis (Ifenthaler & Seel, 2005; Seel, 1999a). Accordingly, research on cognitive structures needs to move beyond the traditional two-wave design in order to capture changes more precisely (Spada, 1983; Willett, 1988). As individuals reinstate and modify their cognitive structures when interacting with the environment (Jonassen, et al., 1993; Piaget, 1976; Seel, 1991), the necessity of conducting multiwave longitudinal experiments is evident. However, the collection and analysis of longitudinal data implicates various methodological dilemmas which should not be neglected (see Ifenthaler, 2008; Seel, 1999a). Besides general concerns about quantitative studies over time (Collins & Sayer, 2001;

Moskowitz & Hershberger, 2002), tracking changes in cognitive structures requires valid and reliable assessment techniques, adequate statistical procedures, and specific situations which enable the activation of such cognitive structures (Ifenthaler, 2008).

Measures of analyzing cognitive structure

As mentioned above, different approaches and tools can be applied to elicit cognitive structures. Accordingly, there are also various possibilities to measure cognitive structures (Koubek & Mountjoy, 1991). However, available methods are often very time consuming and sometimes limited in their ability to precisely measure cognitive structures (see Kalyuga, 2006a).

Therefore, our measurement technique is computer-based and highly automated, which enables us to analyze even larger sets of data within a few seconds. The foundation for analyzing cognitive structures is based on indicators derived from graph theory (Diestel, 2000; Harary, 1974). Graph theory is a promising approach and its fundamentals have been applied in various fields of research and practice, e.g. decision making, project management, network problems, etc. (Chartrand, 1977). A graph is constructed from a set of *vertices* whose relationships are represented by *edges*. Basics of graph theory are necessary to describe externalized cognitive structures as graphs (Bonato, 1990).

A graph $G(V,E)$ is composed of vertices V and edges E. If the relationship between vertices V is directional, a graph is called a directed graph or digraph D. A graph which contains no directions is called an undirected graph.

The position of vertices V and edges E on a graph G are examined with regard to their proximity to one another. Two vertices x, y of G are adjacent if they are joined by an edge e. Two edges $e \neq f$ are adjacent if they have a common end or vertex x.

A path P is a graph G where the vertices x_i are all distinct. The length of a path P is calculated by the number of its edges e_j. The vertices x_0 and x_k are called the ends of the path P.

A graph G is indexed when single vertices V and edges E are distinguished by their names or content.

Every connected graph G contains a spanning tree. A spanning tree is acyclic and includes all vertices of G. Spanning trees can be used for numerous descriptions and calculations concerning the structure of a graph.

By describing externalized cognitive structures as graphs, including associated vertices and edges, we are able to apply various measures from graph theory to analyze individual cognitive structures and, in addition, to track the development of cognitive structures over time (see Table 6.1).

TABLE 6.1
Measures for analyzing the organization of cognitive structures

Measure	Operationalization	Computation
Surface Structure	The overall number of propositions (node-link-node) is an indicator for the development of a cognitive structure.	Computed as the sum of all propositions (node-link-node) of a cognitive structure. Defined as a value between 0 (no proposition) and *N* (*N* propositions of the cognitive structure).
Matching Structure	The complexity of a cognitive structure indicates how broad the understanding of the underlying subject matter is.	Computed as the quantity of edges of the shortest path between the most distant nodes (diameter) of the spanning tree of a cognitive structure. Defined as a value between 0 (no edges) and *N*.
Connectedness	A connected cognitive structure indicates a deeper understanding of the underlying subject matter.	Computed as the possibility to reach every vertex from every other vertex in the cognitive structure. Defined as a value between 0 (not connected) and 1 (connected).
Ruggedness	Non-linked vertices of a cognitive structure point to a lesser understanding of the phenomenon in question.	Computed as the sum of subgraphs which are independent or not linked. Defined as a value between 1 (all vertices are linked) and *N*.
Average degree of Vertices	As the number of incoming and outgoing edges grows, the complexity of the cognitive structure is taken as more complex.	Computed as the average degree of all incoming and outgoing edges of the cognitive structure. Defined as a value between 0 and *N*.
Cyclic	A non-cyclic cognitive structure is considered less sophisticated.	A cyclic cognitive structure contains a path returning back to the start vertex of the starting edge. Defined as a value between 0 (no cycles) and 1 (is cyclic).
Number of Cycles	A cognitive structure with many cycles is an indicator for a close association of the vertices and edges used.	Computed as the sum of all cycles within a cognitive structure. Defined as a value between 0 (no cycles) and *N*.
Vertices	A simple indicator for the size of the underlying cognitive structure.	Computed as the sum of all vertices within a cognitive structure. Defined as a value between 0 (no vertices) and *N*.
Edges	A simple indicator for the size of the underlying cognitive structure.	Computed as the sum of all edges within a cognitive structure. Defined as a value between 0 (no edges) and *N*.

Table 6.2 provides additional measures for analyzing and comparing the semantic content of the cognitive structures.

Besides the three core measures (surface structure, graphical structure, propositional matching), we implemented the graph theory based measures as supplementary indicators into our computer-based analysis tool SMD Technology (Surface, Matching, Deep Structure). In an automated iterative process, the SMD Technology (Ifenthaler, 2010c) calculates numerical indicators for all measures described in Tables 6.1 and 6.2 and stores them in a database.

TABLE 6.2
Measures for analyzing the semantic content of cognitive structures

Measure	Operationalization	Computation
Vertex Matching	The use of semantically correct concepts (vertices) is a general indicator of an accurate understanding of the given subject domain.	Computed as the sum of vertices of a cognitive structure which are semantically similar to a domain specific reference cognitive structure (e.g. expert structure). Defined as a value between 0 (no semantic similar vertices) and *N*.
Propositional Structure	The use of semantically correct propositions (vertex-edge-vertex) indicates a correct and deeper understanding of the given subject domain.	Calculated as the semantic similarity of a cognitive structure and a domain specific reference cognitive structure. Defined as a value between 0 (no similarity) and 1 (complete similarity).

Additionally, standardized graphical re-representations of the externalized cognitive structures are generated. Figures 6.1 and 6.2 show two standardized re-representations constructed by a participant at time points 1 and 5 of our experiment. In the following, we will briefly expound on the above described measures for analyzing the organization and semantic content of cognitive structures using the examples in Figure 6.1 and 6.2.

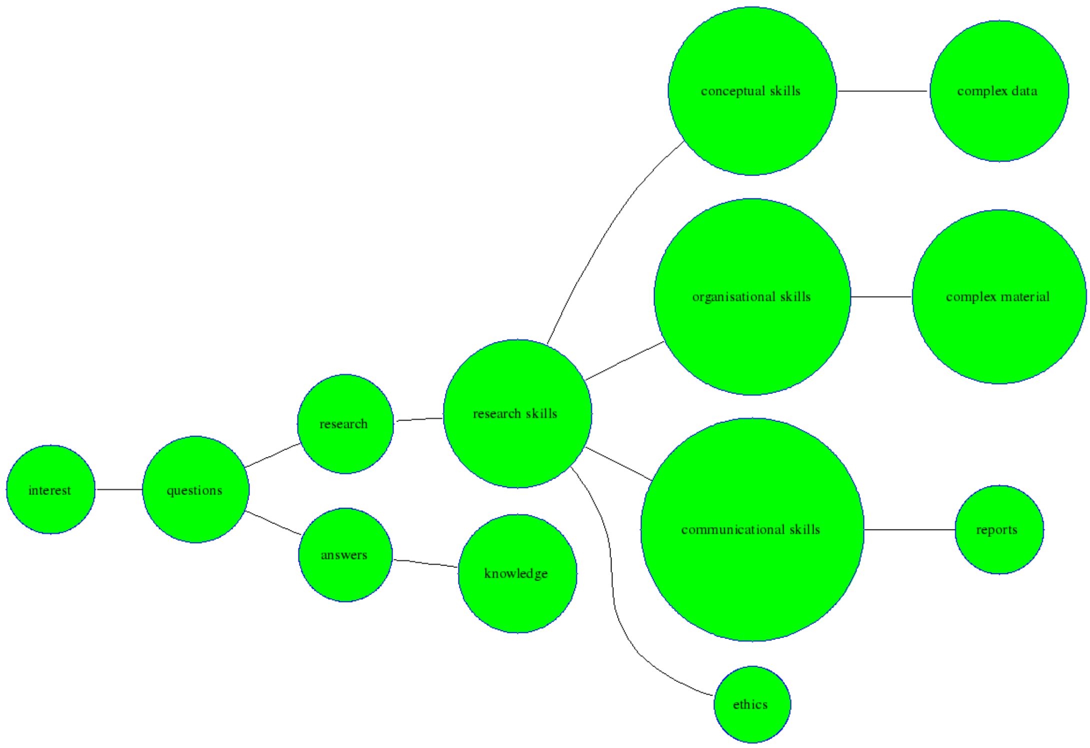

FIGURE 6.1*. Standardized re-representation of a participant's cognitive structure at time point 1*

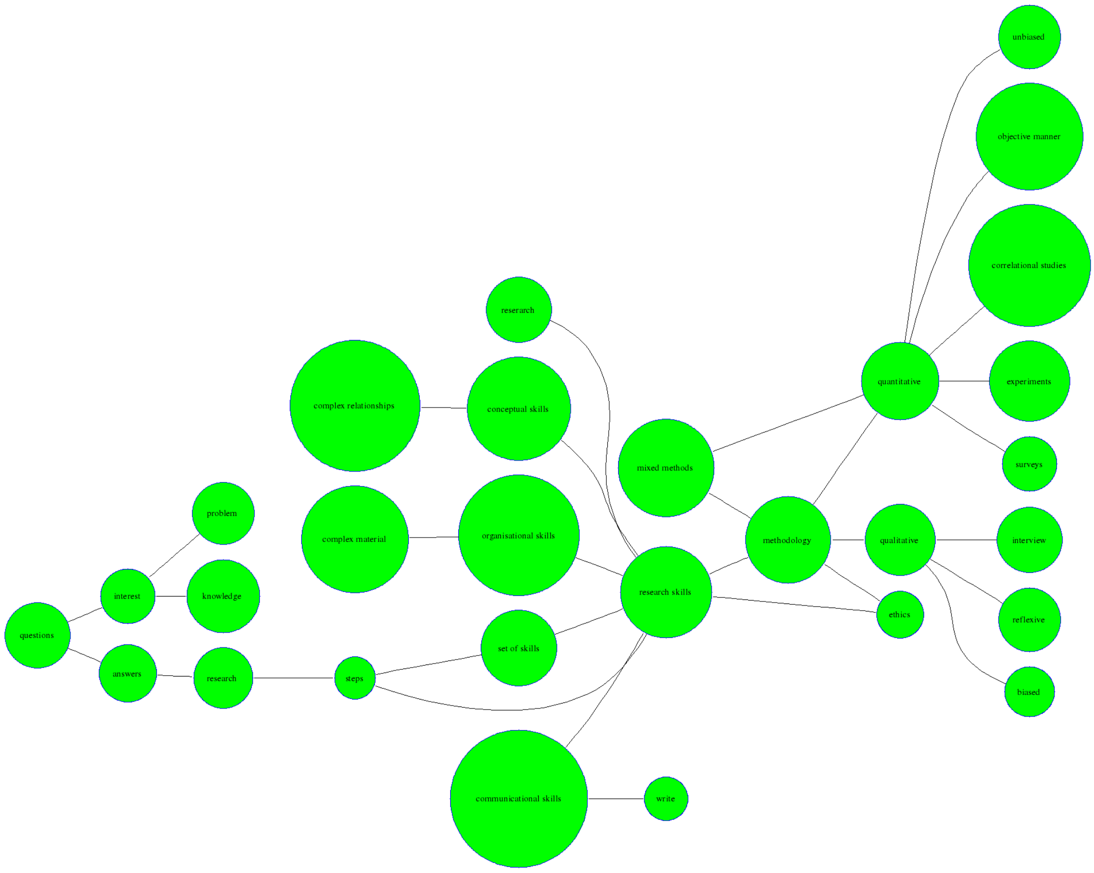

FIGURE 6.2. *Standardized re-representation of a participant's cognitive structure at time point 5*

Table 6.3 shows the calculated measures for quantitatively describing the organization and semantic content of the two examples. The *surface structure* more than doubles during the learning process. This is also indicated by the measure *vertices*, which increases from 13 to 29. We conclude that the cognitive structure of the participant develops during the learning process. With the help of the measure *graphical structure*, we are able to find out whether the complexity of the cognitive structure also increases. In order to calculate the *graphical structure* of the two examples, a spanning tree is generated first. A spanning tree of Figure 6.1 or 6.2 contains all vertices but no cycles. Then, the diameter of the spanning tree (shortest longest path) is calculated. As shown in Table 6.3, the diameter increases from 6 to 9 in our two examples. Corresponding to this result, the measures *connectedness* and *ruggedness* give further information about the complexity of the cognitive structure. In both cases, the re-representations are connected – every vertex can be reached from every other vertex. This means that the participant has a deep understanding of the underlying subject matter and is able to connect various concepts (vertices) together. Accordingly, the measure *ruggedness* is 1. If this indicator were greater than 1 it would indicate that the cognitive structure is divided into subsections (subgraphs). Thus, a less connected cognitive structure points to a poorer understanding of the subject matter. Furthermore, the measures *cyclic* and *number of cycles* point to an interesting difference between the two examples. The re-representation in Figure 1 has no cycles; our example in Figure 6.2 has three cycles. This means that our participant added more associations of concepts to her cognitive structure while studying the subject matter. The *average degree of vertices* in both examples indicates that most concept have an incoming and an outgoing link.

TABLE 6.3
Measures calculated for the example re-representations in Figures 1 and 2

Measure	*Result Figure 1*	*Result Figure 2*
Surface Structure	14	31
Graphical Structure	6	9
Connectedness	1	1
Ruggedness	1	1
Average degree of Vertices	2.11	2.14
Cyclic	0	1
Number of Cycles	0	3
Vertices	13	29
Vertex Matching	0.12	0.52
Propositional Matching	0.04	0.19

However, not all organizational indicators include information about the correctness of the concepts and links within the re-representation. Our measures *vertex* and *propositional matching* provide this information about the semantic content. The number of semantically correct vertices and propositions (compared to an expert re-representation) increases during the learning process. Accordingly, not only does the organization of the cognitive structure grow more complex, it also becomes more correct in comparison with that of an expert.

Assumptions and hypotheses

As they are able to automatically describe and analyze large sets of data, we assume that these indicators are applicable for tracking the development of externalized cognitive structures over time. This leads to the following assumptions and hypotheses, which were tested in our experimental study.

$H_{1.1}$: The *organization* of the externalized cognitive structures changes during the learning process.
$H_{1.0}$: The *organization* of the externalized cognitive structures does not change during the learning process.
$H_{2.1a}$: The *numbers of semantic correct vertices* of the externalized cognitive structures become more similar to the expert structure during the learning process.
$H_{2.0a}$: The *numbers of semantic correct vertices* of the externalized cognitive structures have no or only little similarity to the expert structure.
$H_{2.1b}$: The *numbers of semantic correct propositions* of the externalized cognitive structures become more similar to the expert structure during the learning process.
$H_{2.0b}$: The *numbers of semantic correct propositions* of the externalized cognitive structures have no or only little similarity to the expert structure.
$H_{3.1}$: The development of the *organization* of the externalized cognitive structures influences the course learning outcomes.
$H_{3.0:}$ The development of the *organization* of the externalized cognitive structures has no or only little influence on the course learning outcomes.

The (a) *organization* and (b) *semantic* nature of the cognitive structures changes during the learning process. Further, we assume (c) a correlation between the course learning outcome and the organization / semantics of the externalized cognitive structures.

Method

Participants

Twenty-five students (18 female and 7 male) from the University of Freiburg, Germany, participated in the study. Their average age was 24.7 years (*SD* = 1.9). All students attended an introductory course on *research methods* in the winter semester 2007. A total of 125 concept maps were collected at 5 measurement points during the semester.

Procedure

Data were collected through concept maps using the software *CmapTools* (Cañas, et al., 2004). According to Novak (1998), a concept map is a graphical two-dimensional representation of communicated knowledge and its underlying structure. A concept map consists of *concepts* (graph theory: vertices) and *relations* (graph theory: edges). Research studies on the application of *CmapTools* indicate a wide acceptance of our theoretical assumptions on using this software (e.g. Coffey, et al., 2003; Derbentseva, Safayeni, & Cañas, 2004). Since our research study focuses on the development of cognitive structures, our longitudinal procedure included five measurement points. The main parts of our study were as follows:

In a 60 minute introductory lesson, the subjects were introduced to the concept mapping technique and taught how to use the *CmapTools* software. Additionally, the instructor collected demographic data and delivered documentation on concept maps and the software, including examples.

At five measurement points (MP, see Figure 3) during the course on research methods, the subjects were asked to create an open concept map relating to her or his understanding of *research skills*. Every subject needed to upload the concept map at a specified date and time during the course.

The course learning outcome was measured through (1) five written assignments, (2) a written exam, and (3) a written research proposal. The score of the course learning outcome was rated between 0 and 100 points (Spearman-Brown-Coefficient, *r* = .902).

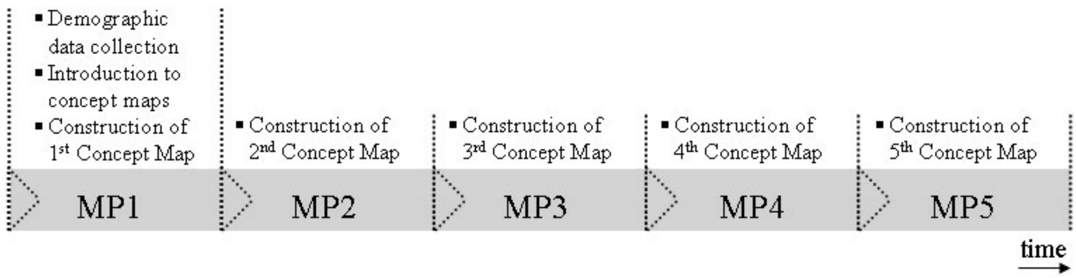

FIGURE 6.3. *Longitudinal research design*

After uploading the concept maps, the instructor gave the students a brief feedback to notify them that their maps had been successfully uploaded and that they should carry on with their studies in the course. As we used open concept maps in our research study, the subjects were not limited to specific words while annotating the concepts and relations.

Analysis procedure

Using the export function of *CmapTools*, we were able to store the subjects' concept maps pairwise (as propositions) in a raw data table, including the (a) *subject number*,

(b) *measurement point*, (c) *vertex 1*, (d) *vertex 2*, and (e) *edge* connecting the two vertices. Having the raw data at hand, we uploaded all information onto the SQL database of our own SMD Technology (Ifenthaler, 2010c). We used the computer-based analysis tool SMD Technology to calculate the above described graph theory based measures. Accordingly, the automated analysis process provides 11 indicators (see Table 1) for each subject representation. The SMD Technology has been tested extensively for reliability (e.g., test-retest reliability for $r_{surface}$ = .824*; $r_{graphical}$ = .815*; $r_{propositional}$ = 901*) and validity (convergent and divergent validity $r_{surfaceXmitocar}$ = .610**; $r_{graphicalXmitocar}$ = .527**).

However, the statistical analysis of such longitudinal data requires a sharpened awareness of the problems involved in the measurement of change (e.g., Collins & Sayer, 2001; Harris, 1963; Ifenthaler, 2008). Accordingly, besides standard statistical procedures, we used HLM (Hierarchical Linear Models), which offers a wide spectrum of data analysis for longitudinal data (Raudenbush & Bryk, 2002). The HLM analysis is realized in two analysis steps. The first growth model (*Level 1*; equation 1.1) tests the intraindividual change of the dependent variables.

$$Y_{ti} = \pi_{0i} + \pi_{1i}(TIME) + e_{ti} \qquad [1.1]$$

The second growth model (*Level 2*; equation 1.2) tests for possible effects of additional variables (e.g., student performance).

$$\begin{aligned} \pi_{0i} &= \gamma_{00} + \gamma_{01}\, PREDICTOR + \xi_{0i} \\ \pi_{1i} &= \gamma_{10} + \gamma_{11}\, PREDICTOR + \xi_{1i} \end{aligned} \qquad [1.2]$$

Results

Our in-depth analysis of *N*=125 cognitive structures (5 re-representation of each of the 25 participants) revealed several patterns that helped us to better understand the construction and development of these constructs over time. To describe our results, we will first present descriptive results and corresponding figures (see Figures 6.4 and 6.5). We will then show the outcomes of our HLM and correlation analysis.

Descriptive analysis

The average *course learning outcome* of all subjects was *M*=84.68 (*SD*=10.53, *Min*=46, *Max*=96). The results of our cognitive structure measures (organization and semantic content) are described in Tables 6.4 and 6.5.

The sum of propositions (*Surface Structure*) increases throughout the five measurement points (*Min*=1, *Max*=247). Equally, the sum of *vertices* increases from MP1 to MP5. A total of *n*=57 (45.6 %) cognitive structures were fully *connected* (the possibility to reach every vertex from every other vertex). However, the average number of sub graphs (*Ruggedness*) nearly doubled from MP1 (*Min*=1, *Max*=3) to MP5 (*Min*=1, *Max*=8).

TABLE 6.4
Average scores (standard deviations in parenthesis) of graph theory based measures (organization) for measurement points 1 – 5 (*N*=25)

		MP1	MP2	MP3	MP4	MP5
Surface Structure	M	14.64	27.34	45.84	67.72	71.80
	(SD)	(7.99)	(14.13)	(23.85)	(48.94)	(46.71)
Graphical Structure	M	5.52	7.62	9.48	12.08	11.72
	(SD)	(2.83)	(3.57)	(3.42)	(4.91)	(4.19)
Connectedness	M	.68	.80	.44	.44	.36
	(SD)	(.48)	(.41)	(.51)	(.51)	(.49)
Ruggedness	M	1.44	1.32	2.12	2.28	2.72
	(SD)	(.71)	(.74)	(1.42)	(1.49)	(2.01)
Average Degree of Vertices	M	1.93	2.06	2.12	2.11	2.09
	(SD)	(.43)	(.53)	(.39)	(.24)	(.26)
Number of Cycles	M	2.52	3.38	4.12	4.76	4.48
	(SD)	(2.37)	(2.59)	(2.68)	(3.95)	(3.00)
Number of Vertices	M	14.40	24.65	42.24	63.96	68.16
	(SD)	(6.69)	(11.76)	(22.60)	(45.85)	(44.33)

Additionally, the increase in complexity of the cognitive structures is described by the *Graphical Structure* (*Min*=1, *Max*=24) and the *Degree of Vertices* (*Min*=1, *Max*=3.44). 76.8 % (*n*= 96) of all cognitive structures contained a *cycle* (a path returning back to the start vertex of the starting edge). We found also an increase in the average number of cycles from MP1 (*Min*=0, *Max*=8) to MP5 (*Min*=0, *Max*=12).

TABLE 6.5
Average scores (standard deviations in parenthesis) of graph theory based measures (semantic content) for measurement points 1 – 5 (*N*=25)

		MP1	MP2	MP3	MP4	MP5
Vertex Matching	M	7.00	12.76	17.16	21.00	21.24
		(3.97)	(6.11)	(7.33)	(8.12)	(8.19)
Propositional Matching	M	.0099	.0288	.0247	.0379	.0383
		(.0186)	(.0363)	(.0316)	(.0370)	(.0399)

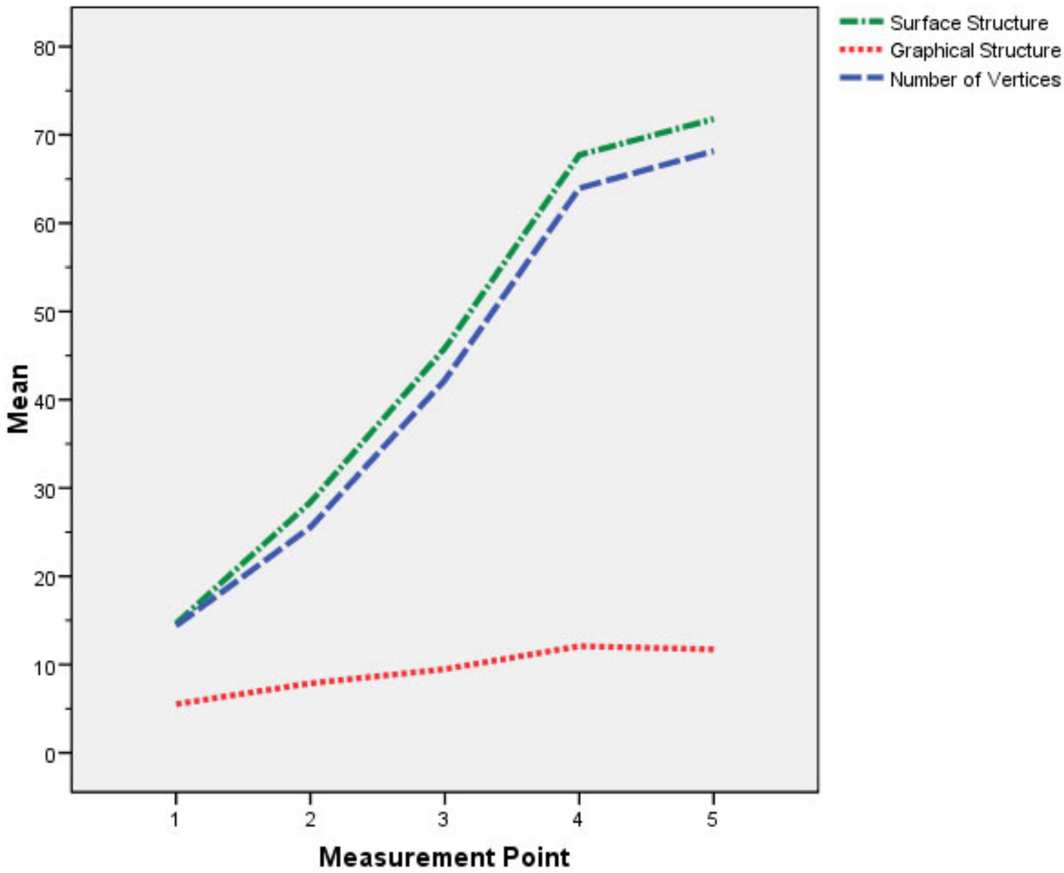

FIGURE 6.4. Development of cognitive structures over time

The *Vertex Matching* (semantically similar vertices) increases throughout the five measurement points (*Min*=0, *Max*=34). The *Propositional Matching*, which describes the semantically similar propositions between an individual cognitive structure and an expert representation, also increases, but the overall similarity to the expert representation is rather low.

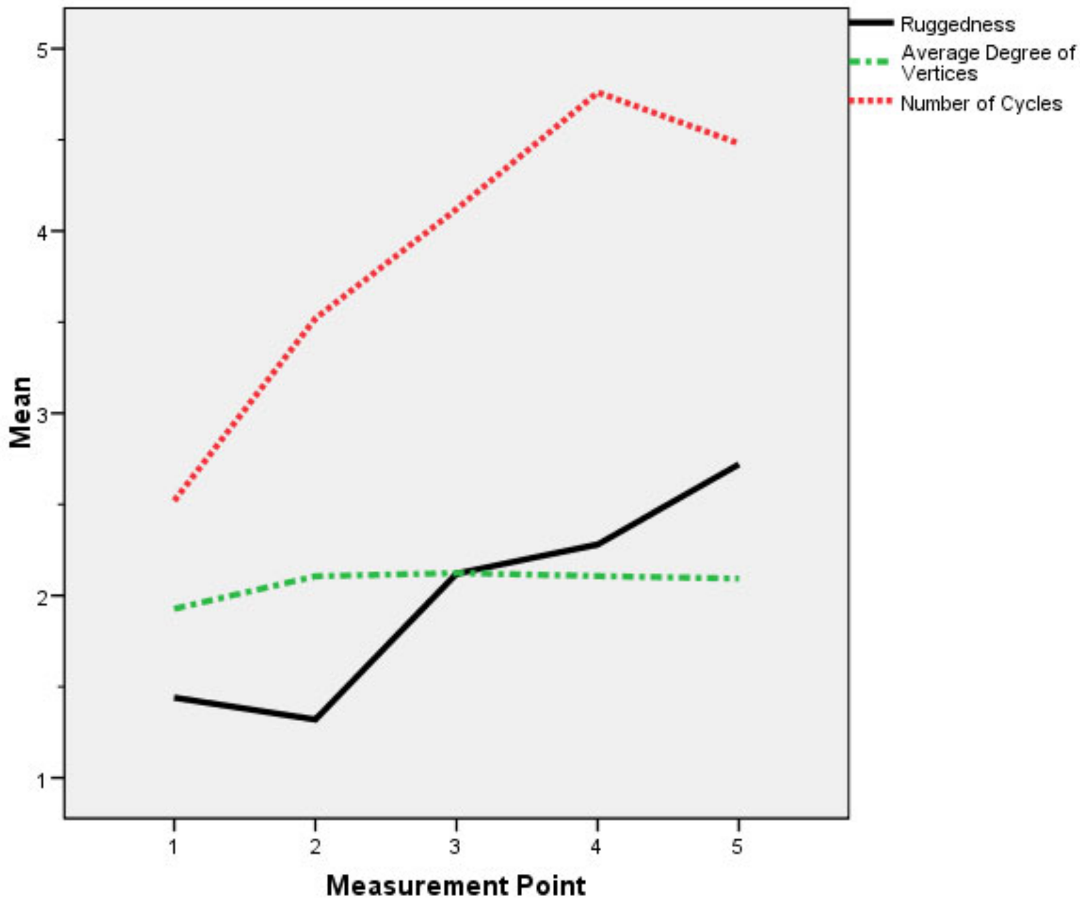

FIGURE 6.5. *Development of cognitive structures over time*

HLM analysis

To test our hypothesis we computed several HLM analyses. According to Hox (2002), the sample size of our study is just adequate. However, in order to validate our initial findings we suggest further studies with larger sample size. The results of our *Level-1* HLM analysis (intraindividual change of cognitive structures over time) are described in Tables 6.6 and 6.7. The *Mean Initial Status* π_{0i} indicates that all corresponding measures are significantly higher than 0. Although this is a rather trivial effect (see Renkl & Gruber, 1995), we think it is useful to examine all HLM results. Except for *Average Degree of Vertices*, all other measures reveal a significant positive linear *Mean Growth Rate* π_{1i} per measurement point (e.g. *Surface Structure* = 15.36).

Therefore, we accept $H_{1.1}$: The *organization* (Surface Structure, Graphical Structure, Ruggedness, Number of Cycles, and Number of Vertices) of the externalized cognitive structures changes during the learning process, except for the measure *Average Degree of Vertices*. The *Average Degree of Vertices* indicates the average number of incoming and outgoing edges. Accordingly, as most of the externalized cognitive structures are very broad and do not center in one vertex, each vertex takes two edges in average (see Table 6.4). This does not change during the learning process, as the subject domain (research skills) does not change and does not seem to be organized around one central vertex.

Likewise, our HLM analysis revealed a significant positive linear *Mean Growth Rate* π_{1i} per measurement point for the measure *Vertex Matching* (3.67). This means that the subjects used more and more correct concepts (vertices) compared to the expert cognitive structure.

TABLE 6.6
Level-1 linear growth models of cognitive structures (organizational measures)

		Coefficient	*SE*	*t*	*df*	*p*
Surface Structure	Mean Initial Status π_{0i}	14.95	1.95	7.64	24	<.001
	Mean Growth Rate π_{1i}	15.36	2.72	5.65	123	<.001
Graphiical Structure	Mean Initial Status π_{0i}	6.02	0.49	12.09	24	<.001
	Mean Growth Rate π_{1i}	1.66	0.29	5.62	123	<.001
Ruggedness	Mean Initial Status π_{0i}	1.27	0.11	11.48	24	<.001
	Mean Growth Rate π_{1i}	0.35	0.11	3.32	123	.002
Average Degree of Vertices	Mean Initial Status π_{0i}	2.01	0.08	24.19	24	<.001
	Mean Growth Rate π_{1i}	0.03	0.03	1.32	123	.189
Number of Cycles	Mean Initial Status π_{0i}	2.85	0.44	6.49	24	<.001
	Mean Growth Rate π_{1i}	0.52	0.19	2.69	123	.008
Number of Vertices	Mean Initial Status π_{0i}	13.68	1.79	7.65	24	<.001
	Mean Growth Rate π_{1i}	14.59	2.63	5.56	123	<.001

Therefore, we accept $H_{2.1a}$: The *numbers of semantic correct vertices* of the externalized cognitive structures become more similar to the expert structure during the learning process.

TABLE 6.7
Level-1 linear growth models of cognitive structures (semantic measures)

		Coefficient	*SE*	*t*	*df*	*p*
Vertex Matching	Mean Initial Status π_{0i}	8.49	0.85	9.94	24	<.001
	Mean Growth Rate π_{1i}	3.67	0.41	8.99	123	<.001
Propositional Matching	Mean Initial Status π_{0i}	0.0317	0.0056	5.63	24	<.001
	Mean Growth Rate π_{1i}	-0.0019	0.0016	-1.15	123	0.253

Contrary to our expectations, we found no significant growth (*Mean Growth Rate* π_{1i}) for the semantic measures *Propositional Matching* (see Table 6.7). The cognitive structures became only slightly more similar to the expert structure during the five measurement points.

Therefore, $H_{2.1b}$ had to be rejected in favor of $H_{2.0b}$: The numbers of semantic correct propositions of the externalized cognitive structures had no or only little *semantic similarity* with the expert structure.

For all graph theory based measures, we computed a *Level-2* HLM analysis for the predictor *learning* (course learning outcome; median split: 0 = low learning outcome, 1 = high learning outcome). We found no significant difference between subjects with low learning outcomes and high learning outcomes in an analysis of the development of their cognitive structures using the graph theory based measures. The general *Level*-2 equation results through substitution as follows (e.g., Surface Structure):

$$\begin{aligned} \pi_{0i} &= 11.98 + 6.18 LEARNING_i \\ \pi_{1i} &= 13.00 + 4.93 LEARNING_i \end{aligned} \qquad [1.3]$$

The *Surface Structure* of subjects with *low learning outcomes* scores an average of 11.98. Subjects with *high learning outcomes* score an average of 18.16 (11.98+6.18). However, this difference is not significant. Additionally, the *Surface Structure* of subjects with *low learning outcomes* increases significantly by 13.00 per measurement points. However, the higher increase of the *Surface Structure* of subjects with *higher learning outcomes* by 17.93 (13.00+4.93) is not significantly different from that of the subjects with *lower learning outcomes*. Details for all graph theory based measures of the *Level-2* HLM analysis are reported in Appendix A (Tables 6.9 and 6.10). Therefore, $H_{3.1}$ had to be rejected in favor of $H_{3.0}$: The development of the *organization* of the externalized cognitive structures has no or only little influence on the course learning outcomes.

Correlational analysis

Table 6.8 shows the correlations for the course learning outcomes and the characteristics of the cognitive structures at the fifth measurement point. We found no significant correlation between the measures *surface structure*, *graphical structure*, *connectedness*, *ruggedness*, *number of vertices*, and *propositional matching*. However, the higher the learners' course learning outcome was, the higher was the *average degree of vertices*, $r = .58$, $p = .002$. Equally, the higher the course learning outcome was, the higher were the *number of cycles* measured in the cognitive structures, $r = .51$, $p = .009$.

Additionally, our analysis revealed a significant correlation between the course learning outcomes and the measure *vertex matching*, $r = .41$, $p = .038$ (i.e., the higher the course learning outcome was, the higher was the number of similar vertices between the subject and expert externalization).

TABLE 6.8
Pearson's correlations between cognitive structure (organization and semantic content) characteristics (MP 5) and course learning outcomes (*N*=25)

	r	p
Surface Structure	.22	.291
Graphical Structure	.31	.127
Connectedness	.31	.137
Ruggedness	-.34	.102
Average Degree of Vertices	.58**	.002
Number of Cycles	.51**	.009
Number of Vertices	.16	.438
Vertex Matching	.42*	.038
Propositional Matching	.23	.270

Note: * $p < .05$; ** $p < .01$

Discussion

The aim of this study was to diagnose the development of cognitive structures over time. For this purpose, we applied different measures derived from graph theory to precisely score the changes in the externalized cognitive structures.

According to the subjects, the software *CmapTools* applied to externalize the cognitive structures was user-friendly and motivated them to continue using it. Additionally, the export function of *CmapTools* enabled us to automatically include all assessed individual cognitive structures in our SQL database. Therefore, we conclude that the data transformation process from the *CmapTools* to our analysis database has a very high reliability.

Contrary to other non-automated and time-consuming techniques for scoring open-ended concept maps (e.g., Al-Diban, 2002), our automated analysis procedure is expeditious and computes the different measures within seconds. As shown in previous experiments, the core measures of the *SMD Technology* have a high reliability and validity (see Ifenthaler, 2006, 2010c). The additionally implemented graph theory based measures allow us to more precisely diagnose changes in the externalized cognitive structures.

The in-depth analysis of all 125 cognitive structures revealed several patterns that help us to better understand their construction and development during learning processes. We distinguish between two types of measures: The (1) *organizational measures* (Surface Structure, Graphical Structure, Ruggedness, Number of Cycles, and Number of Vertices) help us to exactly locate changes in the composition of the externalized cognitive structure. On the other hand, the (2) *semantic measures* (Vertex Matching, Propositional Matching) indicate whether the content of the vertices and propositions used by an individual is correct compared to an expert's cognitive structure.

The result of our HLM analysis revealed a significant growth in the *organizational measures* between measurement points one and five. The overall size of the cognitive structures (*Surface Structure*) increased many times over. Accordingly, this is an indicator for an accommodation process (see Piaget, 1976; Seel, 1991), i.e. the individuals continuously added new concepts (*Number of Vertices*) and links between concepts (*Surface Structure*) to their cognitive structures while learning. As a consequence, the complexity of the externalized cognitive structures also increased, which is indicated by the growth of the measure *Graphical Structure* and *Number of Cycles*. Therefore, we conclude that while learning and understanding more and more of a given subject matter, individuals are able to more tightly integrate single concepts and links. However, we also found a significant growth in the measure *Ruggedness* (i.e., non-linked concepts within the entire cognitive structure). The significant decrease in the measure *Connectedness* supports this result. This indicates that newly learned concepts are not immediately integrated into the cognitive structure. This delay of integrating concepts into the cognitive structure should be kept in mind when constructing instructional materials and learning environments. We also suggest analyzing this phenomenon in a future study more precisely.

Contrary to the results of the *organizational measures*, our HLM analysis revealed only a significant growth in the semantic measure *Vertex Matching*. The individuals use more and more semantic correct concepts (vertices) during the learning process. As individuals become more familiar with the terminology of the subject domain (in our study research methods), they use these concepts more frequently. This learning process enables individuals to communicate their cognitive structures more precisely and more expert like. To reaffirm our assumptions, we also found a significant positive correlation between the course learning outcomes and the number of semantically correct concepts (*Vertex Matching*).

However, we found no significant growth in the semantic measure *Propositional Matching*. This result indicates that the individuals in our experiment were far from using the same proposition for describing the phenomenon in question. Nevertheless, the semantic analysis of cognitive structures is still a challenging endeavor. Therefore, we suggest improving the validity of the semantic measures using other heuristics (e.g., Pirnay-Dummer, et al., 2010).

Besides the quantitative measures, our own *SMD Technology* generates standardized graphical re-representations of all assessed cognitive structures as well as *similarity* and *contrast* re-representations. A *similarity re-representation* includes only the semantically correct concepts (vertices) and links (edges). On the other hand, the *contrast re-representation* includes all concepts (vertices) and links (edges) which are semantically incorrect (Ifenthaler, 2010c).

The quantitative measures and graphical re-representations generated by *SMD Technology* have various potential applications within a learning environment, such as *knowledge diagnosis*, *self-assessments*, *rich feedback*, *prediction of performance on tasks*, and *knowledge sharing*.

In order to provide effective instruction, it is important for students' prior knowledge to be identified since the subsequent construction and organization of knowledge structures as well as mental models in a particular situation depends on the students' preconceptions and naïve theories (Seel, 1999a). Knowing where the students are in terms of their initial cognitive states and the eventual progression of learning enables the teacher to make adjustments at the right time to enhance instructional effectiveness (Ifenthaler & Seel, 2005) or to make necessary changes to the instructional materials as part of a formative feedback process (Shute & Zapata-Rivera, 2008).

Automated knowledge diagnosis can also play an important role in an adaptive learning environment or intelligent tutoring systems (ITS) by integrating student performance data (using the abovementioned quantitative measures or graphical re-representations) into the student model of an ITS, thus enabling the system to tailor instructions to students' individual needs. The system could identify gaps or discrepancies between the students' and the experts' re-representations; then provide the appropriate instructional content to overcome the deficiencies.

Another advantage of knowledge diagnosis is in relation to the possibility of self-assessment within an adaptive learning system (Ifenthaler, 2010c). The various quantitative indicators provide immediate information in terms of the range and complexity of the students' knowledge structures. Then by comparing their structures to an expert or other students', learners can make judgments about their own learning progress and identify areas of self-improvement. The immediacy of such comparisons can increase motivation by suggesting a course of action for the learners as well as the provision of constructive feedback (see Ifenthaler, 2009).

If the assessment of knowledge is carefully synchronized with specific tasks to be performed by the students, the SMD Technology can also be applied to provide detailed and individualized feedback for the execution of those tasks (Ifenthaler, 2010c). This would be more helpful for student performance compared to a general feedback indicating success or failure since the teacher or the computer system can not only point out the errors but also provide suggestions on how to correct them (Shute & Zapata-Rivera, 2008).

Additionally, a person's performance on a cognitive-oriented task can be predicted based on the characteristics of his or her knowledge structure (Koubek & Mountjoy, 1991). For example, a student with more complex knowledge structures may be ready for (and thus perform better) in higher-level problem solving tasks involving abstract domain-specific content, compared to a student whose knowledge structure is simpler. This can help the teacher or learning system allocate the appropriate level of assignment or the grouping of students as team members according to similar abilities.

In relation to team dynamics, the quantitative indicators and graphical re-representations could also be used to facilitate knowledge sharing among team members (Ifenthaler, 2010c). Team understanding for the completion of a task could be compared with each individual's understanding, thus differences can be identified and the task completed in an effective manner. *SMD Technology* outputs can also be used to identify tacit knowledge that exist within individuals so that it can then be communicated and integrated into the team knowledge structures. Such an application is especially useful when you have new group members who need to get up to speed quickly within team projects.

In summary, a precise and stepwise diagnosis of cognitive structures helps us to better understand the differences within and between individuals as they develop over time. This will enable us to identify the most appropriate instructional materials and instructor feedback to be provided at suitable times during the learning process. We also suggest diagnosis of developing cognitive structures in different subject domains in order to detect variations in terms of how cognitive structures develop between different content areas.

Conclusion and Future Work

Our future work will involve validating our results in various subject domains and larger sample sizes. The core measures and the newly developed graph theory based measures of the SMD Technology will be further developed and implemented as a standard analysis tool for web applications. We will mainly concentrate on developing a new alternative for analyzing the semantic content of the externalized cognitive structures. Additionally, we are highly motivated to combine our tool with other existing analysis techniques in order to increase the reliability and validity of the diagnosis of changing cognitive structures.

Appendix A

TABLE 6.9
Level-2 linear growth models of cognitive structures (organization) and course learning outcomes

		Coefficient	*SE*	*t*	*df*	*p*
Surface Structure	Mean Initial Status π_{0i}	11.98	1.54	7.77	23	<.001
	learning	6.18	3.82	1.62	23	0.119
	Mean Growth Rate π_{1i}	13.00	2.49	5.21	23	<.001
	learning	4.93	5.47	0.90	23	0.378
Graphical Structure	Mean Initial Status π_{0i}	5.28	0.53	9.82	23	<.001
	learning	1.54	0.96	1.61	23	0.122
	Mean Growth Rate π_{1i}	1.76	0.41	4.28	23	<.001
	learning	-0.21	0.59	-0.36	23	0.723
Ruggedness	Mean Initial Status π_{0i}	1.48	0.15	10.17	23	<.001
	learning	-0.43	0.20	-2.09	23	0.048
	Mean Growth Rate π_{1i}	0.29	0.14	2.04	23	0.053
	learning	0.12	0.21	0.59	23	0.562
Average Degree of Vertices	Mean Initial Status π_{0i}	1.79	0.09	18.00	23	<.001
	learning	0.46	0.14	3.29	23	0.004
	Mean Growth Rate π_{1i}	0.07	0.03	2.43	23	0.023
	learning	-0.07	0.05	-1.48	23	0.153
Number of Cycles	Mean Initial Status π_{0i}	1.68	0.54	3.12	23	0.005
	learning	2.44	0.73	3.36	23	0.003
	Mean Growth Rate π_{1i}	0.77	0.28	2.77	23	0.011
	learning	-0.53	0.37	-1.44	23	0.162
Number of Vertices	Mean Initial Status π_{0i}	12.35	1.23	10.09	23	<.001
	learning	2.76	3.65	0.76	23	0.456
	Mean Growth Rate π_{1i}	12.42	2.25	5.51	23	<.001
	learning	4.53	5.31	0.86	23	0.402

TABLE 6.10
Level-2 linear growth models of cognitive structures (semantics) and course learning outcomes

		Coefficient	*SE*	*t*	*df*	*p*
Vertex Matching	Mean Initial Status π_{0i}	6.89	0.89	7.75	23	<.001
	learning	3.32	1.59	2.08	23	0.048
	Mean Growth Rate π_{1i}	3.84	0.63	6.07	23	<.001
	learning	-0.36	0.81	-0.45	23	0.656
Propositional Matching	Mean Initial Status π_{0i}	0.0291	0.0082	3.52	23	0.002
	learning	0.0053	0.0111	0.48	23	0.635
	Mean Growth Rate π_{1i}	-0.0023	0.0023	-1.01	23	0.323
	learning	0.0011	0.0032	0.33	23	0.741

7

BETWEEN-DOMAIN DISTINGUISHING FEATURES OF COGNITIVE STRUCTURE 🕮

This research aims to identify domain-specific similarities and differences of externalized cognitive structures. Cognitive structure, also known as knowledge structure or structural knowledge, is conceived as the manner in which an individual organizes the relationships of concepts in memory. By diagnosing these structures precisely, even partially, the educator comes closer to influencing them through instructional settings and materials. The assessment and analysis of cognitive structures is realized within the HIMATT tool, which automatically generates four quantitative indicators for the structural entities of written text or causal maps. Participants worked on the subject domains biology, history, and mathematics. Results clearly indicate different structural and semantic features across the three subject domains.

🕮 This chapter is based on: Ifenthaler, D. (accepted). Identifying between-domain distinguishing features of cognitive structures. *Educational Technology Research and Development*.

Introduction

Knowledge representation is a key concept in psychological and educational diagnostics. Existing models for describing the fundamentals of knowledge representation are multifaceted. The distinction which has received the most critical attention is that between declarative ("knowing that") and procedural ("knowing how") forms of knowledge (see Anderson, 1983; Ryle, 1949). Closely associated with these concepts is the term *cognitive structure*, also known as knowledge structure or structural knowledge (Jonassen, et al., 1993). It refers to the manner in which an individual organizes the relationships between concepts in memory (Shavelson, 1972). Hence, an individual's cognitive structure is made up of the interrelationships between concepts or facts and procedural elements. Furthermore, it is argued that the order in which information is retrieved from long-term memory and externalized will reflect in part the individual's cognitive structure within and between concepts or domains (e.g., Strasser, 2010). Researchers and educators thus have immense interest in assessing and analyzing cognitive structures and comparing them with others in order to identify the most appropriate ways to facilitate learning and problem solving (Ifenthaler, et al., in press). By diagnosing cognitive structure precisely, or even partially, the educator can come closer to influencing it through instruction. It will help to organize materials, identify knowledge gaps as well as misconceptions, and relate new materials to existing slots or anchors within the learners' cognitive structures (Jonassen, 1987).

Characteristics of cognitive structures have been researched and described for various subject domains. The majority of this research is concerned with domains in the natural sciences, e.g., physics (Chi, Glaser, & Rees, 1982) and biology (Baird & White, 1982). Other empirical studies have focused on within-domain specific features and the learning-dependent development of cognitive structure (e.g., Clariana & Wallace, 2007; Ifenthaler, et al., in press; Koubek, et al., 1994). However, as interdisciplinary learning and teaching is becoming more important (e.g., Nikitina, 2005), a comprehensive understanding of cognitive structures across different subject domains is inevitable.

In this chapter, an empirical study in which similarities and differences in externalized cognitive structure across three domains is reported: biology, history, and mathematics. It is also intended to show an automated, reliable, and valid measurement technique that would make this identification possible.

Background

Researchers in the field of cognitive and developmental psychology have proposed a logic-based universal cognitive structure (e.g., Johnson-Laird & Byrne, 1991; Rips, 1994), and there is hardly any doubt that the concept of cognitive structure is applicable to every domain of knowledge (Jonassen, et al., 1993). In addition, educational researchers have described the characteristics of cognitive structure for different domains, e.g. physics (Chi, et al., 1982) or biology (Baird & White, 1982). During the 1980s and 1990s, educational and cognitive psychology focused on domain specificity within cognitive structure. The objective was to identify the meaning or impact of different knowledge structures for specific domains of knowledge. Ennis (1989, 1990) and McPeck (1990) debated on and described domain specificity in their discussion on critical thinking. As a result, three principles of domain specificity have been developed: (1) It needs prior knowledge, (2) it cannot be transferred to other domains without explicit instructions focusing on transfer, and (3) it cannot be deduced from general critical thinking instructions. These principles constitute the foundation for ongoing research on domain specificity of cognitive structure.

Based on the above-described assumptions, many studies published in the past decades have focused on domain-specific knowledge, prior knowledge, and the structure of knowledge in various fields, such as physics (Clement, 1981; Moeira, 1983), chemistry (Taber, 1995), science in general (Bliss, 1996; Watts, 1988), logic (Chase & Simon, 1973), and the social sciences (Voss, Greece, Post, & Penner, 1983). Other studies have indentified the development of cross-domain scientific reasoning processes (Kuhn, Schauble, & Garcia-Mila, 1992), complex mathematical problem solving (Vye, Goldman, Voss, Hmelo, & Williams, 1997), and text processing in history (Wolfe & Goldman, 2005). However, our extensive literature review shows that previous studies focused primarily on knowledge structures in specific domains.

Furthermore, many insights about the nature of cognitive structure in different domains are influenced by research on expertise (e.g., Chi, Feltovich, & Glaser, 1981). Here, the objective is to identify the essential differences in cognitive structures between novices and experts in a specific domain (Gruber & Ziegler, 1996). Some approaches see expertise as being caused by giftedness (Sternberg, 1993), others see it as a general, learnable phenomenon (Glaser, 1999). However, there is general agreement on the point that expertise is usually restricted to one domain (e.g., Gruber, 1994). This is mainly explained by the large amount of time a person needs to become an expert (Gruber, 1994). Empirical results show that a well-organized cognitive structure is an essential factor for expertise (Gruber, 1994). Moreover, experts recognize meaningful patterns and relevant information for a problem faster than novices and spend a lot more time representing the core problem. Another characteristic is fast information processing, which can be explained by multifaceted elaboration supported by experience (Gruber & Ziegler, 1996). Accordingly, these findings confirm the assumption that cognitive structure may be context bound.

In contrast, interdisciplinary learning and teaching is widely discussed and claimed (e.g., Holley, 2009; Woods, 2007). Still, we were not able to identify empirical studies that compared cognitive structure across different subject domains. Therefore, our current research goes beyond the focus on cognitive structure within a single domain. More specifically, we aim to identify similarities and differences in externalized cognitive structure between three distinct subject domains: biology, history, and mathematics. These three domains represent different types of domains. History is regarded as 'soft' domain that lacks a central body of theory (Biglan, 1973). On the other hand, mathematics is regarded as 'hard' domain with a central body of theory (Biglan, 1973). Biology can be classified in between the hard and soft domains. Additionally, these three subject domains were chosen due to their different instructional methods and because they are taught in nearly every grade. In the following sections we discuss unique features of these three domains and suggest possible cross-domain distinguishing features.

Biology

Biology, a natural science, is concerned with the study of life and includes interdisciplinary fields such as zoology, botany, physiology, medicine, and psychology (Nason & Goldstein, 1969). The scientific methods used in biology are multifaceted, including physical, mathematical, sociological, and psychological techniques. Empirical research on learning in biology has dealt with motivation, and interest as well as cognitive structure (Baalmann, 1997; Bayrhuber, 2001; Mintzes, Yen, & Barney, 2008). Findings show that wide generalization of facts in biology has negative effects on highly elaborated knowledge structure (Eschenhagen, Kattmann, & Rodi, 2008). Additionally, knowledge in biology has strong correlations with the specific attitudes and interests of learners (Trumper, 2006).

Thompson and Mintzes (2002) showed that affective learning and teaching objectives are very important for biology education. This is evident in the large amount of topics involving ethical issues, like sexuality, the natural environment, and health education (Eschenhagen, et al., 2008). Domain-dependent learning objectives include issues involving plants, animals, and human beings, e.g., the variety of ecosystems, changes in populations, ecological sequencing, and interactions between the climate and living organisms (Tamir & Jungwirth, 1972). Additionally, basic concepts and techniques of the natural sciences are also elements in biology education. Hence, biology instruction focuses on transferring already existing (preschool) prior and general knowledge to a scientifically correct hierarchical order and specifying it during the learning process. In summary, knowledge structure in biology can be characterized as hierarchical, well-connected, but not very fine-grained.

History

Methodologically speaking, history moved from pure descriptive historicism to a social science perspective to meet the requirements of modern society (Iggers, 1996). Empirical studies in history learning focus on the analysis of attitudes and affective dispositions towards specific events or people, e.g.: "Who is responsible for WWII?" (Hasberg, 2001). However, empirical research on cognitive structure in history is rare (von Borries, 2001). The few existing empirical investigations concerned with cognitive structure are limited to qualitative methods (Mirow, 1991; Pape, 2006). According to Mirow (1991), cognitive structure in history consists of unconnected *knowledge islands* developed from different sources. Moreover, there are fatal misinterpretations or misconceptions concerning the importance or the historical background of events (Donovan & Bransford, 2005; Mirow, 1991). For example, learners do not seem to be oriented towards canonized content; rather, they mobilize different content and situation-dependent memorizations, which leads to different *"histories"* (Rüsen, Fröhlich, Horstkötter, & Schmidt, 1991, p. 343).

From an instructional point of view, the overall learning and teaching objective of history is to cultivate a critical historical consciousness, e.g., a sense of time, a sense of reality, moral sense, and a sense of history and politics (Pandel, 1987). To sum up, cognitive structure in history can be characterized as linear, unconnected, and oversimplified.

Mathematics

Mathematics is one of the oldest sciences and is organized around many branches. Numbers, logic, geometry, algebra, and statistics are just a small part of the broad spectrum (Courant & Robbins, 2000). Mathematics is used as an ancillary science in nearly all other sciences. In contrast to biology and history, research in mathematics has long been focused on the cognitive structure of learners (de Corte, Greer, & Verschaffel, 1996). Findings concerning cognitive structure in mathematics have been discussed in research on psychology (Piaget, 1972) and artificial intelligence (J. Johnson, McKee, & Vella, 1994) as well as in other branches. They suggest that mathematical knowledge develops when coping with real world problems. These real world problems are abstracted to mathematical problems in a step-by-step process through assimilation and accommodation (Piaget, 1972). However, most of these studies analyze deterministic skills like counting (de Corte, et al., 1996). Empirical findings focusing on the cognitive structure of complex mathematical phenomena (e.g., differential and integral calculus) are not available. Overall, it is assumed that mathematical knowledge is strongly connected to a person's mathematical reality, i.e. personal perceptions and experiences (Kitcher, 1983). Cognitive structures for mathematics may be very complex and have rich connections. Additionally, hierarchical as well as linear principles play a fundamental role in mathematical thinking (de Corte, et al., 1996; Kleinert, 2005).

From an instructional point of view, learning and teaching objectives have been a cause for controversy due to the wide range of available instructional methods and for ideological reasons. Still, general learning and teaching objectives include, e.g., the application of mathematics to other fields, creativity, and rational argumentation (Winter, 1975). As in biology, visual demonstration and application to the real-life situations of learners are typical instructional methods. To sum up, cognitive structure in mathematics can be characterized as linear and hierarchical, well-connected, and very specific.

Cross-domain distinguishing features

The above-described theoretical and empirical assumptions of the three subject domains allow us to describe possible cross-domain distinguishing features: (1) Students' knowledge in the domain biology is well structured and is ordered in hierarchical fashion. However, general knowledge and specific details are not well developed. (2) Historical knowledge is characterized by separate knowledge islands, is less structured, and oversimplified. Additionally, it often includes misconceptions, e.g., historical events are dated incorrectly. (3) Mathematical knowledge tends to be very complex and rich in relations. Moreover, it has a strong hierarchical organization and is characterized by everyday mathematical experiences.

The clear structural organization in biology and mathematical knowledge leads us to the assumption that it might be significantly different from the fragmented knowledge in history. Hence, one might expect a more complex cognitive structure in biology and mathematics. However, biology includes less abstracted cognitive structure, whereas mathematics is characterized by more specified and complex cognitive structure.

Our research

Our research builds on the verdict that cognitive and educational researchers use theoretical constructs, e.g., mental models, schemata, etc., to explain complex cognitive structure and procedures for learning, reasoning, and problem solving (e.g., Gentner & Stevens, 1983; Johnson-Laird, 1983; Jonassen, et al., 1993; Lehrer & Romberg, 1996; Schauble, Klopfer, & Raghavan, 1991; Seel, et al., 2009; Snow, 1989, 1990). However, these internal cognitive structures and functions are not directly observable.

Accordingly, the assessment and analysis of internal cognitive structure and functions requires that they be externalized. Therefore, we argue that it is essential to identify economic, fast, reliable, and valid techniques to elicit and analyze these cognitive structures (see Ifenthaler, 2008, 2010d). Methodologies include standardized questionnaires and interviews, think-aloud protocols (e.g., Ericsson & Simon, 1993), the assessment of log files or click streams (e.g., Chung & Baker, 2003; Dummer & Ifenthaler, 2005), eye-tracking measures (e.g., Mikkilä-Erdmann, Penttinen, Anto, & Olkinuora, 2008), and Pathfinder networks (Durso & Coggins, 1990; Schvaneveldt, 1990), as well as mind tools (e.g., Jonassen & Cho, 2008;

Spector, et al., 2006). Accordingly, the possibilities for externalizing cognitive structure are limited to a few sets of sign and symbol systems (Seel, 1999b) – characterized as *graphical-* and *language-based approaches* (Ifenthaler, 2010d). A widely accepted application for the assessment and analysis of cognitive structure is a concept, causal, or knowledge map which can be automatically scored and compared to an expert's solution (Herl, et al., 1996; Spector, et al., 2006; Spector & Koszalka, 2004). On the other hand, there are convincing arguments indicating that natural language representations (e.g., written texts) are a good basis for assessing and analyzing cognitive structure (Ifenthaler & Pirnay-Dummer, 2009).

As not every available methodology is suitable for this research (e.g., lack of reliability and validity, too labor intensive, etc.), we utilize the web-based assessment and analysis platform HIMATT (Highly Integrated Model Assessment Technology and Tools; Pirnay-Dummer, et al., 2010).

HIMATT is a combined toolset which was developed to convey the benefits of various methodological approaches in a single environment and which can be used by researchers with only little prior training (Pirnay-Dummer & Ifenthaler, 2010). Methodologically, the tools integrated into HIMATT touch the boundaries between qualitative and quantitative research methods and build bridges between them. First of all, written text can be analyzed very quickly without loosening the associative strength of natural language. Furthermore, causal maps can be annotated by experts and compared to other solutions. The automated analysis function produces measures which range from surface-oriented structural comparisons (e.g., number of used concepts, complexity of representation) to integrated semantic (e.g., correctness of concepts or propositions) similarity measures. There are four *structural* (surface, graphical, structural, and gamma matching) and three *semantic* (concept, propositional, and balanced propositional matching) measures available (see the Method section for a detailed description of them). All of the data, regardless of how it is assessed, can be analyzed quantitatively using the same comparison functions without further manual effort or recoding.

The central research objective in this study is to identify cross-domain distinguishing features of externalized cognitive structures. First, we look at two specific sources of externalization of cognitive structure, written text and causal maps. We expect these different forms of externalization to represent the same structural and semantic content within each subject domain (Hypothesis 1). More specifically, due to the short time between writing texts and constructing causal maps, we expect a close match between the structural and semantic HIMATT measures (Pirnay-Dummer, et al., 2010; a description of all of the applied measures will be provided in the following section).

Secondly, previous empirical studies have focused on domain-specific features and the learning-dependent development of cognitive structure (e.g., Clariana & Wallace, 2007; Ifenthaler, et al., in press; Koubek, et al., 1994). However, an empirical analysis and comparison of the organization of cognitive structures across different domains has not been conducted so far. Accordingly, this study will identify similarities and differences in externalized cognitive structures between three different subject domains: biology, history, and mathematics. These three subject domains were chosen due to their different instructional methods and because they are taught in nearly every grade. Based on prior research (de Corte, et al., 1996; Kleinert, 2005; Mirow, 1991; Thompson & Mintzes, 2002), we hypothesize that the externalizations of the three subject domains have different structural features (Hypothesis 2.1). Additionally, we assume that the externalizations of biology knowledge are strongly organized in a hierarchy (Hypothesis 2.2), that the externalizations of mathematics knowledge are also strongly organized in hierarchical order (Hypothesis 2.3), and that the organization of externalizations of historical knowledge are less hierarchical (Hypothesis 2.4). We also assume that the externalizations in the history domain are less connected than those in biology and mathematics (Hypothesis 2.5). Last, on the basis of equal difficulty level of the learning material, we expect that the declarative knowledge (assessed with a domain-specific knowledge test) does not differ across the three domains (Hypothesis 2.6).

Finally, previous research studies on cognitive structure have found contradictory results concerning learners' cognitive abilities in association with learning outcomes (e.g., Hilbert & Renkl, 2008; Ifenthaler, et al., 2007; O'Donnell, Dansereau, & Hall, 2002). Hence, our final research question will contribute to this vague empirical basis. We assume that learners with higher mathematical abilities will outperform those with lower mathematical abilities with regard to their learning outcomes in the mathematics domains (Hypothesis 3.1). Additionally, we assume that verbal and spatial abilities will have no effect on learning outcomes in the three subject domains biology, history, and mathematics (Hypothesis 3.2).

Method

Participants

Seventy-one students (61 female and 10 male) from a European university participated in the study. Their average age was 22.2 years (SD = 2.3). They were all enrolled in an advanced course on diagnostics in schools and further education and had studied for an average of 2.5 semesters (SD = 2.1). The first language of 85% of the participants was German. 15% of the participants spoke German as their second language. None of the participants were specially trained in the three subject domains biology, history, or mathematics.

Materials

The materials consisted of three domain-specific articles for the domains biology, history, and mathematics. Additional materials included knowledge tests for each domain, a test for experience with causal maps, three subscales of an intelligence test, and tools for eliciting the participants' understanding of the phenomenon in question.

Domain-specific articles

Selection of the three domain-specific articles was based on (a) an equal difficulty level, (b) a similar text length, and (c) the integration into the high school curriculum. A German-language article on the *human brain* with 546 words was used as the first learning material for the biology domain. A German-language article on the *European boarders* with 720 words was used for the history domain. For the mathematics domain, a German-language article on the *statistical procedures of the t-test* with 500 words was used.

Domain-specific knowledge tests

Each knowledge test (biology, history, mathematics) included 10 multiple-choice questions with four possible solutions each (1 correct, 3 incorrect). They were developed on the basis of the domain-specific articles. In a pilot study (N = 5 participants, independent from the participants of the main study), we tested the average difficulty level to account for ceiling effects. All participants had low prior knowledge in the three domains. They scored M = 3.2 correct answers (SD = 1.2) on the biology test, M = 3.4 correct answers (SD = 1.7) on the history test, and M = 2.1 correct answers (SD = .9) on the mathematics test. In our experiment we administered two equivalent versions (in which the 10 multiple-choice questions appeared in a different order) of the domain-specific knowledge tests (pre- and posttest). Participants did not receive feedback on the scores or on the correctness of their answers for the pre- and posttest. It took about five minutes to complete each test.

Experience with causal maps test

The participants' experience with causal maps was tested with a questionnaire including eight items (Ifenthaler, 2009; Cronbach's alpha = .87). The questions were answered on a five-point Likert scale (1 = totally disagree; 2 = disagree; 3 = partially agree; 4 = agree; 5 = totally agree), e.g., "I used causal maps to structure learning content", "The construction of causal maps is easy." (translated from German).

Mathematical, spatial, and verbal abilities

Three subscales of the I-S-T 2000 R (Amthauer, Brocke, Liepmann, & Beauducel, 2001) were used to test the participants' mathematical, spatial, and verbal abilities. This test is a widely used intelligence test in Germany with high reliability (r = .88 to r = .96; split-half reliability).

The first subscale was used to test the participants' *mathematical abilities*. A total of 20 arithmetic problems (+, -, *, /) had to be completed. Participants had ten minutes to complete this subscale. The second subscale tested *spatial abilities*. The participants had nine minutes to choose similar cubes from a set of five by rotating them. Subset two included 20 cube problems. The third subscale we used tested *verbal abilities*. A total of 20 sentences with a missing word had to be completed using a set of five words. The participants had six minutes to complete this subset.

HIMATT causal maps and text input tools

The *causal maps tool*, which is part of the HIMATT (Pirnay-Dummer, et al., 2010) environment, was used to assess the participants' understanding of the domain-specific phenomenon in question. The intuitive web-based tool allows participants to create causal maps with only little training (Pirnay-Dummer & Ifenthaler, 2010). Once created, all causal maps are automatically stored on the HIMATT database for further analysis. The HIMATT *text input tool* was also used to assess the participants' understanding of the domain-specific learning content. Participants' written texts are automatically parsed and stored on the HIMATT database for further analysis. Written and on-screen instructions in form of questions were provided for each subject domain.

Procedure

First, the participants completed a *demographic data questionnaire* and the *experience with causal maps test*. Secondly, they completed the test on *verbal, mathematical, and spatial abilities*. Next, the participants were given an introduction to causal maps and were shown how to use the HIMATT software. After a short relaxation phase, they completed the *domain-specific knowledge test* on history. Then they received the text on *European borders*. The participants had 15 minutes to read the text. Then they logged in to the HIMATT system, where they constructed a causal map on their understanding of European borders (ten minutes). Immediately afterwards, they wrote a text about their understanding of European borders (ten minutes). After another short relaxation phase, the procedure was repeated with the domains mathematics and biology (1. domain specific knowledge test, 2. reading of text, 3. construction of a causal map, 4. writing of text). In total, the experiment took approximately two hours.

Data analysis

During our experiment, the participants used the web-based platform HIMATT to externalize their understanding of the three subject domains in the form of a causal map and a written text. The automatically stored data were analyzed using the HIMATT analysis function (see Pirnay-Dummer, et al., 2010). Additionally, we used a qualitative scoring rubric to classify the hierarchical structure of the graphical externalizations.

HIMATT

In order to analyze the participants' understanding of the phenomena in question (biology, history, mathematics), we used the seven measures implemented in HIMATT (see Table 7.1; Ifenthaler, 2010d; Pirnay-Dummer, et al., 2010).

Both written texts and causal maps were analyzed using the seven HIMATT measures. Before the written text can be analyzed, a parsing algorithm must be applied. The written text is tokenized, tagged, and stemmed, and the most frequent concepts and pairwise associations between concepts are determined (Pirnay-Dummer & Ifenthaler, 2010). Accordingly, concepts from the written text are stored pairwise on the HIMATT database along with the strength of association. Additionally, the causal maps are stored on the HIMATT database directly. .

Each of the participants' written texts and causal maps can be compared automatically against each other, across domains, or against a reference map (e.g., an expert representation). The automated analysis generates seven measures of HIMATT (see Table 7.1). They include four structural and three semantic measures (Ifenthaler, 2010c, 2010d; Pirnay-Dummer & Ifenthaler, 2010; Pirnay-Dummer, et al., 2010).

TABLE 7.1
Description of the seven HIMATT measures

Measure [abbreviation] and type	*Short description*
Surface matching [SFM] *Structural indicator*	The surface matching (Ifenthaler, 2010c) compares the number of vertices within two graphs. It is a simple and easy way to calculate values for surface complexity.
Graphical matching [GRM] *Structural indicator*	The graphical matching (Ifenthaler, 2010c) compares the diameters of the spanning trees of the graphs, which is an indicator for the range of conceptual knowledge. It corresponds to structural matching as it is also a measure for structural complexity only.
Structural matching [STM] *Structural indicator*	The structural matching (Pirnay-Dummer & Ifenthaler, 2010) compares the complete structures of two graphs without regard to their content. This measure is necessary for all hypotheses which make assumptions about general features of structure (e.g. assumptions which state that expert knowledge is structured differently from novice knowledge).
Gamma matching [GAM] *Structural indicator*	The gamma or density of vertices (Pirnay-Dummer & Ifenthaler, 2010) describes the quotient of terms per vertex within a graph. Since both graphs which connect every term with each other term (everything with everything) and graphs which only connect pairs of terms can be considered weak models, a medium density is expected for most good working models.
Concept matching [CCM] *Semantic indicator*	Concept matching (Pirnay-Dummer & Ifenthaler, 2010) compares the sets of concepts (vertices) within a graph to determine the use of terms. This measure is especially important for different groups which operate in the same domain (e.g. use the same textbook). It determines differences in language use between the models.
Propositional matching [PPM] *Semantic indicator*	The propositional matching (Ifenthaler, 2010c) value compares only fully identical propositions between two graphs. It is a good measure for quantifying semantic similarity between two graphs.
Balanced propositional matching [BPM] *Semantic indicator*	The balanced propositional matching (Pirnay-Dummer & Ifenthaler, 2010) is the quotient of propositional matching and concept matching. Especially when both indices are being interpreted, balanced propositional matching should be preferred over propositional matching.

HIMATT uses specific automated comparison algorithms to calculate similarities between a given pair of frequencies f_1 (e.g., expert solution) and f_2 (e.g., participant solution), which results in a measure of $0 \leq s \leq 1$, where $s = 0$ is complete exclusion and $s = 1$ is identity. The other measures collect sets of properties using the Tversky similarity (Tversky, 1977). The Tversky similarity also results in a measure of $0 \leq s \leq 1$, where $s = 0$ is complete exclusion and $s = 1$ is identity. Please refer to Prinay-Dummer and Ifenthaler (2010) for a detailed discussion of the comparison algorithms.

Every single measure integrated into HIMATT are tested for reliability. The reliability scores range from r = .79 to r = .94 and are tested for the structural and semantic measures separately and across different knowledge domains (Pirnay-Dummer, et al., 2010). Validity scores are also reported separately for the structural and semantic measures. Convergent validity lies between r = .71 and r = .91 for semantic comparison measures and between r = .48 and r = .79 for structural comparison measures (see Pirnay-Dummer, et al., 2010).

Structural classification

Qualitative classification of the structure of the causal maps was based on the four categories introduced by Ku (2007): (1) hierarchy map, (2) spider map, (3) flowchart map, (4) system map. For each subject domain (biology, history, mathematics), we generated standardized graphical outputs using the HIMATT platform (see Figure 7.1).

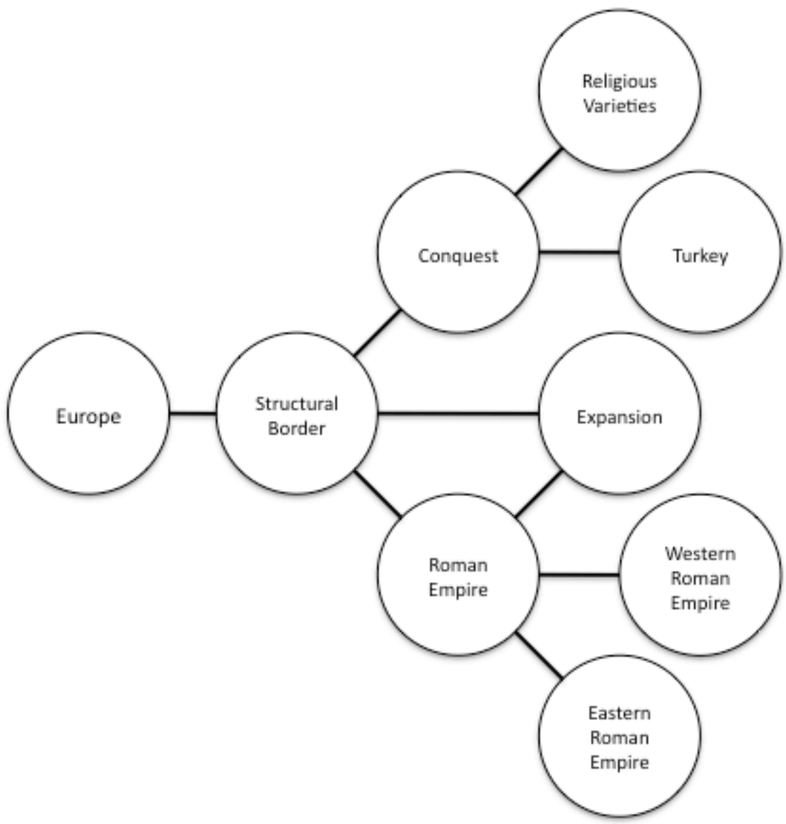

FIGURE 7.1. *Standardized graphical output of the domain history (hierarchical structure)*

All standardized graphical outputs (causal maps; N = 213) were coded using the above-described categories (1 = hierarchy structure; 2 = spider structure; 3 = flowchart structure; 4 = system structure; 5 = other structure). Each coder received a printed set of the standardized graphical outputs (including a subject and domain code; $N = 213$) and a coding sheet, where they had to enter the subject and domain code and in which of the five categories it belonged to. Three independent researchers found an average interrater reliability of $\kappa = .85$ (Fleiss' kappa; Fleiss, 1971).

Results

Initial data checks showed that the distributions of ratings and scores satisfied our assumptions concerning the analysis procedures. All effects were assessed at the .05 level. As effect size measures, we used Cohen's d (small effect: $d < .50$, medium effect $.50 \leq d \leq .80$, strong effect $d > .80$) and partial η^2 (small effect: $\eta^2 < .06$, medium effect $.06 \leq \eta^2 \leq .13$, strong effect $\eta^2 > .13$).

More than two-thirds of the participants (77%) did not use causal maps to structure their own learning materials before our experiment. Only 19% used software to create their own causal maps beforehand. 45% of the participants answered that they did not find it difficult to create a causal map, 55% had difficulties in creating causal maps.

On each domain-specific knowledge test (biology, history, mathematics), participants could score a maximum of 10 correct answers. ANOVA was used to test for differences among the three subject domains (Hypothesis 2.6). The correct answers differed significantly across the three subject domains, $F(2, 210) = 5.51$, $p = .005$, $\eta^2 = .05$. Tukey HSD post-hoc comparisons of the three subject domains indicate that participants had significantly better scores on the biology test ($M = 5.01$, $SD = 1.69$, *95% CI* [4.62, 5.41]) than on the history test ($M = 3.93$, $SD = 1.78$, *95% CI* [3.51, 4.35]), $p = .003$. Comparisons between the correct answers on the mathematics test ($M = 4.34$; $SD = 2.37$) and the biology and history tests were not statistically significant at $p < .05$.

Written text and causal maps

For all three subject domains (biology, history, mathematics), the written texts and causal maps constructed by the participants were automatically compared to domain-specific expert representations by the HIMATT analysis feature (see Table 7.1). Hence, for both written texts and causal maps, seven similarity scores (0 = no similarity; 1 = total similarity; for the measures surface, graphical, structural, gamma, concept, propositional, and balanced propositional matching) were available for further statistical analysis. In order to identify possible expert-novice differences between written text and causal maps, we computed paired-sample t-tests for the seven HIMATT similarity scores between experts' and participants' representations for the three subject domains. (see Table 7.2).

Table 7.2
HIMATT similarity scores (standard deviations in parentheses) between causal maps, texts and expert representations for the three subject domains

HIMATT similarity measure	*Subject domain*					
	Biology		*History*		*Mathematics*	
	Causal map	*Text*	*Causal map*	*Text*	*Causal map*	*Text*
Surface matching [SFM]	.527 (.298)	.474 (.262)	.314 (.234)	.304 (.246)	.460 (.234)	.434 (.233)
Graphical matching [GRM]	.639 (.244)	.522 (.184)	.461 (.261)	.538 (.271)	.597 (.231)	.670 (.230)
Structural matching [STM]	.659 (.210)	.681 (.168)	.551 (.171)	.501 (.186)	.576 (.153)	.489 (.167)
Gamma matching [GAM]	.682 (.244)	.730 (.286)	.547 (.187)	.518 (.246)	.601 (.181)	.448 (.227)
Concept matching [CCM]	.324 (.131)	.052 (.079)	.078 (.105)	.141 (.083)	.064 (.078)	.097 (.081)
Propositional matching [PPM]	.023 (.052)	.007 (.030)	.008 (.021)	.018 (.029)	.005 (.020)	.012 (.026)
Balanced propositional matching [BPM]	.062 (.133)	.032 (.112)	.034 (.089)	.088 (.136)	.023 (.082)	.058 (.115)

Note. HIMATT similarity measures, 0 = no similarity; 1 = total similarity; SFM, GRM, STM, and GAM are structural measures; CCM, PPM, and BPM are semantic measures

Interestingly, written text and causal maps seem to represent different structures and content across the three subject domains when compared to an expert's representation. In the biology domain, the participants' causal maps were significantly more similar to the expert's representation than their written texts were with regard to the graphical matching (GRM) measure, $t(70) = 3.25$, $p = .002$, $d = .54$. Additionally, we found higher similarities between the participants' causal maps and expert representations for the semantic HIMATT measures CCM, $t(70) = 16.14$, $p < .001$, $d = 2.51$, and PPM, $t(70) = 2.27$, $p = .026$, $d = .38$. In the history domain, analysis revealed significant differences for the semantic HIMATT measures. Here, the written texts of the participants were more similar to the expert's representation with regard to CCM, $t(67) = 3.41$, $p = .001$, $d = .67$, PPM, $t(67) = 2.27$, $p = .026$, $d =$

.39, and BPM, $t(67) = 2.52$, $p = .014$, $d = .47$. In the mathematics domain, the participants' written texts were significantly more similar to the expert's representation than their causal maps were with regard to the GRM measure, $t(67) = 1.99$, $p = .050$, $d = .32$. On the other hand, the participants' causal maps were significantly more similar to the expert's representation than their written texts were with regard to the STM measure, $t(67) = 3.09$, $p = .003$, $d = .54$, and the GAM measure, $t(67) = 4.62$, $p < .001$, $d = .75$. Additionally, we found higher similarities between the participants' written texts and expert representations for the semantic HIMATT measure CCM, $t(67) = 2.24$, $p < .028$, $d = .42$.

Therefore, we had to reject Hypothesis 1. The causal maps and text did not represent the same structural and semantic content within the three subject domains.

Cross-domain distinguishing features

In order to identify the hypothesized cross-domain distinguishing features, we computed a MANOVA with the seven descriptive HIMATT measures (SFM, GRM, STM, GAM, CCM, PPM, BPM) as within-subject factors (see Table 7.3). The following between-subject factors were applied for the seven separate analyses: 1. *Subject domain* (biology, history, mathematics); 2. *Elicitation method* (causal map, written text).

MANOVA showed a significant main effect of the subject domain on the descriptive HIMATT measures, Wilks' Lambda = .749, $F(14, 814) = 9.048$, $p < .001$, $\eta^2 = .135$. Univariate ANOVA's revealed that the effect was caused by the dependent variables SFM, $F(2, 413) = 5.561$, $p = .004$, $\eta^2 = .026$, GRM, $F(2, 413) = 7.983$, $p < .001$, $\eta^2 = .037$, STM, $F(2, 413) = 12.420$, $p < .001$, $\eta^2 = .057$, GAM, $F(2, 413) = 11.075$, $p < .001$, $\eta^2 = .051$, and CCM, $F(2, 413) = 17.634$, $p < .001$, $\eta^2 = .079$. Post-hoc comparisons using Tukey's HSD revealed that the re-representations in the biology domain contained a larger surface (SFM) than did those in the history ($p = .007$) and mathematics ($p = .022$) domains. Additionally, the re-representations in the history domain were less complex (GRM) than those in the biology ($p = .001$) and mathematics ($p = .004$) domains. The complete structure (STM) of the re-representations was larger in the biology domain than in the history ($p < .001$) and mathematics ($p = .001$) domains. The connectedness (GAM) of the re-representations in the biology ($p = .002$) and history ($p < .001$) domains was higher than in the mathematics domain. Finally, the number of semantically correct concepts in the biology domain was higher than in the history ($p = .022$) and mathematics ($p < .001$)

domains. Additionally, the number of semantically correct concepts in the history domain was higher than in the mathematics (p = .003) domain.

Table 7.3
HIMATT descriptive measures (standard deviations in parentheses) of participants' causal maps and written texts for the three subject domains

HIMATT descriptive measure	*Subject domain*					
	Biology		*History*		*Mathematics*	
	Causal map	*Text*	*Causal map*	*Text*	*Causal map*	*Text*
Surface matching [SFM]	13.704 (4.086)	24.409 (32.656)	9.294 (3.516)	16.543 (19.880)	10.268 (3.517)	17.471 (11.742)
Graphical matching [GRM]	5.592 (1.769)	4.296 (3.240)	4.368 (1.789)	3.429 (2.801)	5.070 (1.799)	4.500 (2.282)
Structural matching [STM]	13.831 (3.676)	11.803 (9.746)	9.324 (2.985)	9.429 (7.866)	9.972 (3.052)	10.677 (4.952)
Gamma matching [GAM]	.468 (.080)	.469 (.329)	.457 (.130)	.537 (.376)	.429 (.106)	.312 (.216)
Concept matching [CCM]	2.225 (2.349)	2.127 (1.971)	1.206 (1.356)	2.086 (1.726)	.563 (.788)	1.466 (1.165)
Propositional matching [PPM]	.127 (.375)	.296 (.595)	.132 (.420)	.500 (.737)	.056 (.232)	.368 (.710)
Balanced propositional matching [BPM]	.026 (.076)	.091 (.179)	.042 (.139)	.154 (.220)	.026 (.108)	.123 (.230)

Note. SFM, GRM, STM, and GAM are structural measures; CCM, PPM, and BPM are semantic measures (compared to the domain specific expert representation)

In addition, MANOVA revealed a significant main effect of the elicitation method on the descriptive HIMATT measures, Wilks' Lambda = .667, $F(7, 407) = 29.073$, $p < .001$, $\eta^2 = .333$. Univariate ANOVA's revealed that the effect was caused by the dependent variables SFM, $F(1, 413) = 26.669$, $p < .001$, $\eta^2 = .061$, GRM, $F(1, 413) = 16.552$, $p < .001$, $\eta^2 = .039$, CCM, $F(1, 413) = 12.006$, $p = .001$, $\eta^2 = .028$, and PPM, $F(1, 413) = 1.251$, $p = .016$, $\eta^2 = .020$. Written texts (M = 19.47, SD = 1.15) had a larger surface (SFM) than causal maps (M = 11.09, SD = 1.15). Additionally, the written texts contained more semantically correct concepts and propositions (M = 1.89, SD = .11 for CCM, and M = .39, SD = .04 for PPM) than the causal maps (M = 1.33, SD = .11, and M = .11, SD = .04, respectively).

Finally, MANOVA revealed a significant interaction effect of the subject domain and elicitation method on the descriptive HIMATT measures, Wilks' Lambda = .888, $F(14, 814) = 3.562$, $p < .001$, $\eta^2 = .058$. According to univariate

ANOVA's this effect was caused by the dependent variables GAM, $F(2, 413) = 6.139$, $p = .002$, $\eta^2 = .029$, and CCM, $F(2, 413) = 4.192$, $p = .016$, $\eta^2 = .020$. Figure 2a shows the interaction effect on GAM. The connectedness of the re-representation in the history domain is higher for causal maps than for written texts. In contrast, the connectedness of the re-representations in the mathematics domain is higher for written texts than for causal maps. Figure 7.2b shows the interaction effect on CCM. Accordingly, the number of semantically correct concepts is higher for written texts than for causal maps in the subject domains history and mathematics.

Therefore, we accept Hypothesis 2.1. Externalizations of the three subject domains have different structural features.

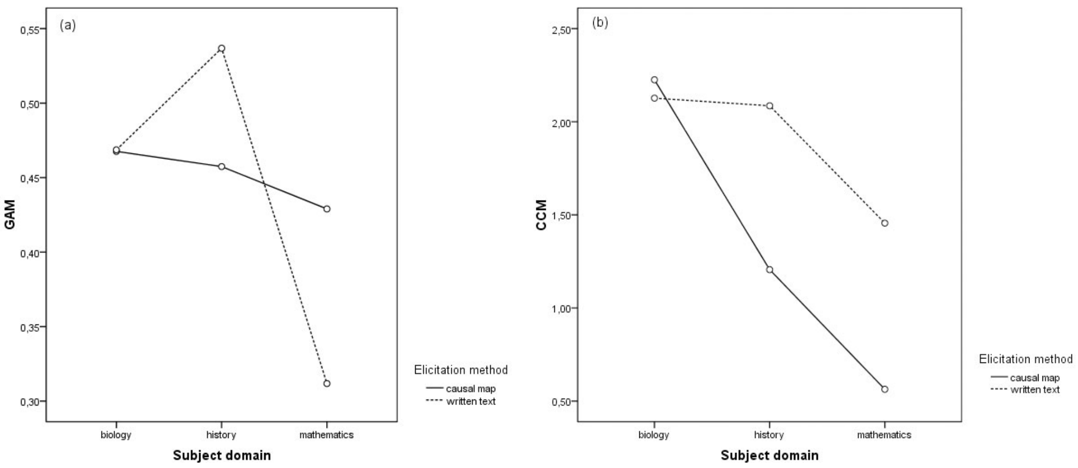

FIGURE 7.2. *Interactions of subject domain x elicitation method on the descriptive HIMATT measures GAM (part A) and CCM (part B)*

Furthermore, a 5 x 3 (structural classification by subject domain) chi-square test was conducted to assess whether the structural classification (hierarchy, spider, flowchart, system, other) is different in the three subject domains (biology, history, mathematics). The results of the chi-square test were significant, χ^2 (8, $N = 71$) = 61.29, $p = < .001$. Additionally, detailed analysis of standardized residuals was conducted in order to find out which structural classifications of the causal maps revealed significant differences (see Table 7.4). The hierarchical structure was the most frequent classification within the domains history and mathematics. In contrast, the spider structure was the most frequent classification in the biology domain. In the biology domain, the proportion of spider structure was much greater than hypothesized, while the proportion of hierarchy structure was lower than hypothesized. In the history domain, the proportion of spider structure was lower than hypothesized. In the mathematics domain, the proportion of hierarchical structure was greater than hypothesized, while the proportion of spider structure was lower than hypothesized.

Therefore, we had to reject Hypothesis 2.2. The causal maps of the biology domain were less organized in hierarchical order than expected. However, we accept Hypothesis 2.3, as the causal maps of the mathematics domain were organized in a strongly hierarchical order. Furthermore, we had to reject Hypothesis 2.4, as the causal maps of the history domain were more strongly hierarchical in structure than expected.

Table 7.4
Frequency (% in parentheses) and standardized residuals of subject domain by structural classification

Structural classification	Subject domain					
	Biology	Standard residual	History	Standard residual	Mathematics	Standard residual
Hierarchy	21 (29.6 %)	- 3.1	47 (66.2 %)	.9	55 (77.5 %)	2.2
Spider	40 (56.3 %)	5.3	8 (11.3 %)	- 2.3	5 (7 %)	-3.0
Flowchart	0 (0 %)	- .8	1 (1.4 %)	.4	1 (1.4 %)	.4
System	0 (0 %)	- .8	1 (1.4 %)	.4	1 (1.4 %)	.4
Other	10 (14.1%)	- .3	14 (19.7 %)	.9	9 (12.7 %)	- .6

Note. Standardized residuals equal to or higher than |1.96| indicate significant differences.

Cognitive abilities

Participants could score a maximum of 20 points on the three subscales of the I-S-T 2000 R on mathematical, spatial, and verbal abilities. On the test for mathematical abilities the participants scored M = 10.46 points (SD = 4.03), on the test for spatial abilities they scored M = 10.65 points (SD = 3.10), and on the test for verbal abilities they scored M = 12.87 points (SD = 3.70). An analysis using Pearson's correlation coefficient was performed to identify correlations between the participants' cognitive abilities (mathematical, spatial, verbal), prior domain knowledge (biology, history, mathematics), and the HIMATT similarity measures. Analysis revealed the following correlations: Mathematical abilities and SFM (written texts) in the history domain, $r(69) = -.30$, $p = .013$; spatial abilities and PPM (causal maps) in the biology domain, $r(71) = .23$, $p = .05$; spatial abilities and GAM (written texts) in the history domain, $r(69) = .28$, $p = .02$; verbal abilities and prior knowledge in the history domain, $r(71) = .37$, $p = .001$; verbal abilities and SFM (causal maps) in the mathematics domain, $r(70) = .30$, $p = .013$; verbal abilities and SFM (written texts) in the mathematics domain, $r(69) = -.29$, $p = .016$; verbal abilities and GAM (causal maps) in the mathematics domain, $r(70) = .30$, $p = .014$; verbal abilities and GAM (written texts) in the mathematics domain, $r(69) = -.24$, $p = .044$; verbal abilities and CCM (written texts) in the mathematics domain, $r(69) = -.39$, $p = .001$; verbal abilities and BPM (written texts) in the mathematics domain, $r(69) = -.31$, $p = .01$.

Therefore, our findings do not support Hypothesis 3.1. Mathematical abilities had no systematic effect on the externalized cognitive structures in the mathematics domain. Additionally, our findings do not completely support Hypothesis 3.2, as we found only non-systematic correlations between the HIMATT similarity measures and cognitive abilities across the subject domains.

General discussion

The aim of our study was to identify cross-domain distinguishing features of cognitive structures. Our experimental design included tasks in three different subject domains: biology, history, and mathematics. Participants were asked to externalize their understanding of the phenomenon in question in the form of causal maps and written texts. The participants' re-representations (causal maps and written texts) were automatically analyzed with the HIMATT analysis features. Accordingly, not only do these automated process have very high objectivity, reliability, and validity (Pirnay-Dummer, et al., 2010), they are also very economical, especially when large data sets need to be analyzed within a short period of time (Ifenthaler, 2010c).

First, we compared the causal maps and written texts to domain-specific expert representations. Due to the short time between the construction of the causal maps and written texts, we expected a close match between the structural and semantic features of the participants' re-representations. However, we found that the written text and concept maps seem to represent different structure and content across the three subject domains when compared to an expert's representation. Participants' causal maps in the biology domain showed higher similarity to the expert representation than the written texts with regard to complexity and semantically correct concepts as well as propositions. In contrast, participants' written texts showed higher similarities to the expert representation than the causal maps with regard to complexity (mathematics domain) and semantically correct concepts (history and mathematics domain). Hence, the type of externalization strategy also influences the knowledge which is represented (structurally and semantically). These findings suggest that instructional approaches, grading, and feedback is highly dependent on the externalization strategy used by learners. Consequently, more empirical research is needed to provide a valid framework for suitable domain-dependent externalization strategies.

Based on these initial findings, we then investigated cross-domain distinguishing features of the participants' re-representations across the subject domains biology, history, and mathematics. As expected, the results of our HIMATT analysis clearly indicate different structural and semantic features across the three subject domains. For example, participants were able to externalize larger cognitive structure (i.e. more concepts and relations) in the biology domain. Furthermore, the externalizations in the history domain were less complex than those in the biology and mathematics domains. Additionally, externalized cognitive structure in the biology domain was more integrated than in the other two domains. As far as semantically correct concepts are concerned, the externalizations in the biology domain included more correct terms than the other two domains. On the other hand, analysis revealed that cognitive structure externalized as written texts had a larger surface and contained more semantically correct concepts than causal maps.

Additionally, the structural classification by subject domain of the externalized cognitive structure revealed that hierarchical structure was the most frequent classification in the history and mathematics domains. In contrast, we found that externalizations in the biology domain were for the most part classified as spider structures.

Furthermore, we looked at the influence of mathematical, spatial, and verbal abilities on the learning outcomes. On the basis of previous studies (Hilbert & Renkl, 2008; Ifenthaler, et al., 2007), we expected no correlation between cognitive abilities and learning outcomes. Indeed, we did not find systematic influences of cognitive abilities on learning outcomes. However, some results suggest that cognitive abilities might have some influence. Accordingly, we recommend for future experimental studies to concentrate on the influence of cognitive abilities on cognitive structure during learning processes.

Instructional implications

Our results indicate that cognitive structures are organized in different ways depending on the subject domain (Johnson-Laird, 1989). Accordingly, identifying the learner's cognitive structure will help to organize instructional materials, discover knowledge gaps, and relate new materials to existing slots or anchors within the learner's cognitive structure (Jonassen, 1987). Hence, the classification of cognitive structure can act as a "topographical map" for identifying key areas of learning difficulties and facilitating instructional interventions (Ifenthaler, et al., in press; Snow, 1989). This might lead to the design of new learning materials which consider the unique features of specific subject domains and their related cognitive structure. Further it might help to design effective feedback methods to facilitate individual learning in a more effective and personalized way (Ifenthaler, 2009; Shute, 2008).

In addition, as the applied elicitation techniques seem to be highly domain-specific, validating results using outside criteria seems unavoidable. These findings may have a major impact on future research and knowledge diagnosis. We strongly suggest investigating these initial findings further in future experimental studies (e.g., Ifenthaler & Pirnay-Dummer, 2009).

To sum up, the findings of our study suggest that a diagnostics of learner's external representations always requires different elicitation techniques, e.g., written texts, verbal communication, or graphical drawings (de Vries, 2006). Clearly, a cognitive structure is internal to the mind, and for obvious reasons not directly observable (Seel, 1999a). Such representations are widely viewed as having a language-like syntax, and a compositional semantic (Spector, 2010; Strasser, 2010). A mental model is a representation of a thing, ideas or more generally, an ideational framework. It relies on language and uses symbolic pieces and processes of knowledge to construct a heuristic for a situation, which is instantiated by the world, or an internal process resembling the world, e.g., a mental simulation (Johnson-Laird, 1983; Schnotz & Bannert, 2003). Its purpose is heuristic reasoning, which leads either to intention, planning, behavior, or to a reconstruction of cognitive processes (Piaget, 1976). The facilitation of model-building processes may lead to enhanced problem-solving strategies and better transfers to near and far subject domains (Anzai & Yokoyama, 1984; Gick & Holyoak, 1980; Ifenthaler, et al., 2007).

Limitations and future research directions

Despite the promising results of this study, some critical remarks are in order. First, our results are limited to three very specific topics within the subject domains biology, history, and mathematics. Since cognitive structure seem to be highly domain dependent, we might also expect contradictory results within a single subject domain. Secondly, to gain more insight into the functions of cognitive structure and their domain-distinguishing features, a comparison across three subject domains is not sufficient by far. We thus suggest expanding our research question to other subject domains and including some topics which are closely related and others which are very different. An advanced research design of this kind would enable us to validate the findings of this initial study. Additionally, we recommend for researchers to reflect on possible elicitation techniques critically when investigating cognitive structure and knowledge in general. Further, in order to validate the structural and semantic measures of HIMATT, we recommend additional validation studies using outside criterions like the categories introduced by Ku (2007). However, in order to gain acceptable validation results, such an outside criterion needs to exactly match the HIMATT measures.

In summary, further studies will be needed to investigate the influence of externalization methodologies on learning and instruction. Also, additional studies concerning domain-distinguishing features are needed across and within various subject domains. This will give us more detailed insight into the functions of cognitive structure and help us to design more effective learning environments and apply more precise diagnosis strategies. The design and development of instruction is not only a matter of the applied methods and technologies; it is also highly dependent on the subject domain and last but not least on the cognitive structure learners already have developed prior to newly implemented instruction.

8

A LONGITUDINAL PERSPECTIVE 🕮

Cognitive scientists have studied internal cognitive structures, processes, and systems for decades in order to understand how they function in human learning. Nevertheless, questions concerning the diagnosis of changes in these cognitive structures while solving logical problems are still being scrutinized. This chapter reports findings from an experimental study in which 73 participants in three experimental groups solved logical word problems at ten measurement points. Changes of cognitive structures are illuminated and significant differences between the treatments are reported. The results also indicate that supportive information is an important aid for developing cognitive structures while solving logical problems.

🕮 This chapter is based on: Ifenthaler, D., & Seel, N. M. (in press). A longitudinal perspective on inductive reasoning tasks. Illuminating the probability of change. *Learning and Instruction*. doi: 10.1016/j.learninstruc.2010.08.004

Introduction

Learning, discussed in terms of constructivist theories, occurs when learners actively construct meaningful mental representations closely related to presented information. In general, a distinction is made between several forms of mental representations such as concepts, images, schemata, and mental models. As a result of the so-called cognitive revolution in cognitive psychology, schemata and mental models emerged as central theoretical constructs which have enriched the psychological knowledge about information processing, logical reasoning, and problem solving (Gick & Holyoak, 1980; Rumelhart, 1980; Rumelhart, Smolensky, McClelland, & Hinton, 1986). The idea that human cognition operates with mental models in thinking and reasoning can be traced back to "picture theories" of British empiricists of the 17th and 18th centuries, and can also be found in epistemology and psychology of the first half of the 20th century as Wittgenstein's (1922) picture theory in his *Tractatus* as well as Craik's (1943) epistemology of the nature of explanation demonstrate. Mental models returned as a powerful theoretical construct when Johnson-Laird (1983) as well as Gentner and Stevens (1983) published their works in the same year. Since then, study after study demonstrates that human reasoning exhibits particular features predicted by mental models which, therefore, emerged as important concept of logical reasoning and of creating plausibility in subject matter learning in various academic disciplines (e.g., Bonatti, 1994a, 1994b; Kalyuga, 2006c; Magnani & Nersessian, 2002; Rasch & Schnotz, 2009; Rumelhart, et al., 1986; Schaeken, Vandierendonck, Schroyens, d'Ydewalle, & Klauer, 2006; Schnotz & Bannert, 2003; Seel, 1991, 2003).

However, the construction of mental models presupposes semantic knowledge which is organized as schemata. Cognitive schemata can be conceived as the building blocks of mental models. As a consequence, some cognitive scientists argue that reasoning is regularly performed by means of pragmatic reasoning schemas (e.g., Cheng & Holyoak, 1985). Advocates of *schema-based reasoning* argue that generalizable knowledge is "stored" in reasoning schemas which contain the records of single cases of past successful reasoning and problem solving. Thus, schema-based reasoning extends the idea of case-based reasoning by referring to generalized "cases" (= schemata) rather than single cases and thus relies on the effective use of generic contextual knowledge to be transferred onto a current problem (Turner, 1994).

In our research we operate with a cognitive architecture which integrates both kinds of reasoning into a comprehensive framework which operates on the learning-dependent progression of mental models and their transition to (pragmatic reasoning) schemata. Thus, the present study was conducted to explore solution strategies of inductive reasoning tasks at ten measurement points.

Cognitive architecture of reasoning

A central assumption of cognitive psychology is that mental representations enable individuals to understand and explain experience and events, process information, and solve problems (Johnson-Laird, 1989). More specifically, Rumelhart et al. (1986) argue that these internal functions of the human mind are dependent on two interacting modules or sets of units: (1) schemata and (2) mental models. The resulting cognitive architecture corresponds to a great extent with Piaget's epistemology (Piaget, 1943, 1976) and its basic mechanisms of assimilation and accommodation.

Clearly, assimilation is dependent on the availability and activation of schemata, which allow new information to be integrated immediately into pre-existing cognitive structures. As soon as a schema can be activated, it runs automatically and regulates information processing in a "top down" manner. This allows information to be processed very quickly, a function which is vital for humans as it enables them to adapt to their environment spontaneously. If a schema does not fit immediately with the requirements of a new task it can be adjusted to meet them by means of accretion, tuning, or reorganization (Seel, et al., 2009). Accordingly, if a schema for any problem type is available, the schema is mapped onto the problem to be solved promptly (Jonassen, 2000). If assimilation is not successful, accommodation must take place in order to reorganized or restructure an individual's knowledge. However, when no schema is available or its reorganization fails, the human mind switches to the construction of a mental model which is defined as a dynamic ad hoc representation of a phenomenon or problem that aims at creating subjective plausibility through simplifying and envisioning the situation, or through analogical reasoning (see Figure 8.1).

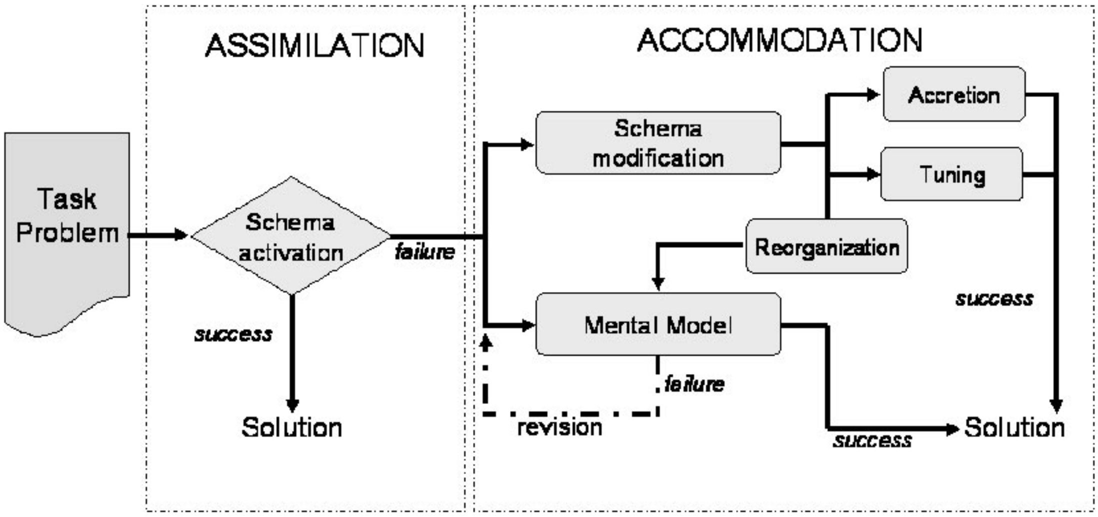

FIGURE 8.1. Cognitive functions of assimilation and accommodation

In accordance with Johnson-Laird's (1983) idea of "fleshing out," we argue that an individual constructs a mental model by integrating relevant bits of domain-specific knowledge into a coherent structure step by step in order to meet the requirements of a phenomenon to be explained or a problem to be solved (Seel, 1991). Understanding this step-by-step process more precisely will help instructors to organize learning materials, identify knowledge gaps, and relate new learning materials to existing slots or anchors within the learners' cognitive structures (Jonassen, 1987, 2000).

Learning-dependent progression of mental models

When humans are confronted with a problem, they can apply either a schema or a mental model that hypothesize mechanisms, either structures or processes, that account for the problem to be solved. However, in order to understand the continuous progression of learning, thinking, reasoning, and problem solving, the underlying mental representations must be assessed carefully at the various stages of the learning process. Evidently, measuring cognitive structures continuously or repeatedly during transitional stages is more effective than only measuring them before and after instruction.

In our current research, we characterize the learning-dependent progression of cognitive structures as a specific kind of transition which mediates between mental models, which describe the initial states of the learning process, and schemata, which are described as the desired end state of learning. Exempli gratia, a novice may not be able to activate a well-developed schema to solve a specific task. Hence, this novice will rely on general schemata and in all probability will fail to successfully solve the task immediately (Jonassen, 2000). Accordingly, the novice will create, through an iterative process, various types of mental models in order to successfully (judged under subjective plausibility) solve the task. In contrast, an expert will recognize the type of task and map an existing schema onto the specific task to solve it (Jonassen, 2000). Therefore, our research focuses on the long-term perspective of changes in mental models and schemata along with the transition of mental representations from mental models to schemata. Specifically, we aim to identify transition points within a learning progression at which the shift of cognitive structures from mental model (fluctuation in probability of change) to schemata (decrease in probability of change) occur (see Figure 8.2).

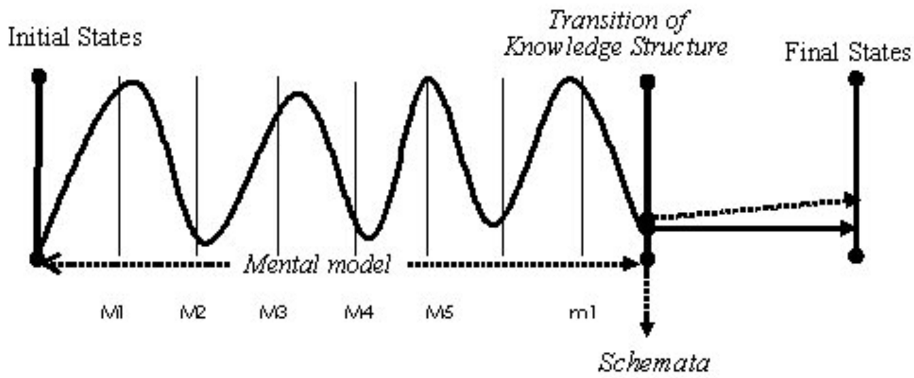

FIGURE 8.2. *Transition of cognitive structures*

Feedback and cognitive structures

Feedback is considered to be any type of information provided to learners with regard to their learning progress (Wagner & Wagner, 1985). Accordingly, feedback can take on many forms depending on a particular theoretical perspective, the purpose it is intended to serve, research goals, and methodological approaches. Moreover, feedback is considered an elementary component for supporting and regulating learning processes. Especially in computer-based and self-regulated learning environments, the nature of feedback is of fundamental importance (Simons & de Jong, 1992). Unlike this initial general understanding of feedback, the term informative feedback refers to all kinds of external post-response information used to inform the learner of his or her current state of learning or performance (Narciss, 2006, 2008). Widely accepted forms of feedback include (a) knowledge of result, (b) knowledge of correct result, (c) knowledge of performance, (d) answer until correct, (e) knowledge of task constraints, (f) knowledge about concepts, (g) knowledge about mistakes, (h) knowledge about how to proceed, and (i) knowledge about metacognition (Narciss, 2008). Feedback on cognitive structures, such as the use of conceptual models (i.e. explicit and consistent causal explanations of a given phenomenon) to help persons to build mental models or schemata of the system being studied, has also been investigated and discussed (e.g., Mayer, 1989; Norman, 1983; Seel & Dinter, 1995). Further, new forms of automated and individualized feedback have been successfully implemented in self-regulated learning environments (e.g., Ifenthaler, 2009).

From an instructional point of view feedback can be provided by internal (individual cognitive monitoring processes) or external (various types of correction variables) sources of information. Internal feedback may validate the externally provided feedback, or it may lead to resistance against it (Narciss, 2008). However, the empirical evidence of effects of different types of feedback is rather inconsistent and contradictory in parts (e.g., Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Clariana, 1993; Kluger & DeNisi, 1996; Kulhavy, 1977; Mory, 2004).

While solving problems of the world, cognitive structures provide subjectively plausible explanations on the basis of restricted domain-specific information (see Ifenthaler, 2010c). Accordingly, such cognitive structures are in many cases resistant to changes as they have a high subjective plausibility which requires special types of feedback. Indeed, various research studies have shown that it is very difficult but possible to influence the generation of plausible mental models by providing specific information (see Anzai & Yokoyama, 1984; Ifenthaler & Seel, 2005; Mayer, 1989; Seel, 1995; Seel & Dinter, 1995). Ifenthaler and Seel (2005) argue that it is important to consider how such feedback is provided to the learner at specific times during the learning process and how it is structured.

Learning experiences and problem solving

Individual differences in problem solving depend on the characteristics of the problem, i.e. its scope, degree of structuredness, and complexity, which correlates with the cognitive operations necessary for solving a problem (Funke, 1991). Problems can be well-structured or ill-structured: well-structured problems, like textbook problems, are composed of few variables, while ill-structured problems may include many factors or variables that may interact in unpredictable ways (Funke & Frensch, 1995). For many people, inductive reasoning tasks are not easy to solve and actually produce a problem (Holland, Holyoak, Nisbett, & Thagard, 1986).

From the perspective of research on problem solving, inductive reasoning tasks may be considered as well-structured problems for which a solution exists and can be found (Feeney & Heit, 2007).

Cognitive psychologists propose that the first thing a person does when confronted with a problem is to try to construct a mental representation of its relevant features (Dörner & Wearing, 1995). Accordingly, problem solving presupposes that people either activate appropriate schemata or actively construct meaningful representations, such as mental models, which represent and communicate subjective experiences, ideas, thoughts, and feelings. By means of such representations an individual is able to simulate real actions in imagination (in the sense of thought experiments) in order to solve problems (Seel, et al., 2009). In this context, a mental model fulfills several functions: (1) It guides the comprehension of the system as well as the concrete operations with it; (2) it allows the system's states to be explained; and (3) it allows predictions about the system's behavior and the effects of intervention in the system to be derived (Greeno, 1989; Young, 1993). As shown above, solving a task requires iterative steps of hypothesis testing as well as an increased time for constructing appropriate cognitive structures (Funke, 1992). This constitutes a problem in itself because cognitive structures are regularly incomplete and constantly evolving. They are usually not an accurate representation of a phenomenon but rather typically contain errors and contradictions. However, especially mental models are parsimonious and provide simplified explanations of complex phenomena. Additionally, they often contain measures of uncertainty concerning their plausibility. This allows mental models to be used even if they are incorrect from an expert's perspective.

These iterative processes of hypothesis testing while solving a task are closely related with learning experiences that are represented in long-term memory as declarative and/or procedural knowledge (Jonassen, et al., 1993). Another indicator for solving tasks is the person's awareness of the problem type (applied strategy). Sweller (1988) argues that experienced problem solvers are able to automatically use strategies to solve familiar tasks. However, transfers of successful strategies to different kinds of tasks are on rare occasions (Gick & Holyoak, 1980; Jonassen, 2000).

Accordingly, two types of change while investigating solving tasks with a longitudinal perspective are of special interest. The first has to do with how experts (experienced problem solvers) adapt their learning experiences or strategies within the solution processes. The second is about how novices become experts over time, how their learning experiences develop / accumulate during this process, and how their strategies change (Seel, et al., 2009).

Research questions and hypotheses

Based on the literature overview, the following research questions and hypotheses were addressed: (1) Do specific transition points within a learning progression exist at which the shift of cognitive structures from mental model (fluctuation in probability of change) to schemata (decrease in probability of change) occurs? For being able to answer our first research question, we argue that there is strong evidence that the research on mental models and schemata has to move beyond the traditional two-wave design in order to capture changes more precisely (Ifenthaler, 2008; Willett, 1988). Another requirement for measuring mental models and schemata precisely is that the diagnosis should be embedded in a complex problem situation (Funke, 1991; Seel, et al., 2009). Hence, participants are confronted with a set of different inductive reasoning tasks at ten measurement points. In inductive reasoning, the premises of an argument indicate some degree of support for the conclusion but not entail it (Feeney & Heit, 2007; Heit, 1998; Holland, et al., 1986; Sternberg & Gardner, 1983). There is an ongoing debate on processes of inductive reasoning focusing e.g., on development of reasoning process of children (e.g., Hayes & Thompson, 2007), teaching of inductive reasoning (e.g., K. J. Klauer, 1996), self-directed learning (e.g., Wilhelm & Beishuizen, 2003), cross-sectional assessments (e.g., Csapo, 1997), and everyday decision making (e.g., Nisbett, Krantz, Jepson, & Kunda, 1983). The longitudinal perspective of our empirical investigation wants to add and complement the available body of literature on inductive reasoning.

In order to identify the specific point at which the transition of cognitive structures from mental models (discussed in terms of the fluctuation in probability of change) to schemata (discussed here as the decrease in probability of change) occurs, our experimental groups receive different types of task classes. One experimental group receives tasks which require identical solution procedures, whereas the other experimental group receives tasks with varying solution procedures. We assume that persons who receive inductive reasoning tasks which require identical solution strategies will have a stronger decrease in the probability of change, while persons who receive tasks which require different kinds of solution strategies will have a stronger fluctuation in probability of change (Hypothesis 1).

Regarding feedback, we wanted to investigate a conservative type of feedback in our longitudinal study which provides information about the strategy in order to solve the task in question: (2) Can feedback effectively support the learning-dependent development of cognitive structures? As shown above, feedback plays a particularly important role in highly self-regulated model-centered learning environments because it facilitates the development of mental models and schemata (Ifenthaler, 2009). Past research studies demonstrate how different forms of feedback can be provided to improve a person's understanding of a specific task in a given context. However, most of these research studies lack a longitudinal perspective (e.g., Mayer, 1989; Norman, 1983; Shute, 2008). We assume that if learners have access to feedback, which guides them in finding a strategy to solve the logical reasoning task, they will perform better than they would without feedback (Hypothesis 2).

Additionally, previous research studies (e.g., Hilbert & Renkl, 2008; Ifenthaler, et al., 2007) have found that verbal and spatial abilities do not affect the quality of model-building processes. However, the above mentioned studies did not include a longitudinal design. Hence, we are interested in replicating these results within a longitudinal perspective. Additionally, as learning in our experimental investigation is highly self-regulated, motivation is another important factor to be taken into account (e.g., Keller, 1983). However, motivationally relevant factors are seldom linked to mental models and schemata. Therefore, a third research question to be explored is: (3) Do verbal abilities and the degree of achievement motivation affect the logical reasoning task outcome? We assume that persons with higher achievement motivation will outperform persons with lower achievement motivation (Hypothesis 3a). Additionally, we assume that verbal abilities will have no effect on the learning outcome (Hypothesis 3b).

Method

Participants

Initially 73 German university students of educational science took part in our experiment. However, as not every student was present at all ten measurement points, we had a total of 64 participants (56 female and 8 male). Their mean age was

22.3 years (*SD* = 2.29). They were enrolled in a research methods course of intermediate level.

Design

The participants were randomly assigned to one of three experimental conditions: self-guided & non-varying strategy (SG-N; n_1 = 21), self-guided & varying strategy (SG-V; n_2 = 21), and scaffolding-based & non-varying strategy (SB-N; n_3 = 22). Varying and non-varying strategy are related to the type of inductive reasoning tasks. Varying strategy means that the solution strategy for the inductive reasoning task changed at every measurement point. Participants in the SG-N groups had to solve four consecutive inductive reasoning tasks in which it was possible to apply the same solution procedure. Figure 8.3 shows the longitudinal research design with ten measurement points and the three experimental groups. Participants in the SB-N group received support on which strategy to apply for the first and sixth task. Participants in SG-N and SB-N received tasks in which the solution strategy was identical for measurement points one to four and six to nine (see Figure 8.3). Participants in SG-V received tasks with varying solution strategies at all ten measurement points. At measurement points one, five, and ten, the inductive reasoning tasks were identical for all experimental groups.

SG-N: R		O_{Ta1}	O_{Ta2}	O_{Ta3}	O_{Ta4}	O_{Tb0}		O_{Tc1}	O_{Tc2}	O_{Tc3}	O_{Tc4}	O_{Te0}
SG-V: R		O_{Ta1}	O_{Tb1}	O_{Tc1}	O_{Td1}	O_{Tb0}		O_{Ta2}	O_{Tb2}	O_{Tc2}	O_{Td2}	O_{Te0}
SB-N: R	X_1	O_{Ta1}	O_{Ta2}	O_{Ta3}	O_{Ta4}	O_{Tb0}	X_2	O_{Tc1}	O_{Tc2}	O_{Tc3}	O_{Tc4}	O_{Te0}

FIGURE 8.3. *Longitudinal research design (SG-N: self-guided & non-varying strategy; SG-V: self-guided & varying strategy; SB-N: scaffolding-based & non-varying strategy; O = measurement of dependent variable; X = treatment; T = task; a, b,c, d, e = strategy to solve the task)*

Our experiment was implemented on a web-based platform, which enabled us to track the participants' behavior and, more importantly, the time needed to solve the ten tasks. Based on the participants' login and experimental condition, our web-based platform assigned the corresponding task (and if required the feedback) at each measurement point. It was not possible to log in again to solve the task a second time.

Materials

- Achievement motivation inventory: The short version of the LMI-K (Leistungsmotivationsinventar; i.e. an achievement motivation inventory) was used to test the participants' achievement motivation. The LMI-K consists of 30 items which are combined to form a global value. Schuler and Prochaska (2001) report high reliability scores for the LMI-K (Cronbach's alpha = .94).

- Verbal abilities: A subscale of the I-S-T 2000 R (Amthauer, et al., 2001) was used to test the participants' verbal abilities. This test is a widely used intelligence test in Germany with high reliability (r = .88 and r = .96; split-half reliability). A total of 20 sentences with a missing word had to be completed using a set of five words. The participants had six minutes to complete this subset on verbal abilities.
- Inductive reasoning tasks and feedback: 14 inductive reasoning tasks in the German language were administered at specific points in time (see Table 8.1 for examples). Solving a task took approximately 15 minutes on average. As shown in our experimental design (see Figure 8.3), we administered tasks which required identical and different solution strategies. Two sets of four tasks required the same solution strategy, and the remaining six tasks required different solution procedures. Table 8.1 shows two examples of tasks, the corresponding feedback which was provided to the subjects in the SB-N group, and the solution. Difficulty of tasks increased slightly during the ten measurement points.
- Logical reasoning rating test: The logical reasoning rating test consisted of five items focusing on the difficulty, motivation, time, solution procedure, and replicability of the tasks (Cronbach's alpha = .83). The questions were answered on a four-point Likert scale (1 = totally disagree; 2 = disagree; 3 = agree; 4 = totally agree).

Procedure

In the first phase of the experiment, the participants completed a *demographic data* questionnaire, the short version of the *LMI-K*, and the subset of the *I-S-T 2000 R*. Additionally, participants were randomly assigned to the three experimental conditions. In the second phase, participants solved ten tasks within five weeks (two tasks per week, Mondays and Thursdays). After logging into the web-based platform with a personal codeword, the participants were provided with the task. Here the participants were asked to type in (a) the *solution* to the task and (b) the *strategy* they applied to solve it. Additionally, the participants had to estimate how long it took them to solve the task (*estimated time on task*). Subsequently, they filled out the five items of the *logical reasoning rating test*.

TABLE 8.1
Two examples of inductive reasoning tasks with different solution strategies, provided feedback, and solutions (translated from German)

Example task	Provided feedback	Solution
A father is the same age as his three sons together. Ten years ago, he was three times as old as his oldest son and five times as old as his second oldest son. The youngest son is 14 years younger than his oldest brother. How old are the three sons?	The problem includes four variables: Father (f), son 1 (s1), son 2 (s2), and son 3 (s3). Accordingly, you need four equations. Equation one would be: f = s1 + s2 + s3. Now find the remaining equations to solve the problem.	Son one = 25 years old, son two = 19 years old, and son three = 11 years old.
All three friends Anton, Hans, and Karl play two musical instruments. Hence, we are able to give everybody two of the following designations: Flautist, drummer, violinist, cellist, trumpeter, and pianist. The flutist likes to take the mickey out of the violinist; the trumpeter and violinist join Anton for watching a soccer game; the cellist is in debt to the drummer; the flutist is engaged with the sister of the cellist; Hans hid the trumpeter's instrument; and Karl has won against Hans and the cellist in the last card game. Now it should be clear which instruments are played by whom?	First create a table with three columns and three rows. The first column is for the names, the second, and third for the corresponding instruments	Anton: pianist, cellist Hans: violinist, drummer Karl: trumpeter, flautist

Scoring

For each participant, an achievement motivation and a verbal ability score were determined. Furthermore, we determined each participant's task solution score, points being awarded for partial or full solution of the tasks at the ten measurement points (0 – 5 points). Additionally, an average score for the logical reasoning rating test was determined.

Task strategy measure: To analyze the strategies for solving the tasks during our longitudinal experiment, a scoring rubric was developed. We determined each participant's task solution score, points being awarded for partial or full solution of the tasks at the ten measurement points (0 – 5 points). The task strategy measure (0 = NS; 1 = WS; 2 = RS) at the ten measurement points was scored as follows: (NS) no strategy for solving the task; (WS) application of an incorrect strategy for solving the task; (RS) application of the correct strategy for the task. For the task solution score and task strategy measure we found a very highly significant correlation, $r = .914$, $p < .001$.

Time spent for solving the tasks: We tracked the time spent on solving the task within the online experimental environment (TT: tracked time), and the participants were asked to estimate how long it took them to solve the task (ET: estimated time).

Using transition probabilities to identify change: A process which develops dependent on time and in accordance with probabilistic principles is a stochastic process. This means that we cannot predict with certainty its future behavior but rather only probabilities as to various possible states for the future. Bartholomew (1967) introduced the application of stochastic models for describing social processes, specifically the growth of different generations within families and societies. In this context, Ifenthaler and Seel (2005) considered the progression of cognitive structures to be comparable to the growth of such social processes.

Thus, we assume that changes in cognitive structures can be characterized by transition probabilities which develop over time. In order to model and analyze the likelihood that one given state of a cognitive structure (mental model or schemata) will be followed by another, we compute transition probabilities from one state to another. The results can be presented in a transitional probability matrix (see Equation 1).

$$P = \begin{bmatrix} .77 .23 \\ .38 .62 \end{bmatrix} \qquad (1)$$

In matrix P, the entries in each row add up to 1. For example, there is a .38 probability that a less elaborated cognitive structure will increase in size or a .23 probability that an elaborated cognitive structure will decrease in size. These transition probabilities can be illustrated by means of a state transition diagram, which is a diagram showing all states and transition probabilities (see Figure 8.4). Possible missing arrows indicate zero probability; the density of the arrows indicates the potency of probability.

In order to identify which transition probability deviates significantly from its expected values, a z-score is computed to test significance. A z-score larger than 1.96 absolute is then regarded as statistically significant at the .05 level (Bakeman & Gottman, 1997). The above-described stochastic models provide the mathematical basis for precisely computing learning-dependent changes in cognitive structures (Ifenthaler & Seel, 2005).

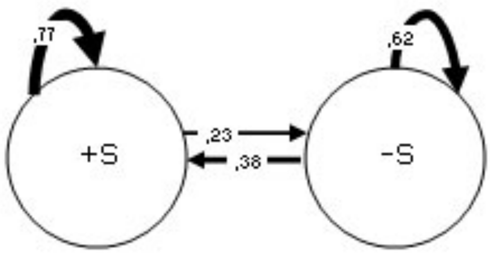

FIGURE 8.4. State transition diagram of Equation (1)

Results

Initial data checks showed that the distributions of ratings and scores satisfied the assumptions underlying the analysis procedures. Main effects of gender were not significant for any measure. All effects were assessed at the .05 level. As effect size measures, we used Cohen's d (small effect: $d < .50$, medium effect $.50 \leq d \leq .80$, strong effect $d > .80$) and partial η^2 (small effect: $\eta^2 < .06$, medium effect $.06 \leq \eta^2 \leq .13$, strong effect $\eta^2 > .13$).

Longitudinal perspective on task solution

Participants spent an average of M = 206.78 (SD = 111.13) minutes solving all ten tasks (tracked time). In order to obtain an overview of overall performance during the ten measurement points, we analyzed the individual answers. Table 8.2 shows the means of task solution score and task strategy measure. An ANOVA showed no significant differences for the overall task solution scores between the SG-N (M = 32.00, SD = 6.85), SG-V (M = 30.00, SD = 7.94), and SB-N (M = 32.54, SD = 8.05) experimental group, $F(2, 63) = .66$, $p = .523$. Also, we found no significant difference for the task strategy measure between the SG-N (M = 12.14, SD = 2.78), SG-V (M = 10.95, SD = 3.22), and SB-N (M = 10.86, SD = 3.43) experimental group, $F(2, 63) = 1.09$, $p = .344$.

TABLE 8.2
Means, standard deviations, minimum and maximum scores of task solution score and task strategy measure (N = 64)

	M	SD	Min	Max
Task solution score	31.53	7.59	15	49
Task strategy measure	11.31	3.16	4	19

Note. For ten measurement points, task solution score (maximum = 50); task strategy measure (maximum = 20).

Learning-dependent progression of task solution score

We computed a repeated-measure MANOVA with the task solution score at ten measurement points as a within-subjects factor, and experimental groups (self-guided & non-varying strategy, self-guided & varying strategy, and scaffolding-based & non-varying strategy) as a between-subjects factor. The sphericity assumption was not met ($\chi 2(44) = 66.17$, $p = .017$), so the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied. The difference between measurements was significant, $F(7.2, 437.5) = 26.85$, $p < .001$, $\eta 2 = .306$ (strong effect). We also found a significant interaction, $F(14.3, 437.5) = 3.06$, $p < .001$, $\eta 2 = .091$ (medium effect). However, the difference between experimental groups was not significant, $F(2, 61) = .66$, $p = .523$.

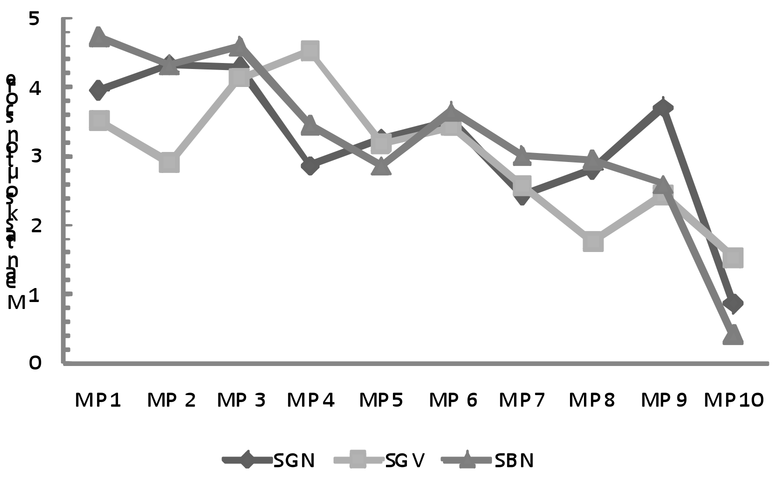

FIGURE 8.5. *Mean task solution score over time, by experimental group*

The results of our MANOVA analysis indicated a significant difference in the mean task solution score over time (see Figure 8.5). Additionally, the significant interaction effect showed that the mean task solution score of the three experimental groups changed differently over time. A pairwise comparison of the task solution score at different times indicated significant differences between experimental groups for the following measurement points (MP): MP 3 – MP 4 ($F(2, 61) = 6.43$, $p = .003$, $\eta 2 = .174$), MP 4 – MP 5 ($F(2, 61) = 4.03$, $p = .023$, $\eta 2 = .117$), and MP 9 – 10 ($F(2, 61) = 4.64$, $p = .013$, $\eta 2 = .132$). However, we found no difference in the mean task solution score between the three experimental groups over time. See Appendix A for means and standard deviations.

Transition probabilities of task strategy measure

In order to model and analyze the likelihood that one given state of a cognitive structure (mental model or schemata) will be followed by another, we computed transition probabilities from one measurement point to another (see Appendix B for the transitional probability matrix, including *z-scores*). Based on the transition probabilities, we were able to illustrate all states and transition probabilities by means of a state transition diagram. Possible missing arrows within the diagrams indicate zero probability; the density of the arrows indicates the potency of probability. These transition state diagrams reveal similarities and differences concerning the *task strategy measure* (NS, WS, RS) of the tasks during the learning process (ten measurement points). Accordingly, these diagrams help us to identify specific points during the task solution process which may give an insight into changes of cognitive structures from mental models to schemata.

Overall, the transition probabilities and state diagrams for participants in the *SG-N* group (see Appendix B) revealed a possible schematization between MP1 and MP4 and between MP6 and MP9, because it was very likely that once they had applied a correct strategy for solving a task they did not revert to an incorrect strategy (see Figure 8.6).

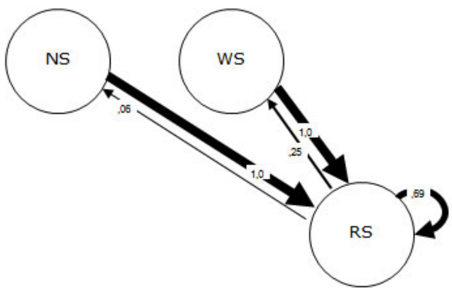

FIGURE 8.6. *State transition diagram for participants in the SG-N experimental group (n_1 = 21), MP 2 – 3*

For participants in the *SG-V* group the transition probabilities and state diagrams revealed a possible construction of mental models between MP1 to MP10 (see Appendix B), because it was very likely that they changed state between each measurement point and also often reverted to incorrect strategies (see Figure 8.7).

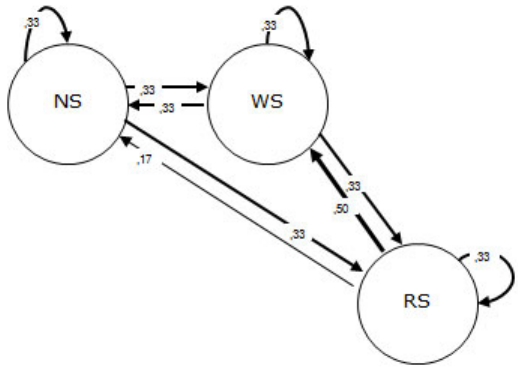

FIGURE 8.7. *State transition diagram for participants in the SG-V experimental group (n_2 = 21), MP 6 – 7*

The transition probabilities and state diagrams for participants in the *SB-N* group revealed a possible schematization between MP1 and MP4 and between MP6 and MP9 (see Appendix B), because it was very likely that once they applied a correct strategy to solve a task they did not revert to incorrect strategies. Additionally, the feedback at MP1 and MP6 caused higher probabilities of change at the following MPs (see Figure 8.8).

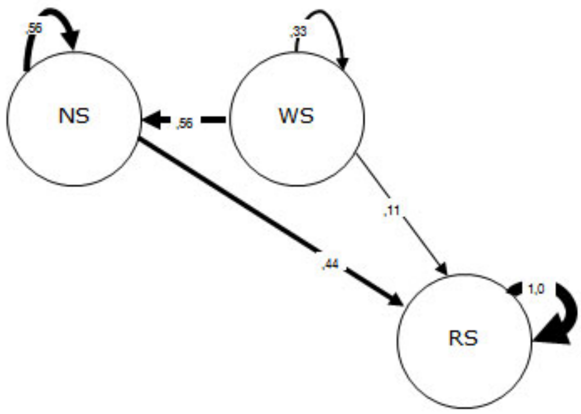

FIGURE 8.8. *State transition diagram for participants in the SB-N experimental group (n_3 = 22), MP 8 – 9*

Finally, we found a high probability in all three experimental groups of solving the task correctly at MP9 and having no solution at MP10. Accordingly, we assume that the task at MP10 was too difficult (including the underlying strategy) to be solved by the participants.

Verbal abilities and achievement motivation

Participants could score a maximum of 210 points on the *achievement motivation test* and 20 points on the subset of the I-S-T 2000 R on *verbal abilities*. On the test for achievement motivation, participants scored *M* = 140.11 points (*SD* = 23.04) and on the test for verbal abilities they scored *M* = 12.97 points (*SD* = 3.94). Table 3 shows the correlations for *the task solution score* and *task strategy measure* with the participants' *achievement motivation* and *verbal abilities* scores. We found no significant correlation between achievement motivation or verbal abilities and the task solution score and task strategy measure. The data were divided into high and low *achievement motivation groups* by a median split. Still, a t-test analysis revealed no significant difference for the task solution score, $t(62) = -.936$, $p = .353$, and task strategy measure, $t(62) = -1.74$, p = n.s. Additionally, we divided the data into high and low *verbal abilities groups* by a median split. Also, the t-test analysis revealed no significant difference for the task strategy measure, $t(62) = -1.70$, p = n.s. However, we found a significant difference for the task solution score between participants with high verbal abilities ($M = 33.41$, $SD = 7.82$) and low verbal abilities ($M = 29.66$, $SD = 7.00$), $t(62) = -2.02$, $p = .048$, $d = .51$ (medium effect).

Accordingly, the results support the hypothesis that verbal abilities are not related to mental model and schematization processes for the *task strategy measure*. However, we have to reject our hypothesis for the *task solution score* since participants with high verbal abilities outperformed those with low verbal abilities. Additionally, we have to reject our hypothesis that achievement motivation has an influence on the task strategy measure and the task solution score.

TABLE 8.3
Correlations between achievement motivation, verbal abilities, task solution score, and task strategy measure (*N* = 64)

	Achievement motivation	Verbal abilities
Task solution score	.163	.205
Task strategy measure	.242	.178

Discussion

This study is part of our current research on model-based reasoning grounded on the theoretical assumptions of cognitive structures. In this paper we examined the progression of cognitive structures that learners produce in solving a series of tasks within a given instructional context. More specifically, we attempt to identify the learning-dependent progression of mental models and their transition to schemata.

On the on hand, mental models enable mental "leaps" in the establishment of truth values and operate only with the premises which are directly consistent with the conclusion (Holland, et al., 1986; Holyoak & Thagard, 1995). Thus, mental models make it possible for people with minimal information to reach correct conclusions since they test the truth value of only the premises which are subjectively plausible and do not contradict the conclusion when combined with one another. On the other hand, Bransford (1984) has pointed out that schema activation and schema construction are two different problems. Although it is possible to activate existing schemata with a given topic, it does not necessarily follow that a learner can use this activated knowledge to develop new knowledge and skills. This can be done by means of constructing and revising explanatory models – as advocated in the mental model hypothesis (Seel, 1991).

Although we do not know how many repetitions of similar experiences will be necessary to develop a schema, we argue that learning experiences with structurally similar tasks will result in a learning-dependent progression of mental models. Snow (1990) identified the learning-dependent mental model progression as a specific kind of transition mediating between preconceptions, which describe the initial states of the learning process, and causal explanations, which are described as the desired end state of learning. We understand the initial states of learning as working models that are condensed – as a result of repeated learning experiences – to a stable mental model or even an inferential schema that can be applied to solve a class of particular problem solving tasks. More specifically, we assume that there is a specific point in the learning process at which a transition from a mental model (indicated by fluctuations in probability of change) to an inferential schema occurs (indicated by a decrease in probability of change).

At specific measurement points we found interesting significant differences between the treatments (Hypothesis 1). We found that learners in the SB-N condition (i.e., scaffolding-based with no variations in the type of task) outperformed learners in the SG-V condition at the first measurement point, $F(2, 63) = 4.97$, $p = .010$, $d = .14$. Hence, at the very beginning of the learning process the feedback (scaffold) was very effective and the learners were able to solve the task significantly better than students who did not receive the feedback (Hypothesis 2). However, at the following nine points of measurement there were only a few significant differences between the experimental groups. This indicates that all subjects were successful – independently of the particular experimental condition – in constructing effective mental models for mastering the tasks provided.

Also, at the second measurement point the learners in the SG-V were outperformed by the learners of the SG-N and SB-N conditions, $F(2, 63) = 7.05$, $p = .002$, $d = .19$. Accordingly, learners who were able to apply the same mental model to the second task (conditions SG-N and SB-N) were more successful than learners who needed to apply another strategy (new mental model) to solve the task (SG-V condition). This supports the assumptions of our first research question.

Additionally, the significant difference between conditions at the fourth measurement point strengthens our hypothesis (Hypothesis 1). Here, learners in the SG-V condition (self-guided with variations of tasks) outperformed the learners in both the SG-N and SB-N conditions, $F(2, 63) = 8.68$, $p < .001$, $d = .22$. Hence, having applied different strategies to solve the tasks enables better performance after a specific learning period. This result supports the assumption that it is more effective to construct flexible mental models like those required by the variation of tasks. Seel, Darabi, and Nelson (2006) have pointed out that within any given domain of activity, the richness and flexibility of a learner's mental model directly influences the quality of his or her task performances in that domain. In other words, a person (for instance, an expert) who has a rich and powerful set of strategies (mental models, related to a particular task domain) will show much greater productivity and diversity with respect to solving tasks than someone (for instance, a novice) who has only weak mental models.

Regarding the *task solution strategy*, we computed transition probabilities to identify fluctuations and stability over time. The state transition diagrams helped to identify differences between the three experimental groups. Actually, transition probabilities and state transition diagrams are good indicators for identifying fluctuation and stability in learning processes. This procedure can be considered a suitable methodology for assessing the learning-dependent progression of cognitive structures.

Furthermore, we looked at the influence of verbal abilities and achievement motivation on the task solution. We expected that learners with higher achievement motivation would outperform other learners (Hypothesis 3a). Additionally, on the basis of previous studies (Hilbert & Renkl, 2008; Ifenthaler, et al., 2007), we expected no differences between learners with high and low verbal abilities in terms of their mean task solution score (Hypothesis 3b). Indeed, the results of our research support the hypothesis that verbal abilities are not related to mental model and schema processes for the *task strategy measure*. However, we have to reject our hypothesis for the *task solution score* since participants with high verbal abilities outperformed those with low verbal abilities. Additionally, we have to reject our hypothesis that achievement motivation has an influence on the task strategy measure and the task solution score.

In addition to extending the research literature on cognitive structure, our study may enhance information available to instructional designers and educators. Most people can cope effectively with cognitively demanding tasks by constructing and maintaining a *mental model* that provides them with enough understanding of the task to be accomplished. In this sense, the notion of mental models is interrelated with the investigation of inductive reasoning and problem solving, which provides a unique challenge for research in the field of learning and instruction (Jacobson & Archodidou, 2000). This can be illustrated by the discussion on higher-order instructional objectives concerning logical reasoning and problem solving. Actually, several scholars such as Lesh and Doerr (2000) and Schauble (1996), encourage the pursuit of higher-order objectives and argue that helping students to develop their own "explanatory models" should be among the most important goals of math and science education. A recommendation often made in recent learning theory and research is to involve students, either individually or in groups, in actively constructing mental models for mastering cognitively demanding tasks, such as inductive reasoning tasks. The construction of a mental model in the course of learning often necessitates both a restructuring of the underlying representations and a reconceptualization of the related concepts. Of course, there is no need for a mental model as long as the learner can assimilate the learning material into the structures of his or her prior knowledge. Therefore, a substantial *resistance to assimilation* is a prerequisite for constructing a mental model, and the degree of this resistance depends greatly on the complexity or difficulty of the tasks to be mastered. An alternative to a model-based approach of inductive reasoning within the realm of instruction is certainly a schema-based approach, such as cognitive load theory which recommends the use of means-end-analysis and worked examples that are presented to students to show them directly, step by step, the procedures required to solve conventional problems, such as inductive reasoning tasks (Sweller, 1988). Both the model-based and schema-based approach agree at the point that learning occurs when people actively construct meaningful representations, such as mental models or schemata (Mayer, Moreno, Boire, & Vagge, 1999).

However, such representations are constructed from significant properties of external information, e.g. well-designed learning environments or materials. This corresponds with a basic assumption of constructivist approaches of learning according to which learners respond sensitively to characteristics of the environment, "such as the availability of specific information at a given moment, the duration of that availability, the way the information is structured" (and presented), "and the ease with which it can be searched" (Kozma, 1991, p. 180). In contrast with schema-based argumentations researchers in the field of mental models argue that context sensitivity occurs consciously and intentionally. Among others, Anzai and Yokoyama (1984) assume that learners encode information on a problem in a mental model as soon as they begin working on it in order to gain a basic understanding of the situation and its demands. This initial experiential model can – and the learner is generally aware of this – be false or insufficient for accurately representing the subject domain in question. However, it is *semantically sensitive* toward key stimuli in the learning environment and can thus be transformed into a new model through accurate processing and interpretation of these key stimuli. The results of the experimental study of Anzai and Yokoyama (1984) as well as those of other studies (e.g., Ifenthaler, et al., in press; Ifenthaler & Seel, 2005; Seel & Dinter, 1995) demonstrate the contextual semantic sensitivity in the learning-dependent progression of mental models. Accordingly, learners search continuously for information in the given learning environment in order to complete or stabilize an initial mental model, also know as a multi-step process of model-building and revision (Penner, 2001). Hence, providing appropriate scaffolds or feedback could influence these complex processes.

With regard to the implemented feedback, we found that our conservative type of feedback (information about the strategy in order to solve the task; see Table 8.1) administered at the first and sixth measurement point did not have a strong effect on the learning process and performance. However, we assume that a more elaborated and repetitive version of feedback could facilitate the development of mental models while solving inductive reasoning tasks. Accordingly, based on these findings, a newly conducted experimental study including 20 measurement points explores the effect of feedback on model-building processes in more detail. The proposed model-based feedback not only includes information about the expert solution strategy but also incorporates the learner's prior knowledge (Ifenthaler, 2009).

In summary, a precise and stepwise assessment and analysis of cognitive structures helps us to better understand the differences within and between individuals as they develop over time. This will enable us to identify which instructional materials and instructor feedback are most appropriate at various times during the learning process in order to help educators struggling to find appropriate teaching tools to enhance learning and retention.

Appendix A

TABLE 8.4
Means (standard deviations in parenthesis) of task solution score over time (N = 64)

	Experimental group			Achievement motivation		Verbal abilities		Tracked time on task		Logical reasoning rating	
	SG-N (*n* = 21)	SG-V (*n* = 21)	SB-N (*n* = 22)	Low (*n* = 31)	High (*n* = 33)	Low (*n* = 32)	High (*n* = 32)	Fast (*n* = 32)	Slow (*n* = 32)	Low (*n* = 32)	High (*n* = 32)
MP 1	3.95 (1.40)	3.52 (1.66)	4.73 (.46)	4.26 (1.18)	3.91 (1.49)	4.09 (1,3)	4.06 (1,41)	3.81 (1.49)	4.34 (1.15)	4.03 (1.45)	4.12 (1.26)
MP 2	4.33 (1.32)	2.90 (1.41)	4.32 (1.52)	3.74 (1.65)	3.97 (1.47)	3.75 (1,74)	3.97 (1,36)	3.78 (1.62)	3.94 (1.5)	3.75 (1.63)	3.97 (1.49)
MP 3	4.29 (1.45)	4.14 (1.39)	4.59 (.85)	4.61 (0.99)	4.09 (1.42)	3.91 (1,51)	4.78 (0,71)	4.31 (1.12)	4.38 (1.39)	4.41 (1.19)	4.28 (1.33)
MP 4	2.86 (1.32)	4.52 (1.03)	3.45 (1.54)	3.65 (1.23)	3.58 (1.68)	3.53 (1,44)	3.69 (1,51)	3.25 (1.48)	3.97 (1.38)	3.50 (1.48)	3.72 (1.46)
MP 5	3.24 (1.64)	3.19 (1.57)	2.86 (2.03)	2.68 (1.9)	3.48 (1.5)	2.88 (1,7)	3.31 (1,79)	3.06 (1.63)	3.13 (1.88)	2.91 (1.87)	3.28 1.61)
MP 6	3.52 (1.69)	3.43 (2.11)	3.64 (1.92)	3.65 (1.76)	3.42 (2.02)	3.09 (1,96)	3.97 (1,73)	2.84 (1.97)	4.22 (1.54)	3.53 (1.87)	3.53 (1.93)
MP 7	2.43 (1.91)	2.57 (1.96)	3.00 (1.98)	2.45 (1.88)	2.88 (2.00)	2.66 (1,98)	2.69 (1,93)	2.66 (1.95)	2.69 (1.96)	2.13 (1.79)	3.22 (1.95)
MP 8	2.81 (1.75)	1.76 (1.90)	2.95 (1.94)	2.29 (1.9)	2.73 (1.93)	2.44 (1,9)	2.59 (1,95)	2.28 (1.94)	2.75 (1.88)	2.00 (1.92)	3.03 (1.79)
MP 9	3.71 (1.93)	2.43 (1.78)	2.59 (2.44)	2.52 (2.17)	3.27 (2.04)	2.66 (2,24)	3.16 (2.00)	2.41 (2.2)	3.41 (1.95)	2.03 (2.04)	3.78 (1.85)
MP 10	.86 (.85)	1.52 (1.40)	.41 (1.18)	.77 (1.06)	1.06 (1.39)	.66 (0,94)	1.19 (1,45)	.84 (1.22)	1.00 (1.27)	.84 (1.11)	1.00 (1.37)

Note. SG-N: self-guided & non-varying strategy; SG-V: self-guided & varying strategy; SB-N: scaffolding-based & non-varying strategy

Appendix B

TABLE 8.5
Transitional probabilities (z-scores in parenthesis) of the task strategy measure (NS, WS, RS) for the ten measurement points (N = 64)

		Self-guided & non-varying strategy (*n* = 21)			*Self-guided & varying strategy* (*n* = 21)			*Scaffolding-based & non-varying strategy* (*n* = 22)		
		NS	WS	RS	NS	WS	RS	NS	WS	RS
MP 1-2	NS	0 (-.23)	0 (-.50)	1 (.57)	0 (.72)	0 (-.98)	1 (2.11)	0 (-.32)	1* (2.58)	0 (-1.89)
	WS	0 (-.72)	.14 (-.39)	.86 (.73)	.33 -	.56 (.63)	.11 (-.80)	.20 (.97)	0 (-1.01)	.80 (.17)
	RS	.08 (.80)	.23 (.60)	.69 (-.95)	.36 (.31)	.46 (-.21)	.18 (-.11)	.06 (-.76)	.13 (-.25)	.81 (.73)
MP 2-3	NS	0 (-.23)	0 (-.50)	1 (.57)	0 -	.14 (-1.02)	.86 (1.03)	0 (-.32)	.50 (1.22)	.50 (-.97)
	WS	0 (-.50)	0 (-1.08)	1 (1.24)	0 -	.30 (.14)	.70 (-.14)	0 (-.41)	.33 (.73)	.67 (-1.04)
	RS	.06 (.57)	.25 (1.24)	.69 (-1.43)	0 -	.50 (1.05)	.50 (-1.05)	.06 (.56)	.12 (-1.44)	.82 (1.05)
MP 3-4	NS	0 (-.57)	1 (.80)	0 (-.42)	0 -	0 -	0 -	1 (1.67)	0 (-1.12)	0 (-.48)
	WS	.25 (.06)	.75 (.60)	0 (-.91)	0 (-.65)	.17 (.20)	.83 (.18)	.25 (-.11)	.75 (.91)	0 (-1.04)
	RS	.25 (.23)	.56 (-.95)	.19 (1.05)	.07 (.65)	.13 (-.20)	.80 (-.18)	.24 (-.73)	.53 (-.28)	.24 (1.20)
MP 4-5	NS	.40 (1.37)	.40 (-.39)	.20 (-.72)	1 (1.62)	0 (-.98)	0 (-.57)	.17 (-.68)	.67 (1.81)	.17 (-1.18)
	WS	.15 (-.54)	.54 (.73)	.31 (-.32)	0 (-1.18)	1* (1.96)	0 (-1.05)	.33 (.70)	.33 (-.32)	.33 (-.32)
	RS	0 (-.91)	.33 (-.54)	.67 (1.32)	.29 (.18)	.41 (-1.22)	.29 (1.24)	.25 (-.11)	0 (-1.67)	.75 (1-78)
MP 5-6	NS	.25 (.06)	.25 (-.18)	.50 (.11)	.33 (.31)	0 (-1.18)	.67 (.56)	.67* (2.15)	0 (-1.35)	.33 (-.96)
	WS	.40 (1.66)	.20 (-.83)	.40 (-.67)	.30 (.14)	.20 (.71)	.50 (-.63)	.38 (.43)	.13 (-.52)	.50 -
	RS	0 (-1.81)	.43 (1.03)	.57 (.62)	.20 (-.49)	.20 (.42)	.60 (.15)	0* (-2.42)	.38 (1.78)	.63 (.89)
MP 6-7	NS	.40 (-.64)	.40 (.97)	.20 (-.23)	.33 (.65)	.33 (-.56)	.33 -	.86* (2.29)	.14 (-1.21)	0 (-1.51)
	WS	.67 (.83)	.33 (.65)	0 (-1.62)	.33 (.42)	.33 (-.36)	.33 -	.25 (-1.11)	.75* (2.05)	0 (-1.04)
	RS	.50 (-.21)	.10 (-1.42)	.40 (1.66)	.17 (-.89)	.50 (.76)	.33 -	.36 (-1.28)	.27 (-.46)	.36* (2.21)
MP 7-8	NS	.27 (1.78)	.46 (-1.14)	.27 (-.14)	1 (1.62)	0 (-.83)	0 (-1.24)	.55 (1.30)	.27 (-1.30)	.18 -
	WS	0 (-1.05)	1* (2.22)	0 (-1.62)	.67 (-.42)	.22 (1.72)	.11 (-.80)	.14 (-1.74)	.86* (2.92)	0 (-1.54)
	RS	0 (-1.05)	.40 (-.89)	.60 (1.78)	.57 (-1.02)	0 (-1.05)	.43* (1.97)	.50 (.41)	0 (-1.84)	.50 (1.82)
MP 8-9	NS	.33 (1.02)	0 (-.91)	.67 -	.33 (.76)	.60 (.42)	.07 (-1.58)	.56 (.79)	0 (-1.55)	.44 (.28)
	WS	.08 (-.90)	.25 (.80)	.67 -	0 (-.94)	.50 (-.21)	.50 (1.52)	.56 (.79)	.33* (2.24)	.11* (-2.37)
	RS	.17 (.20)	.17 (-.18)	.67 -	.25 (-.18)	.50 (-.32)	.25 (.68)	0* (-2.02)	0 (-.88)	1* (2.66)
MP 9-10	NS	1 (1.78)	0 (-1.78)	0 -	1* (2.51)	0* (-2.77)	0 (-.65)	1 (1.70)	0 (-1.35)	0 (-.93)
	WS	.50 (-.11)	.50 (.11)	0 -	.33* (-2.55)	.58* (2-21)	.08 (.88)	.67 (-1.07)	.33 (1.57)	0 (-.41)
	RS	.43 (-1.24)	.57 (1.24)	0 -	.67 (.36)	.33 (-.18)	0 (-.42)	.78 (-.98)	.11 (.27)	.11 (1.23)

Note. * indicate transitional probabilities whose values significantly exceed expected, $p < .05$.

9

FACILITATING LEARNING THROUGH GRAPHICAL REPRESENTATIONS 🕮

This experimental study integrates automated natural language-oriented assessment and analysis methodologies into feasible reading comprehension tasks. With the newly developed toolset, prose text can be automatically converted into an association net which has similarities to a concept map. The "text to graph" feature of the software is based on several parsing heuristics and can be used both to assess the learner's understanding by generating graphical information from his or her text and to generate conceptual graphs from text which can be used as learning materials. The study investigates the effects of association nets made available to learners prior to reading. The results reveal that the automatically created graphs are highly similar to classical expert graphs. However, neither the association nets nor the expert graphs had a significant effect on learning, although the latter have been reported to have an effect in previous studies.

🕮 This chapter is based on: Pirnay-Dummer, P., & Ifenthaler, D. (in press). Reading guided by automated graphical representations: How model-based text visualizations facilitate learning in reading comprehension tasks. *Instructional Science*. doi: 10.1007/s11251-010-9153-2

Introduction

Notwithstanding the tremendous efforts of research, design, and development for e-learning, online learning, blended learning, and multimedia learning environments, text still holds the key position within learning environments. Learning has a strong connection to reading and always will. The material ranges from small annotations to whole textbooks. The technologies used in this study to support reading and understanding were initially developed as alternative assessment methods for finding out what a learner knows as opposed to what he or she does not know (e.g., counting errors in classical testing). Like all methodologies they have strengths and weaknesses with respect to what they account for and what features they convey. They never describe states of the mind directly but rather through the medium of external artifacts which correspond to internal states and allow some (but not all) conclusions about what is going on internally. This is a constraint for every empirical approach which addresses cognition. After using and validating the assessment technologies in many studies, we found that the graphical artifacts from the output of the new assessment tools may be used not only for assessment but also as a feedback component for learners. One reason for this is that they are comparatively easy to read, even for non-experts. In this study we investigate an immediate effect of the availability of these artifacts when they are used to support a typical short reading task.

Model supported strategies for reading and understanding

When learners are confronted with medium-sized or long texts, conceptual representations can help them to navigate the meaning – to assimilate the content or navigate the text more efficiently (Crinon & Legros, 2002; Seel & Schenk, 2003). While abstracts, indexes, and sequential information (e.g., tables of content) and their counterparts in text layout are very common aids for navigating the logical sequences of a text, semantic structures are (if at all) only embedded locally. For instance, many texts contain a table of contents, an index, or a glossary, all of which help the reader to navigate the logic (overview) of the text. Semantic structures, on the other hand, only illustrate local content. They can be found in pictures and graphs which illustrate the meaning of locally discussed information (e.g., Eliaa, Gagatsisa, & Demetriou, 2007; Hardy & Stadelhofer, 2006). Expert representations (e.g., models, concept maps, or graphs invented and drawn by experts) help the reader to understand text as well as to assimilation its information into prior knowledge. The integration of new knowledge (assimilation) and the rearrangement of existing knowledge in order to incorporate new and conceptually different aspects (accommodation) are paramount to learning. Thus, the learning process itself uses heuristic resources of reasoning.

A theoretical framework for describing this interrelation is the theory of mental models, and assessment methods from this area of research may provide external graphical structures for visualizing structural content. The role of mental models in deductive and inductive reasoning within learning environments has a strong theoretical foundation (Dinter, 1993; Johnson-Laird, 1983; Seel, 1991, 2003) as well as a sound empirical basis (Al-Diban, 2002; Ifenthaler, 2010c; Ifenthaler & Seel, 2005; T. E. Johnson, et al., 2009; Jonassen & Cho, 2008; Jonassen, et al., 1997; Schnotz, 2001; Seel & Dinter, 1995). The general use of model representations in the form of concept maps for reading has already been investigated and discussed (e.g., Mayer, 1989). According to research findings, the best time to present graphical representations to learners is before the first reading, i.e. before they access the text. One of the major practical problems with this approach is that there is not always an expert available to provide the learners with an expert model because such a model has to be related solely to the specific text. Furthermore, not every expert is trained in reflecting an internal model in the format of a concept map (see Novak, 1998). Therefore, the quality may vary widely depending on the concept mapping skills of the experts one selects. Of course, such skills could be monitored or controlled. However, this involves additional manual effort, and it usually takes too long to work for normal classroom applications. Unfortunately, this is one reason why concept maps which are directly related to a text are rarely used in classrooms.

Therefore, our work integrates automated natural language-oriented assessment and analysis methodologies, e.g., SMD Technology (Surface, Matching, Deep, Ifenthaler, 2010c), T-MITOCAR (Text-Model Inspection Trace of Concepts and Relations, Pirnay-Dummer & Ifenthaler, 2010), into feasible reading comprehension tasks comparable to those implemented in an everyday classroom setting. Our studies have also already shown that the graphical assessment outputs exert considerable influence on ongoing writing (Pirnay-Dummer & Ifenthaler, 2011) and learning (Ifenthaler, 2009, 2010a).

Re-representation

A model is a representation of a thing or a fact or sets thereof. It always has a purpose, yet the purpose can vary. A model may serve more than one purpose. Representation formats can be diverse, ranging from analog (e.g., a miniature model of a house) to symbolic, from simple (where aspects are few and mostly constant) to complex (where aspects, variables, and functions change over time).

A mental model is a specific kind of model. It is inherent to a mind. It is either a representation of something which is outside the mind (the world) or something which is inside the mind (a representation of representations, e.g., a simple guess or a mental simulation). The purpose of mental models is to facilitate decision making, be it inductive, deductive, or different sets thereof. Decision making supports action in the world, including simple and complex problem solving. Human decision making uses a set of heuristics which provide shortcuts for problems which cannot be solved in a sufficient amount of time. Mental models support these heuristics and are thus considered to be at the center of human cognition.

Mental models cannot be observed directly. In order to study them, we need to represent them externally. The externalization process is a heuristic, as is the mental model construction process itself. Because two representation processes are involved – one leading from the world to the mind and the other from the mind back to the world again – we call external representations re-representations to underline the objects we are describing. Re-representations are of course not mental. However, they allow inferences about what is going on inside. Re-representation formats can be based on any objects which allow us to convey at least a part of the mental model. This may be done through language, formalisms, and arranging (e.g., graphical parts), but also by way of art or music.

Thus, the re-representations have purposes that transcend diagnostics (Ifenthaler, 2010d). First, they interact with the inherent model and are therefore often considered to be interesting objects during learning. Second, they are used to communicate. In fact, they are the only known means of establishing communication between minds. Most of the time this is done in natural language. Mental models cannot be shared; they can only be communicated by external means. In our studies, we rely on re-representations in different formats to assess the complex worlds of mental models. However, we also use the same formats to relay content back to the learners.

Automated graphical representations from texts

Whereas classical methods like concept maps (e.g., Cañas, et al., 2004), conceptual graphs (e.g., Sowa, 1984), causal diagrams (e.g., Jensen, 2001), and structure formation techniques (e.g., Scheele & Groeben, 1984) are used to let the learner (or expert) conceptualize his or her knowledge graphically, natural language-oriented methodologies like T-MITOCAR (Pirnay-Dummer & Ifenthaler, 2010) use multiple phases from text to graph. T-MITOCAR automatically converts prose text to an association network using a heuristic.

To illustrate how far we can get by analyzing texts directly, it will be useful come back to an old axiom from research on association and sequences: What is closely related is also closely externalized (Pollio, 1966; Smith, 1894, 1918; Wells, 1911). Texts contain model structures. Closer relations tend to be presented more closely within a text. This does not necessarily work within single sentences, since syntax is more expressive and complex. But texts which contain 350 or more words may be used to generate associative networks as graphs. The re-representation process is carried out in multiple stages. The goal of this approach is to improve the availability of graphical representations of written text across all subject domains (in schools, in companies, in learning management systems, in forums, in chats) and of course also for additional use within qualitative research. It can easily interface with other automated analysis tools, e.g., with the SMD Technology (Ifenthaler, 2010c) or ACSMM (Analysis Constructed Shared Mental Models, T. E. Johnson, et al., 2006). The SMD Technology uses pairwise list forms of graphical drawings (e.g., concept maps) or natural language statements to automatically generate two structural and one semantic measure for quantitatively assessing individuals' re-representations. Besides these quantitative measures, SMD generates four standardized concept map-like representations which can be used for qualitative analysis and as ready-to-use instructional materials: 1) individual or team representation, 2) reference or expert representation, 3) similarity representation (only including semantically similar propositions between individuals/teams and experts), and 4) contrast representation (including propositions which individuals/teams and experts do not share). The ACSMM technology aggregates individual models to group models by means of propositional frequencies which constitute a probability of "sharedness." For a selectable probability value an aggregated model can be constructed by looking at which propositions are commonly shared on this level within a group. Depending on the context, different values are selected. The T-MITOCAR text-to-graph process can be divided into four different stages (see Figure 9.1). Stage 1 is the text input interface, where text is taken into the system (e.g., through a browser interface or at the back end of learning software). In stage 2 the actual model is created by means of parsing and the calculation of association measures. Stage 3 contains the visual output and graphical analysis of the model, and stage 4 allows multiple structural and semantic methods of comparing the graphs.

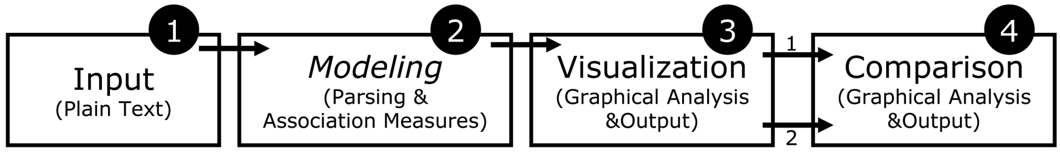

FIGURE 9.1. *Process from text to graph*

When text is pasted to T-MITOCAR from any text source, it may contain characters which could disturb the re-representation process. Thus, all characters which are not part of a specific character set are deleted. The same happens to tags (e.g., HTML tags) and other expected meta-data within each text. When generating the model, we do not want to have formatting code in our way. After the whole text has been prepared in this fashion, it is split into sentences and tokens consisting of words, punctuation marks, quotation marks, and so on. This process is called "tokenizing" and is somewhat language dependent, which means that we need different tokenizing methods for each language we want to use. We only want nouns and names to be part of the final output graph. Hence, we need to find out which words are nouns or names. There are many different approaches and heuristics for tagging sentences and tokens. We found a combination of rule-based and corpus-based tagging to be most feasible when the subject domain of the content is not known in advance, and since T-MITOCAR is designed to work domain independently, this is an important factor. Tagging and the rules for it is a quite complex field of linguistic methods. An explanation of our tagging technique would go beyond what is presentable in this paper. Please see Brill (1995) for a good discussion on mixed rule-based and corpus-based tagging.

Usually we would prefer for different inflexions of a word to be treated as one (e.g., the singular and plural forms "fire" and "fires" should appear only once in the re-representation). Stemming solves this problem by reducing all words to their word stems for the following stages leading to the output graph. Therefore, all words within the initial text and all words within the tagged list of nouns and names are stemmed. After tagging and stemming, the most frequent noun stems are listed from the text. The amount of terms fetched from the text depends on its length in words and sentences. Thus, larger texts also generate larger models. There is, however, a ceiling value. In the running versions of T-MITOCAR no more than 30 single terms are fetched from a text. This value can of course be set for the software. The core algorithms of T-MITOCAR calculate associatedness:

- The default length is calculated. The words are counted for each sentence. The default length is the longest sentence in the text plus one.
- All fetched terms are paired so that all possible pairs of terms are in a list.
- All sentences are analyzed for each pair. If the pair appears within a sentence, the distance for the pair is the minimum number of words between the terms of the pair within the sentence: If at least one term occurs more than one time in the sentence, then the lowest possible distance is taken.
- If a pair does not appear in a sentence (also true if only one of the two terms is in the text), then the distance will be the default length.
- The sum of distances is calculated for each pair.
- The N pairs with the lowest sum of distances find their way into the final output model. Like the list of terms, N depends on the number of words and sentences within the text (exact values can be controlled by the software settings).
- This process automatically cuts the maximum distance from re-representation, even if pairs would normally be presented on the basis of the number of sentences and words. This prevents the algorithm from just deriving random pairs which do not really have any association evidence within the text.

The weights are calculated from the pair distances. They are to some extent comparable to the *combined* measure of the MITOCAR toolset. All weights ($0 \leq w \leq 1$) are mapped linearly so that 1 is the pair with the lowest sum of distances and 0 is the pair with the maximum sum of distances. Linguistic word stems sometimes look strange to untrained viewers. Although one can still guess which words they come from, deriving the output directly from the word stems is no help in reading the re-

representations. Hence, lists of words and their stems are created during stemming for the specific text at hand.

After determining the associatedness and the weight, the procedures use this table to determine which word led most frequently to the stem: If it was the plural, then the plural moves into its place. If it was the singular, then the singular is presented. Thus, the final output model contains a real word in that it uses the inflexion which was most frequently used in the text. The list form is a table which accounts for an undirected graph containing all N pairs (see Table 9.1). It is sorted by weight (descending).

TABLE 9.1
List form of the graph output

Term 1	Term 2	Sum of Distances	Weight
economy	trade	3428	1
exchange	goods	5710	.60
…	…	…	…

The weights ($0 \leq w \leq 1$) at the edges describe the overall weight for the whole noun-distance oriented matrix generated *from the text*. The weights inside the brackets show the weights *within the graph*. This weight is also taken to generate the color of the edges. The strongest edge is red, while the weakest (compared to the graph, not to the text matrix) is blue.

The "text to graph" feature of the software is based on several parsing heuristics and can be used to assess the learner's understanding by generating graphical information from his or her text as well as to generate conceptual graphs from texts which are used as learning materials. It may simply help to have the option of avoiding the effort of an expert model in everyday classroom settings, even if expert models turn out to work better than the automated representations. To create a graphical model from a text, all teachers need to do is upload the text and attach a label to it – in order to find it later on. Additionally available features to make the analysis easier are word counts (of nouns), tables (list form) of the models, and a comparison section that allows comparison of different text based models. The comparison contains measures for graph comparison and graphical representations (pictures), e.g., to represent intersections and difference models.

The output models comply with most of the quality indicators suggested by Mayer (1989). They are *complete* because they represent the text – and only the text is used to build up the structure. This is also the reason why we consider them to be *concise* as regards the task: They only present the associations within the text and therefore have the same scope as the text. However, if the text itself does not correspond to the learning goal or the group then the model that is based on the text will also fail. Thus, the possibility of creating such a model does not obviate the need for the instructional task of selecting a fitting learning text. The models are directly related to the text by design. If the text is compatible with the learners then it will also be *coherent*, as long as it also includes a sufficient amount of words (≥ 350 words).

Pirnay-Dummer, Ifenthaler, and Rohde (2009) provided a study which showed a positive effect of available models on writing when the learners' own text was visualized for the experimental condition. We interpret this as an indicator for *coherence*. In order to decide whether the models are *conceptual*, it is important to know which basis they stand on. Within this study, the experts selected a text on an encyclopedic level. Thus, both the initial authors and the experts thought that it covered correct content and was still able to address a common audience – the models are *conceptual* to that extent. Whether the models are also *considerate* is not yet fully understood. We do not believe that this criterion can be fulfilled a priori by means of the algorithm.

Measures of graph-comparison

The measures for comparison can be applied to any graph, not only to re-representations from T-MITOCAR. There are six core measures for the comparison of conceptual graphs from the SMD Technology (Ifenthaler, 2010c) and from MITOCAR (Pirnay-Dummer, 2006). The indices measure features of graphs. Of all the available measures from graph theory we picked the ones which are theoretically most likely to correspond to the constructs we are trying to describe. We also constructed new algorithms where necessary. In the course of our studies they have shown empirical stability on different occasions. Over time some of the measures may converge, and new ones will certainly also emerge as a result of discussions on future studies. Some of the measures count specific features of a given graph. For a given pair of frequencies f_1 *and* f_2, the similarity results in a measure of $0 \leq s \leq 1$, where $s=0$ is complete exclusion and $s=1$ is identity. The other measures collect sets of properties from the graph (e.g., the vertices = concepts or the edges = relations). In this case, the Tversky similarity (Tversky, 1977).

The four structural and two semantic measures are defined as follows: (1) The *surface* measure (Ifenthaler, 2010c) compares the number of vertices within two graphs. It is a simple and easy way to calculate values for surface complexity. (2) The *graphical matching* (Ifenthaler, 2010c) compares the diameters of the spanning trees of the graphs and is an indicator for the range of conceptual knowledge. It corresponds with structural matching as it is also a measure for structural complexity only. (3) The *density of vertices* measure (also often called "*gamma*") (Pirnay-Dummer & Ifenthaler, 2010) describes the quotient of terms per vertex within a graph. Since both graphs which connect every term with each other term (everything with everything) and graphs which only connect pairs of terms can be considered weak models, a medium density is expected for most good working models. (4) The *structural matching* measure (Pirnay-Dummer & Ifenthaler, 2010) compares the complete structures of two graphs without regard to their content. This measure is necessary for all hypotheses which make assumptions about general features of structure (e.g., assumptions which state that expert knowledge is structured differently from novice knowledge).

(5) *Concept matching* (Pirnay-Dummer & Ifenthaler, 2010) compares the sets of concepts (vertices) within a graph to determine the use of terms. It counts how many concepts are alike. This measure is especially important for different groups operating in the same domain (e.g., using the same textbook). It determines differences in language use between the models. (6) The *propositional matching* (Ifenthaler, 2010c) value compares only fully identical propositions (concept-link-concept) between two graphs. It is a measure for quantifying semantic similarity between two graphs.

The individual measures usually correlate differently. There are significantly higher correlations within each classification (convergent, structure between $r=.48$ and $r=.79$ and semantics between $r = .68$ and $r = .91$) and lower correlations between them (divergent, between $r = -.24$ and $.36$). The density of vertices (gamma) usually stands alone and only rarely correlates with the other structural measures because it accounts for a different feature of structure (correlations between $r=.37$ and $r=.38$).

Pirnay-Dummer et al. (2010) provide a full validation study. The validation study was conducted with N = 1,849,926 model comparisons in 13 different subject domains ranging from common knowledge to scientific subject domains. There is not yet any indication of an interpretable convergence of the measures. They measure different features. Depending on the research question, they either need to be reported completely or selected to fit with the hypotheses if possible, e.g., for research aiming only at the semantic level the structural indices may be omitted or treated as a covariate.

Research questions and hypotheses

We assume that conceptual graphs generated by the T-MITOCAR system can be used to improve reading comprehension in the same way as graphical representations from experts would. This assumption has two aspects. The first has to do with the re-representation object: If the automated graphical representations and expert re-representations share the same central features then they should induce similar effects because the objects are alike. The second aspect is directed at the source of the re-representation. If an expert solution is not available for a specific text, teachers only have a general representation to rely on, if at all. The alternative would be for them to invest the time to create a representation on their own. This is less likely if a large amount of learning texts are at hand, i.e. if the prototype is replaced by a real everyday classroom intervention. In this case the automated text representation may be feasible and still convey the model of the text – maybe even better than a general expert model in the field, because it is directly related to the content of the texts. Thus, we believe that the examination of the model representation influences the model building process in favor of the learning goals as long as the external representation corresponds closely to the selected text basis: Regardless of the learning goal, the text and the representations should correspond to each other as much as possible and share the same properties. This should result in semantic redundancy, which is known to support learning (Christmann & Groeben, 1999).

First, we want to show that the automated representations have high similarities to expert representations – to be on the safe side for interventions. If they are similar it makes sense to assume that they also have similar effects on learning because they share the same structural and semantic properties. This leads to the following first set of hypotheses we tested in our study (each presented as a classical pair of null and alternative hypotheses).

$H_{1.1}$: T-MITOCAR graphs have high semantic similarities to the expert models.

$H_{1.0}$: T-MITOCAR graphs have only little or no semantic similarity to the expert models.

$H_{2.1}$: T-MITOCAR graphs have high structural similarities to the expert models.
$H_{2.0:}$ T-MITOCAR models have only little or no structural similarity to the expert models.

Second, we want to compare the effects of the graphical representations on reading comprehension directly to see whether they have an influence and whether this influence is comparable to the effect that expert models have.

$H_{3.1}$: T-MITOCAR graphs lead to the same performance gain as expert models or more.
$H_{3.0}$: T-MITOCAR graphs lead to less performance gain than expert models.

In a control group we investigated the reading itself without providing any representation. Another control group was presented with a graph which was constructed from the terms but whose relations were completely arbitrary (randomized). With the second control group we wanted to see whether the effects were based on the relational structure of the re-representation or if they could be explained by the availability of the terms only – regardless of how they may have been organized. This allowed us to see how much of the effect was due to the organization of the knowledge:

$H._{4.1}$: T-MITOCAR graphs lead to more performance gain than random graphs and no conceptualizations
$H._{4.0}$: T-MITOCAR graphs lead to the same performance gain as random graphs and no conceptualizations or less

Method

Participants

The experiment was conducted with 60 undergraduate students (34 female and 26 male) from the University of Freiburg. Their mean age was 20.8 years (*SD* = 1.76). They were all students of fields which did not contain any content trained in this experiment. It took the subjects about 1.5 hours to complete the full experiment. They were paid 10 Euros each as compensation for their participation.

Materials

- Three *texts* for the subject domains geodesy, English literature, and pharmacy were provided by three domain experts. Each text was selected to be used for training non-experts on the specific topic. The experts on geodesy and pharmacy chose texts from www.wikipedia.org, the text on literature was taken from Abrams (1993).
- The conceptual graphs (*expert model*) for each subject domain (economy, English literature, and pharmacy) were provided by the domain expert. Each text (economy, English literature, and pharmacy) was processed by T-MITOCAR, which also resulted in a graph (*T-MITOCAR model*). The similarity indices between the expert model and T-MITOCAR model were calculated for each of the three subject domains (see Table 3). Similarity indices are between 0 and 1 ($0 \leq s \leq 1$): 1 is identity and 0 is exclusion. To simplify the reading of the similarity values, the measure of similarity may to some extent be interpreted as being similar to correlations or contingencies (although they may of course not drop below zero).
- *Random models* for each subject domain were created from the most frequent terms. Instead of using meaningful relations, the "propositions" were randomly assigned to pairs of terms. The number of randomly created links was derived on the basis of the distribution of link numbers within the expert models and the T-MITOCAR models. The models were randomized for every participant.
- Test on general *reading comprehension*: The test was constructed on the theoretical basis of Groeben (1992) and Langer, Schulz von Thun, and Tausch (1974). All items on this test are measured on five point Likert scales. The four scales (45 items) of the test are: *simplicity* [12 items], (e.g., ease of reading, Cronbach's alpha = .84); *order* [12 items], (e.g., structure and design, Cronbach's alpha = .94); *length* [13 items], (e.g., appropriateness of length, Cronbach's alpha = .83); *motivational aspects* [8 items], (e.g., mood of the text, writing style acts as stimulant, Cronbach's alpha = .88)

- Three *domain dependent knowledge tests* (economy, English literature, and pharmacy, pretest and posttest versions), each including six multiple-choice questions (higher order) with six alternatives (one correct, five incorrect). The *knowledge gain* is measured as the difference between posttest and pretest in order to account for intra-individual differences (individual gain from reading). Table 9.2 shows one example question from the test for each domain. It contains the correct answer and two of the five incorrect answers.

TABLE 9.2
Example items of the domain dependent knowledge tests

	Item	*Correct answer*	*Incorrect answer (selection)*
Geodesy	Given an average GPS-receiver, why is it very well possible that it shows “- 15m” while you are standing on top of a hill, 40m above sea level?	GPS uses reference ellipsoid, differs from geoid by ± 110m	With GPS, height is measured as "potential energy“, which needs to be translated into "meters above sea level“, which is not possible with absolute accuracy.
English Literature	Which term is related to the convention that the narrator knows everything that needs to be known about the agents, actions, and events and also has privilege to access to the characters’ thoughts, feelings, and motives?	Omniscient point of view	Self-Conscious narrator Self-effacing author
Pharmacy	What is the function of a filler in the manufacturing of tablets?	A filler provides a quantity of materials which can accurately be formed into a tablet.	A filler is added to reduce friction between the tablet and the punches during pressing of the tablet. A filler is used to speed up the disintegration of the tablet in the gastric tract.

Design

The three different subject domains (economy, English literature, and pharmacy) and the four sources of graphical representation (no conceptualization, random model, automated T-MITOCAR model, expert model) resulted in a total of 12 different experimental conditions for the 60 participants in our *Latin square* experimental design. In each experimental condition the participants read the domain dependent text and received a standardized graphical representation from an expert, a random model (including concepts from the subject domain connected randomly), an automated T-MITOCAR model, or no conceptualization .

Procedure

First, every participant completed a domain dependent pretest. After completing the pretest, they received either an expert model, an automated T-MITOCAR model, a random model, or no graphical conceptualization. After five minutes of study time with the graphical representation, the participants read the text. They were given 20 minutes for reading. After the reading, the participants took the reading comprehension test and the domain dependent posttest.

Results

Graphically, the expert models look different from the T-MITOCAR models (see Figure 9.2). The expert uses different shapes, but only to distinguish between the topic and the rest of the content. Some but not all of the links are annotated. Link annotations are partly hierarchical, causal, or procedural/commenting. Also, some but not all of the links have directions. Thus, from a formalistic perspective, the graph would have to be analyzed as a non-hierarchical and undirected graph.

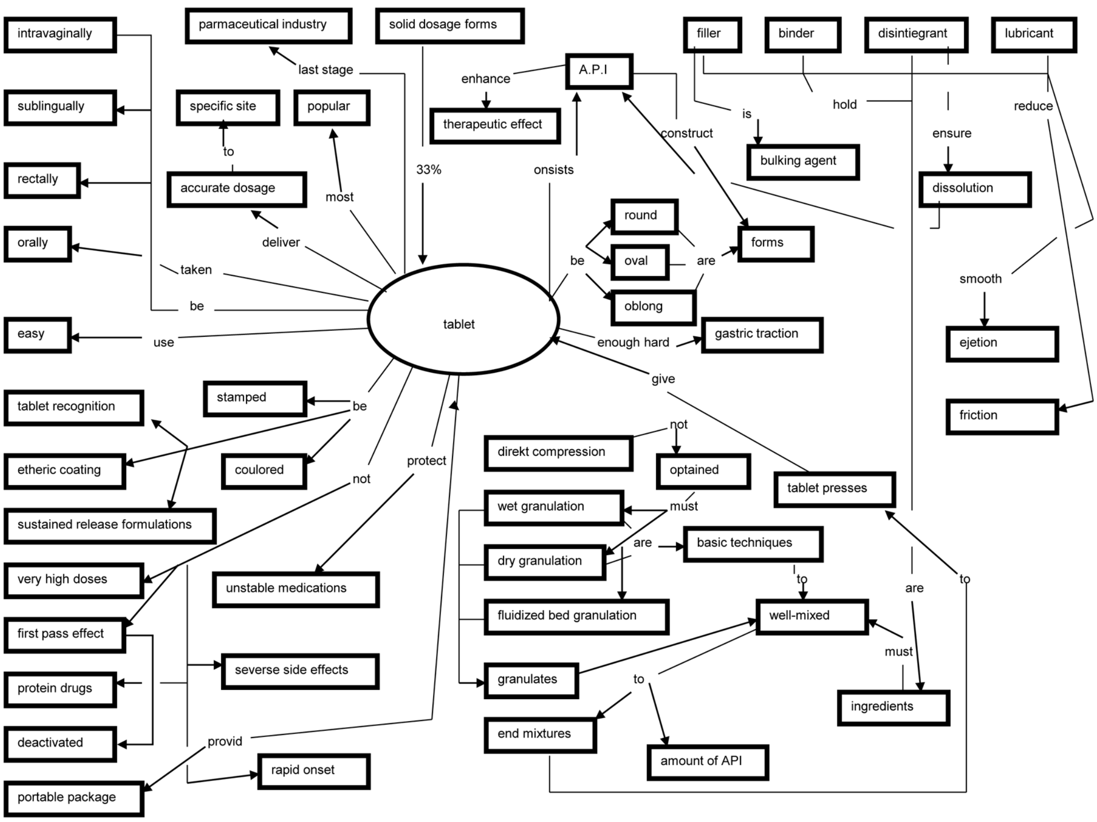

FIGURE 9.2. *Sample graph created by the expert on pharmacy*

To test the first two hypotheses, we calculated the similarity measures. Semantic and structural similarities (relationships) between the expert's model and the T-MITOCAR generated model are shown in Table 9.3. The results can be interpreted in the form of correlations to determine whether a value may be considered to indicate weak, medium, or high similarity (see Williams, 1968, for the interpretation of correlations and Tversky, 1977, for the interpretation of similarities).

Both semantics (concept matching and propositional matching) and structure have high similarities. Only the surface matching values have a medium similarity. All similarity indices are statistically significant on the level of graph-feature comparison (within each model comparison). Therefore we accept $H_{1.1}$: T-MITOCAR graphs have high semantic similarities to the expert models. We can also accept $H_{2.1}$: T-MITOCAR graphs have high structural similarities to the expert models.

TABLE 9.3
Similarity measures between expert graph and T-MITCAR graph

	Matching Index	Pharmacy	Literature	Geodesy	M
Structural Measures	Surface	0.72**	0.60**	0.50**	0.61
	Graphical	1.00**	0.92**	0.70**	0.87
	Structural	0.77**	0.74**	0.92**	0.81
	Gamma	0.96**	0.70**	0.67**	0.78
Semantic Measures	Concepts	0.86**	0.91**	0.58**	0.78
	Propositional	0.84**	0.77**	0.67**	0.76
Overall		0.86	0.77	0.67	0.77

Additionally, we asked the experts who originally provided the expert models whether the T-MITOCAR models represent the content in a good way. Since there were only three experts (one for each domain), there is no systematic way to aggregate the answers reliably.

The *pharmacy* expert said (answer provided in German, translated into English by the authors): "Graphically, the two models do not look alike. However, their content is very similar. My own model is more detailed than the other [T-MITOCAR] model, but the other model is more clearly arranged."

The *literature* expert said (answer provided in English): "The model I provided includes more specific concepts than the other [T-MITOCAR] model. However, the core concepts and most important propositions are also represented in the automatically generated model. It seems to me that this technique could save a lot of time."

The *expert on geodesy* said (answer provided in English): "I was surprised to find most of the core concepts of the matter represented in the automatically generated model. Furthermore, the connections between these concepts are remarkably similar in the automatically generated model and the one made by me. Thus, it seems to me as though both models represent the important information equally well."

Overall, it seems that the experts see a close relationship between the model they constructed on their own and the automatically created T-MITOCAR model. Additionally, the experts pointed out that the associations between individual concepts are correctly represented. The difference between the pretest and the posttest was considered to accurately reflect the performance gain.

There are no meaningful differences between the conditions as regards to performance gain. The differences shown in Table 9.4 are also not statistically significant (ANOVA: $F(3, 176) = 0.2294$, $p > .05$). No pairs have individually significant differences either. Neither the pretest nor the posttest showed any ceiling effects. Ironically, this still corresponds to $H_{3.1}$: T-MITOCAR graphs lead to the same performance gain as expert models or more. Of course this is not the kind of outcome we were expecting. But at least T-MITOCAR graphs do not differ from the expert graphs.

TABLE 9.4
Performance gain within the experimental variation

	No Conceptualization	Random Model	Automated T-MITOCAR Model	Expert Model
M	0.67	0.88	0.67	0.87
SD	1.49	1.80	1.72	1.82

We had to reject $H_{4.1}$ in favor of $H_{4.0}$: T-MITOCAR graphs lead to the same performance gain as random graphs and no conceptualizations or less. However, the text has a high influence on knowledge gain, as can be seen in Table 9.5.

TABLE 9.5
Knowledge gain depending on text/content

	Pharmacy	Literature	Geodesy
M	1.82	1.07	-0.56
SD	1.49	1.31	1.35

This has nothing to do with the fact that reading a text has an influence on learning (which should be obvious because text is the only media in this experiment). Rather, it means that different texts influence learning differently. The performance gain depending on the text is statistically significant (ANOVA: $F(2, 177) = 46.426$, $p < .01$). The text on geodesy caused a systematic knowledge loss. The pharmacy text offered the best chance to increase knowledge. As mentioned above, the tests were constructed by the experts who selected the texts, and they were instructed to create the test items to match the texts. A further analysis did not raise any suspicion that the tests did not correspond sufficiently to the texts.

To account for any possible hidden interaction effects, including effects from the (systematically varied) position of the subject domain and the models, we also conducted a multifactor variance analysis (see Table 9.6).

TABLE 9.6
Multifactor Variance Analysis

	SS	df	F value	p
Modeltype	2.017	3	0.3408	0.7959
Position	5.300	2	13.436	0.2642
Text	159.834	2	405.161	<0.001**
Modeltype:Position	3.312	6	0.2799	0.9457
Modeltype:Text	13.046	6	11.023	0.3639
Position:Text	4.051	4	0.5135	0.7259
Modeltype:Position:Text	25.567	12	10.801	0.3811
Residuals	284.036	144		

As shown in Table 9.6, nothing but the text had an effect on the knowledge gain ($\eta^2 = 0.563$). There were also no interactions between the experimental variation (position as varied by the Latin square design) and the outcome. We also compared the subjective readability of the texts using the above-mentioned four scale test (see Table 9.7).

TABLE 9.7
Subjective mean readability (standard deviations in parenthesis) of the texts

	Pharmacy	Literature	Geodesy
Simplicity	3.33 (0.56)	3.45 (0.51)	2.41 (0.52)
Order / Layout	3.92 (0.61)	3.94 (0.65)	2.79 (0.75)
Length	3.40 (0.57)	3.46 (0.45)	2.58 (0.50)
Motivational Aspects	2.37 (0.74)	2.53 (0.84)	1.71 (0.54)

Whereas the texts on pharmacy and literature were well accepted, the text on geodesy had obvious acceptance problems throughout all scales.

This may explain at least a part of the negative effect the text had on learning. All differences are statistically significant according to an ANOVA (see Table 9.8 for details). There were no factor effects from the type of model presented (no model, random model, T-MITOCAR, and expert model) on the subjective readability ratings. The scale reliabilities within this study were between $\alpha=.84$ and $\alpha=.94$. The position in which a text had been presented during the experiment had an effect on motivation (see Table 9.9).

TABLE 9.8
The influence of the text on the text ratings

Simplicity					
	df	SS	F	p	η^2
Text	2	38.843	69.052***	<2.2e-16	.780
Residuals	177	49.783			
Length					
	df	SS	F	p	η^2
Text	2	28.979	55.978***	<2.2e-16	.633
Residuals	177	45.815			
Order / Design					
	df	SS	F	p	η^2
Text	2	51.451	57.107***	<2.2e-16	.645
Residuals	177	79.734			
Motivation / Stimulation					
	df	SS	F	p	η^2
	2	23.126	22.341***	<2.231e-09	.252
	177	91.608			

Interestingly, the motivational aspects rose during work on the experiment (ANOVA: $F(2, 177) = 3.4074$, $p < 0.5$, $\eta^2 = 0{,}039$). However, the effect is very low and the position did not have effects on any other subjective text ratings (see Table 9.9).

TABLE 9.9
Mean effect (standard deviations in parenthesis) of the position on motivational/stimulant rating of the text

	Position 1	Position 2	Position 3
Motivation / Stimulant	1.99 (0.63)	2.28 (0.76)	2.34 (0.95)

To sum up, we found an overall knowledge gain in the domain dependent multiple choice tests. However, we found no effects indicating that conceptual models support reading comprehension, neither with the T-MITOCAR graphs nor with the expert models.

Discussion

The newly developed T-MITOCAR toolset enables researchers and instructors to convert prose text directly to an association net. The application of T-MITOCAR is also feasible for practitioners. After any text is submitted to the system, the re-representation process is carried out in multiple stages. As a result, the system (1) *provides a list of the most frequent terms*, (2) *displays a thumbnail* and *a full size picture of the graphical model*, (3) *displays the model in list form* and *generates a spreadsheet file for download*, and (4) *allows quantitative pairwise comparisons of two or more models*. The automated quantitative analysis generates six core measures, ranging from surface over structure to semantic indicators (surface, graphical matching, concept matching, density of vertices, structural matching, and propositional matching). With the help of these six indicators, we are able to describe and track changes in students' and experts' representations. An earlier pilot study raised high hopes for the efficiency and feasibility of the T-MITOCAR models for facilitating learning in reading comprehension. Irrespective of which graphical representation was provided (no conceptualization, random model, T-MITOCAR model, expert model), we revealed an overall knowledge gain in the domain dependent multiple choice tests. However, we found no effects in which conceptual models supported reading comprehension, neither with the T-MITOCAR graphs nor with the expert models. However, as we used an expert model constructed by only one expert, this may limit our results on this side. Accordingly, in future studies it could be helpful to ask more than one expert to generate a model, or to ask additional experts rating their colleagues expert model, as we did with the T-MITOCAR models.

The second prediction in Mayer (1989) assumes a reduction of verbatim retention when models are used to support understanding of novice or low achieving learners. However, we could not find this effect in our study. We cannot yet determine whether the models will improve problem-solving transfer either, since we did not incorporate a problem-solving performance test. We will have to address this aspect in a future study, since this may be an important blind spot for the use of T-MITOCAR generated models.

Finally, administering a Latin square experimental design allowed us to control for hidden interaction effects, including the position of the text with foci on different subject domains (geodesy, English literature, pharmacy) and the type of model representation (no conceptualization, random model, T-MITOCAR model, expert model). The only significant effect which influenced the learning outcome was the text. Additional analysis revealed a high acceptance of the *pharmacy* and *English literature* texts, while the text on *geodesy* was not well received by the subjects. The overall motivational rating of the texts rose during our 1.5 hour experiment.

Applications

The T-MITOCAR technology can automatically generate graphs with only the text at hand. These graphs are structurally and semantically very similar to graphs conceptualized by human experts. Irrespective of the subject domain, we found a high similarity between the computer-generated graph and the expert's re-representation. This could still allow a variety of applications. E.g., learners can use them in online learning environments to enhance their text understanding whenever they like.

The technology can be used on any texts or parts of texts to instantly generate a graphical conceptualization. It can also be used by instructors and teachers preparing for class or assignments (or for other homework) with an almost negligible amount of effort. Whereas human experts are not always available for a certain domain, T-MITOCAR can provide the necessary graph any time. Additionally, human experts require an extensive amount of time to re-represent a domain specific expert model. The T-MITOCAR graph thus saves researchers and instructors valuable time. Once our effects have been verified in international studies, the T-MITOCAR technology will be ready for use in learning environments wherever expert models can be implemented to improve the quality of learning. Unfortunately, this does not work with simple text reading.

Future projects

One of the future projects will therefore concentrate on problem-solving transfer and also use a more learner-oriented technology. The technology has already been developed and implemented with interfaces to selected research tools like DEEP, SMD, MITOCAR (Pirnay-Dummer, et al., 2010). When measures are applied to re-representations it helps methodologically to look at them from different perspectives (Jonassen & Cho, 2008). The different effects from the texts still need to be explained. The experts choose the texts by applying the same instructions. The texts all had equal basic layouts and were about the same length. Nonetheless, there have to be identifiable features within the text that explain the differences between the effects. It would be useful to identify these features on the basis of the texts and test them in a further study, also taking a closer look at features of layout, syntax and semantics. This would not only help us to understand the reading comprehension task better but could also provide criteria for text development for learning and instruction.

10

FACILITATING LEARNING THROUGH INDIVIDUALIZED AUTOMATED FEEDBACK 🕮

Feedback is considered an elementary component for supporting and regulating learning processes. Feedback plays a particularly important role in highly self-regulated model-centered learning environments because it facilitates the development of mental models, thus improving expertise and expert performance. In this chapter, different types of model-based feedback are investigated. Seventy-four participants were assigned to three experimental groups in order to examine the effects of different forms of model-based feedback. With the help of seven automatically calculated measures, changes in the participants' understanding of the subject domain "climate change", represented by causal diagrams, are reported. The results strengthen our assumption that the mental model building process for experts and expert performance should be trained in a more direct way, such as with simulation environments.

🕮 This chapter is based on: Ifenthaler, D. (2009). Model-based feedback for improving expertise and expert performance. *Technology, Instruction, Cognition and Learning, 7*(2), 83-101.

Introduction

In the field of learning and instruction, feedback is considered an elementary component for supporting and regulating learning processes. Especially in computer-based and self-regulated learning environments, the nature of feedback is of fundamental importance (Simons & de Jong, 1992). However, the empirical evidence of effects of different types of feedback is rather inconsistent and contradictory in parts (e.g., Bangert-Drowns, et al., 1991; Clariana, 1993; Kluger & DeNisi, 1996; Kulhavy, 1977; Mory, 2004).

In a broader sense, feedback is considered to be any type of information provided to learners (see Wagner & Wagner, 1985). Accordingly, feedback can take on many forms depending on theoretical perspective, the role of feedback, research goals, and methodological approaches. Unlike this initial general understanding of feedback, the term *informative feedback* refers to all kinds of external post-response information used to inform the learner of his or her current state of learning or performance (Narciss, 2006, 2008). Furthermore, from an instructional point of view feedback can be provided by internal (individual cognitive monitoring processes) or external (various types of correction variables) sources of information. Internal feedback may validate the externally provided feedback, or it may lead to resistance against the externally provided feedback (see Narciss, 2008).

Feedback plays a particularly important role in highly self-regulated model-centered learning environments because it facilitates the development of mental models, thus improving expertise and expert performance (Johnson-Laird, 1989; Seel, 2003). However, this requires for the person to be sensitive to characteristics of the provided environment, such as the availability of certain information at a given time, the ease with which this information can be found in the environment, and the way the information is structured and mediated (Ifenthaler & Seel, 2005). Feedback on mental model construction, such as the use of *conceptual models* to help persons to build mental models of the system being studied, has already been investigated and discussed (e.g., Mayer, 1989). Conceptual models highlight the most important objects and associated causal relations of the phenomenon in question. However, not only do new developments in computer technology enable us to dynamically generate simple conceptual models and expert representations; they may also be used to generate direct responses to the learner's interaction with the learning environment. We define this as *model-based feedback*.

In this chapter, different types of model-based feedback generated automatically with our own HIMATT (Highly Integrated Model Assessment Technology and Tools) methodology will be investigated. The following section focuses on mental model development and model-based feedback. In the next section we present our newly developed HIMATT methodology, which enables us to generate different types of model-based feedback on the fly. Then we will describe the research design we used to investigate effects of different types of model-based feedback and present our results. We conclude with a discussion of our findings and suggestions for further development of our approach.

Model building and feedback

Since the beginnings of mental model research (e.g., Gentner & Stevens, 1983; Johnson-Laird, 1983; Seel, 1991) many research studies have provided evidence that "mental models guide and regulate all human perceptions of the physical and social world" (Seel & Dinter, 1995, p. 5). Accordingly, mental models are dynamic ad hoc constructions which provide subjectively plausible explanations on the basis of restricted domain-specific information (Ifenthaler, 2010c). Various research studies have shown that it is very difficult but possible to influence such subjectively plausible mental models by providing specific information (see Anzai & Yokoyama, 1984; Ifenthaler & Seel, 2005; Mayer, 1989; Seel, 1995; Seel & Dinter, 1995). Ifenthaler and Seel (2005) argue that it is important to consider how such information is provided to the learner at specific times during the learning process and how it is structured. In accordance with the general definition of feedback introduced above (see Wagner & Wagner, 1985), such information for improving individual mental model building processes provided purposely and on the fly is referred to as model-based feedback.

The importance of feedback for improving knowledge and skill acquisition has been discussed controversially in educational research (e.g., Azevedo & Bernard, 1995; Bangert-Drowns, et al., 1991; Narciss, 2008; Narciss & Huth, 2004; Shute, 2008). Widely accepted forms of feedback include (a) knowledge of result, (b) knowledge of correct result, (c) knowledge of performance, (d) answer until correct, (e) knowledge of task constraints, (f) knowledge about concepts, (g) knowledge about mistakes, (h) knowledge about how to proceed, and (i) knowledge about metacognition (see Jacobs, 1998; Narciss, 2008). Additionally, Schimmel (1983) found that feedback is most effective under conditions that encourage the learner's conscious reception.

In accordance with empirical findings on feedback (see Schimmel, 1983) and mental model theory (see Ifenthaler, Pirnay-Dummer, & Spector, 2008; Seel, 1991), we argue that effective model-based feedback is composed of externalized representations (re-representations) of mental models. An externalization of a mental model of a learner or expert could be a causal model, concept map, written or spoken text, etc. (Ifenthaler, 2010c). Such externalized representations induce positive effects on internal information processing (see Galbraith, 1999). Additionally, model-based feedback aims at the development of mental models for the improvement of expertise and expert performance (Johnson-Laird, 1989). Accordingly, model-based feedback is highly associated with necessary expertise and expert performance in the specific subject domain.

Past research studies have shown how *conceptual models* (i.e. explicit and consistent causal explanations of a given phenomenon) can be provided to improve a person's understanding of a specific problem in a given context (e.g., Mayer, 1989; Norman, 1983; Seel & Dinter, 1995). However, we argue that model based-feedback should not only include an expert's solution of the given phenomenon. Rather, in order to be more effective the feedback should also take into account the person's prior understanding (initial mental model, preconception), because such preconceptions are in many cases resistant to change as they have a high subjective plausibility (Ifenthaler & Seel, 2005; Seel & Dinter, 1995). In order to fulfill this requirement, we introduce two new forms of model-based feedback in this article: (1) cutaway model-based feedback and (2) discrepancy model-based feedback. These two forms of model-based feedback are considered as graphical re-representations constructed from a set of *vertices* whose relationships are represented by *edges* (Ifenthaler, et al., in press).

The *cutaway model-based feedback* is based on the individual's preconception or on a more elaborated mental model constructed during the learning process. Additionally, an expert's understanding of the phenomenon in question is taken into account. By combining both, the individual's re-representation (preconception) and the expert's re-representation, we create the cutaway model-based feedback re-representation. This re-representation includes all propositions (vertex-edge-vertex) of the individual's re-representation and highlights semantically correct vertices (as compared to the expert's re-representation); see Figure 10.1.

The *discrepancy model-based feedback* is also based on the individual's preconception or on a more elaborated mental model constructed during the learning process. However, it includes only the propositions (vertex-edge-vertex) which have no semantic similarity to the expert's re-representation. Additionally, semantically correct vertices (compared to the expert's re-representation) are highlighted (see Figure 10.1).

Hence, model-based feedback aims at a restructuring of the underlying representations and a reconceptualization of the related concepts (vertices and edges). This is in following with Piaget's epistemology (1950, 1976). New information provided through model-based feedback can be assimilated through the activation of an existing schema, adjustment by accretion, or tuning of existing schema. Otherwise it is accommodated by means of a reorganization process which involves building new mental models (Ifenthaler, et al., in press; Seel, et al., 2009).

In order to fulfill the requirement that model-based feedback be provided to the learner on the fly, it is necessary to implement the cutaway and discrepancy feedback in a computer-based environment. Accordingly, the automated model-based feedback generation is described in the following section.

Automated model-based feedback generation

HIMATT (Highly Integrated Model Assessment Technology and Tools) is a combined toolset conveying the benefits of various methodological approaches in one environment. It is implemented and runs on a Web server using Apache, MySQL, PERL, and additional packages (Pirnay-Dummer, et al., 2010). The HIMATT architecture consists of two major platforms: The HIMATT *Research Engine* (functions for conducting and analyzing experiments) and the HIMATT *Subject Environment* (functions for dynamically providing assigned experiments to individual subjects). Methodologically, the tools integrated into HIMATT touch the boundaries of qualitative and quantitative research methods. Text and conceptual graphs can be analyzed quantitatively with the comparison function of the SMD Technology (Ifenthaler, 2010c) and MITOCAR (Pirnay-Dummer & Ifenthaler, 2010). Additionally, Ifenthaler (Ifenthaler, 2010c) introduced an automated feature of the SMD Technology to generate standardized graphical re-representations of subjects' data with the help of the open source graph visualization software *GraphViz* (Ellson, et al., 2003). This algorithm, the newest add-on to the HIMATT toolset, enables us to generate automated model-based feedback.

The *feedback function* of the SMD Technology (Ifenthaler, 2010c), which we implemented in HIMATT (Pirnay-Dummer, et al., 2010), automatically generates standardized reference (e.g., expert), participant (e.g., learner), cutaway, and discrepancy re-representations. A cutaway re-representation includes all propositions (vertex-edge-vertex) of the individual's re-representation. Additionally, the semantically correct vertices (compared to a reference re-representation, e.g., expert solution) are graphically highlighted as circles (ellipses for dissimilar vertices). The discrepancy re-representation of an individual only includes propositions (vertex-edge-vertex) which have no semantic similarity to a reference re-representation. Additionally, the semantically correct vertices (compared to a reference re-representation) are graphically highlighted as circles (ellipses for dissimilar vertices). Figure 1 shows an example of a reference (1), participant (2), cutaway (3), and discrepancy (4) re-representation.

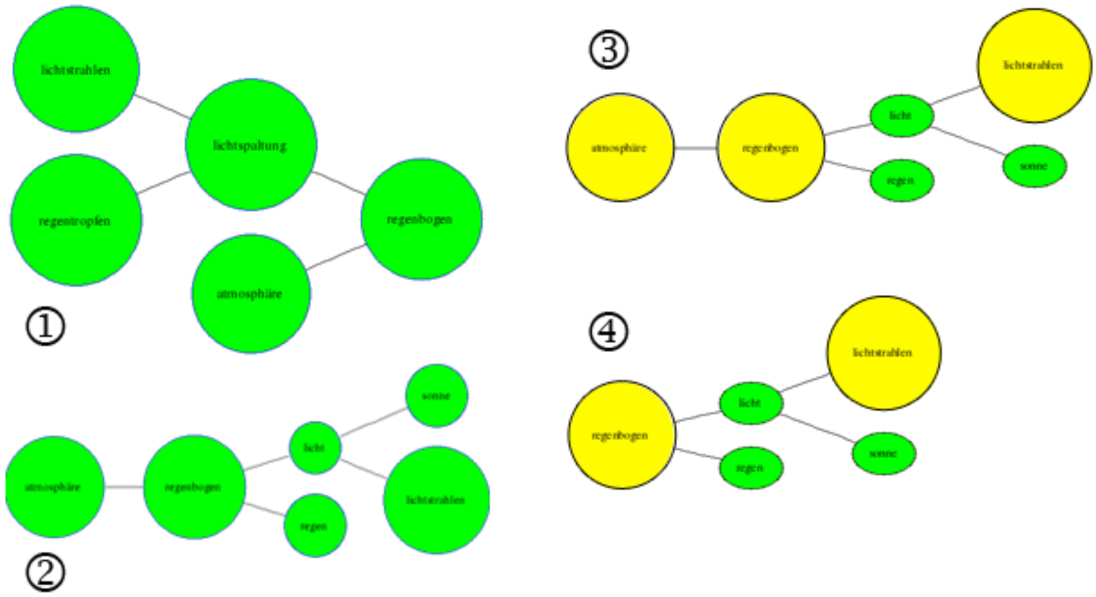

FIGURE 10.1. *Reference, subject, cutaway, and discrepancy re-representations*

These automated and standardized re-representations are generated on the fly while participants work within the HIMATT environment. They are then used for individual model-based feedback during work on a learning task.

The reference model (1) represents a best practice solution by an expert for the task to be completed. The participant's model (2) is a solution found after a specified time of work on the task. With the reference (1) and participant (2) models at hand, HIMATT automatically generates the cutaway (3) and discrepancy (4) feedback models. The cutaway model allows the learner to see how many vertices are semantically correct (graphically highlighted circles compared to the expert solution). Additionally, the cutaway model provides information about the semantically incorrect vertices (ellipses). The discrepancy model only provides information about the semantically incorrect propositions compared to the expert solution (vertex-edge-vertex). Additionally, semantically correct vertices are highlighted. We argue that either feedback model (3) or (4) will have different effects when presented during the learning process. As the cutaway feedback model (3) helps to confirm the correct understanding of the phenomenon in question (compared with an expert), the discrepancy feedback model (4) causes a cognitive conflict, because correct propositions (vertex-edge-vertex) of the person's understanding are deleted from the re-representation.

Each of the above described feedback models could help to improve expertise and expert performance in various subject domains. Therefore, we conducted an experimental study to investigate the effects of different types of model-based feedback. The research questions of this empirical investigation are as follows.

Research questions

Feedback plays a particularly important role in highly self-regulated model-centered learning environments because it facilitates the development of mental models, thus improving expertise and expert performance (see Ifenthaler & Seel, 2005). Past research studies demonstrate how *conceptual models* can be provided to improve a person's understanding of a specific problem in a given context (e.g., Mayer, 1989; Norman, 1983; Seel & Dinter, 1995). Conversely, model-based feedback includes not only a conceptual or expert solution to the given phenomenon; it also includes the person's prior understanding (initial mental model, preconception). Therefore, we introduced two forms of model-based feedback: (1) cutaway model-based feedback and (2) discrepancy model-based feedback. Accordingly, our first research question investigated in this chapter is:

Does model-based feedback (cutaway and discrepancy) facilitate the understanding of a specific phenomenon in question?

Since it is possible to generate different forms of model-based feedback, we wanted to investigate which form of feedback is most accepted among participants. Thus, our second research question investigated in this article is:

Do participants valuate the forms of model-based feedback differently?

Additionally, previous research studies (e.g., Hilbert & Renkl, 2008; Ifenthaler, et al., 2007) have found that verbal and spatial abilities do not affect the quality of model-building processes and declarative learning outcomes. Therefore, a third and last research question to be explored in this article is:

Do verbal and spatial abilities affect the declarative learning outcome and the quality of model-building processes?

Method

Participants

Seventy-four students (66 female and 8 male) from the University of Freiburg, Germany, participated in the study. Their average age was 21.9 years (*SD* = 2.3). The participants were randomly assigned to the three experimental groups (1) cutaway feedback (*n* = 26), (2) discrepancy feedback (*n* = 24), and (3) expert feedback (*n* = 24).

Materials

- A German-language article on climate change (Schönwiese, 2005) with 1,417 words was used as learning content.
- HIMATT causal diagram and text input tools were used to assess the participants' understanding of the subject domain climate change. First, the participants constructed a causal diagram using vertices and edges in order to describe the phenomenon of climate change. Secondly, they had to write a text about their understanding of climate change. The causal diagrams and texts of all participants were stored in the HIMATT database for further analysis.
- Two subsets of the I-S-T 2000 R (Amthauer, et al., 2001) were used to test the participants' verbal and spatial abilities. This test is a widely used intelligence test in Germany with high reliability (*r* = .88 and *r* = .96; split-

half reliability). The first subset we used tested the verbal abilities of the participants. A total of 20 sentences with a missing word had to be completed using a set of five words. The participants had six minutes to complete this subset. The second subset tested spatial abilities. Within nine minutes, the participants had to choose similar cubes from a set of five by rotating them. Subset two included 20 cube problems.

- The participants' experience with concept mapping and causal diagrams was tested with a questionnaire including eight items (Cronbach's alpha = .87). The questions were answered on a five-point Likert scale (1 = totally disagree; 2 = disagree; 3 = partially agree; 4 = agree; 5 = totally agree).
- The domain specific knowledge test included 27 multiple-choice questions on climate change. In a pilot study with 5 female und 5 male participants (average age 26.3 years, SD = 3.49), we tested the average difficulty level in order to account for ceiling effects. The participants scored 10.5 out of 27 possible points on average (SD = 3.54, Min = 5, Max = 17). In our experiment we administered two versions (in which the 27 multiple-choice questions appeared in a different order) of the domain-specific knowledge test (pre- and posttest). It took about 10 minutes to complete the test.
- The feedback model quality test consisted of nine items on whether the provided feedback model helped the participant to understand the text better (Cronbach's alpha = .66). The questions were answered on a five-point Likert scale (1 = totally disagree; 2 = disagree; 3 = partially agree; 4 = agree; 5 = totally agree).

Procedure

First, the participants completed a demographic data questionnaire. Secondly, they completed the concept map and causal diagram experience questionnaire. Next, the participants completed the test on verbal (six minutes) and spatial abilities (nine minutes). Then they answered the 27 multiple choice questions of the domain specific knowledge test on climate change (pretest). After a short relaxation phase, the participants were given an introduction to concept maps and causal diagrams and were shown how to use the HIMATT software. Then, the participants used the username and password they had been assigned to log in to the HIMATT system, where they constructed a causal diagram on their understanding of climate change (ten minutes). Immediately afterwards, they wrote a text about their understanding of climate change (ten minutes). A short relaxation phase followed, during which we automatically generated the individual feedback models for each participant. After that, the participants received the text on climate change and the automatically generated feedback model (cutaway, discrepancy, or expert model – depending on the assigned experimental group). All three types of feedback models were automatically generated with HIMATT. The *cutaway feedback model* (see Figure 10.2) included all propositions (vertex-edge-vertex) of the participant's pre-test causal diagram. Additionally the semantically correct vertices (compared to the expert re-representation) were graphically highlighted (circles are semantically correct to the expert; ellipsis are semantically incorrect compared to the expert re-representation). The *discrepancy feedback model* included only propositions (vertex-edge-vertex) of the participant's pre-test causal diagram which had no semantic similarity compared to the expert re-representation. The *expert feedback model* consisted of a standardized re-representation of an expert on climate change. The participants had 15 minutes to read the text and examine their feedback model. Immediately after working on the text, the participants completed the model feedback quality test.

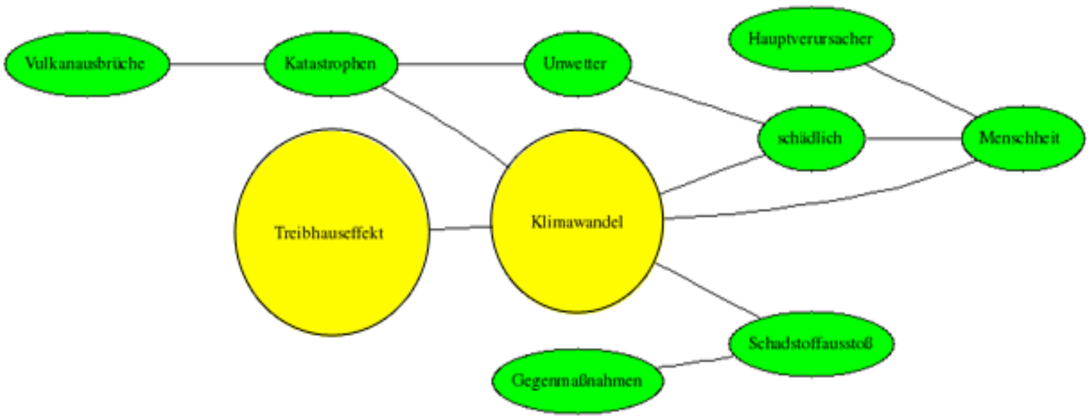

FIGURE 10.2. *Example of an automatically generated cutaway feedback model used in our experiment*

Then they answered the 27 multiple choice questions of the posttest on declarative knowledge. After another short relaxation phase, the participants used their username and password to log in to the HIMATT system for the second time. In the HIMATT posttest, they constructed a second causal diagram on their understanding of climate change (ten minutes) and wrote a second text regarding their understanding of climate change (ten minutes). Finally, the participants had to complete a short usability test regarding HIMATT features.

Analysis

To analyze the causal diagrams constructed by the participants in the HIMATT environment, we used the seven core measures implemented in HIMATT (Pirnay-Dummer, et al., 2010). Figure 10.3 shows the seven measures of HIMATT, which include four structural and three semantic indicators.

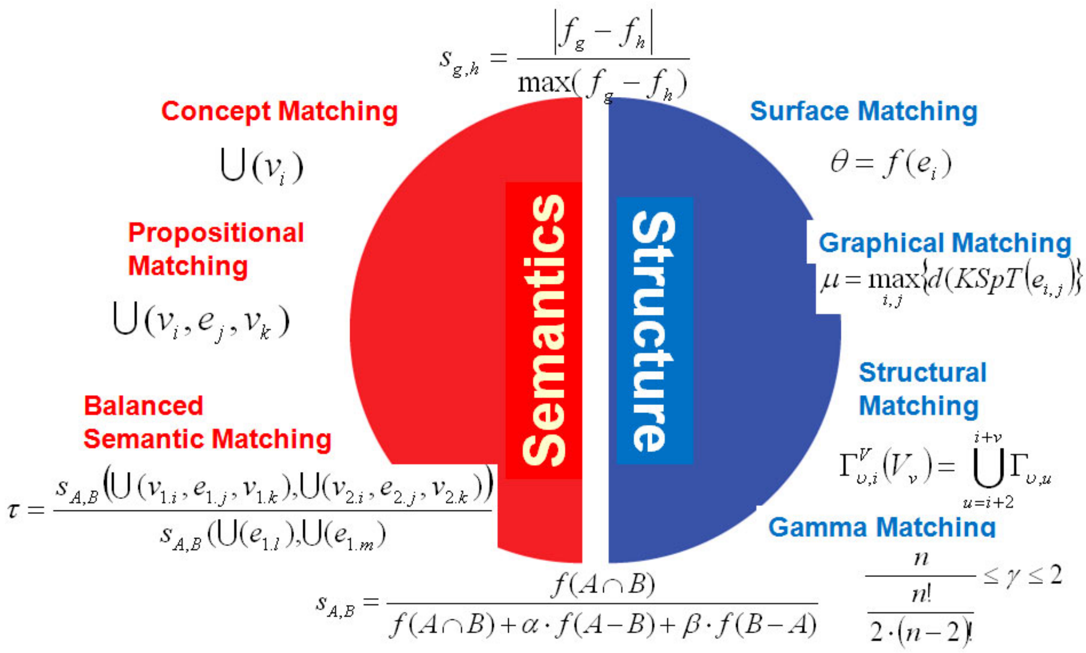

FIGURE 10.3. *HIMATT measures*

These seven measures are defined as follows (see Ifenthaler, 2006, 2010c, 2010d; Pirnay-Dummer, et al., 2010):

Surface Matching: The surface measure compares the number of vertices within two graphs. It is a simple and easy way to calculate values for surface complexity.

Graphical Matching: The graphical matching compares the diameters of the spanning trees of the graphs, which is an indicator for the range of conceptual knowledge. It corresponds to structural matching as it is also a measure for structural complexity only.

Structural Matching: The structural matching compares the complete structures of two graphs without regard to their content. This measure is necessary for all hypotheses which make assumptions about general features of structure (e.g., assumptions which state that expert knowledge is structured differently from novice knowledge).

Gamma Matching: The gamma or density of vertices describes the quotient of terms per vertex within a graph. Since both graphs which connect every term with each other term (everything with everything) and graphs which only connect pairs of terms can be considered weak models, a medium density is expected for most good working models.

Concept Matching: Concept matching compares the sets of concepts (vertices) within a graph to determine the use of terms. This measure is especially important for different groups which operate in the same domain (e.g. using the same textbook). It determines differences in language use between the models.

Propositional Matching: The propositional matching value compares only fully identical propositions between two graphs. It is a good measure for quantifying semantic similarity between two graphs.

Balanced Propositional Matching: The balanced propositional matching index is the quotient of propositional matching and concept matching.

Results

Over two-thirds of the participants (68%) did not use concept maps or causal diagrams to structure their own learning materials before our experiment. Only 12% of the participants used concept mapping software to create their own concept maps before. On the other hand, over 40% of the participants answered that they did not find it difficult to create a concept map or causal diagram. Consequently, there was no significant difference in the learning outcome as measured by the domain-specific knowledge posttest between participants who used concept mapping software before the experiment and those who did not use concept mapping software at all, $t(72) = .508$, *ns*.

Domain specific knowledge

On the domain specific knowledge test (pre- and posttest), participants could score a maximum of 27 correct answers. In the pretest they scored an average of $M = 7.78$ correct answers ($SD = 2.10$) and in the posttest $M = 18.16$ correct answers ($SD = 3.80$). The increase in correct answers was significant, $t(73) = 28.32$, $p < .001$, $d = 3.096$ (strong effect). The cutaway feedback group ($M = 10.88$, $SD = 3.32$) outperformed the discrepancy ($M = 10.42$, $SD = 2.92$), and expert group ($M = 9.79$, $SD = 3.23$) concerning their knowledge gain. However, these differences were not significant.

Verbal and spatial abilities

Participants could score a maximum of 20 points in both subsets of the I-S-T 2000 R on verbal and spatial abilities. On the test for verbal abilities, participants scored M = 12.76 points (SD = 3.66) and on the test for spatial abilities they scored M = 10.39 points (SD = 3.15). As reported in Table 1, we found no significant correlations between the seven HIMATT measures and verbal and spatial abilities. However, the higher the learners' spatial abilities were, the higher was their increase on the domain specific knowledge test (see Table 10.1).

TABLE 10.1
Correlations between learning outcomes, HIMATT similarity measures, and verbal and spatial abilities

	Verbal abilities	Spatial abilities
Domain specific knowledge increase	.108	.290*
Surface Matching	-.075	.051
Graphical Matching	-.213	-.139
Structural Matching	-.028	.056
Gamma Matching	.057	-.063
Concept Matching	-.139	-.004
Propositional Matching	.011	.130
Balanced Propositional Matching	-.004	.177

Note. * $p < .05$

Quality of feedback models

An explorative factorial analysis (varimax rotation) was carried out by means of selected variables of the *feedback model quality test* (see Table 10.2).

TABLE 10.2
Factor analysis component matrix for nine items of the *quality of feedback models* instrument (N = 72)

Nr	Item (translated from German)	Factor 1	Factor 2
1	The model is clearly laid out.	**.787**	.212
2	The model is well-structured.	**.733**	-.261
3	The concepts in the model are comprehensible.	**.725**	
4	The links between the concepts are comprehensible.	**.663**	
5	The model helped me understand the text.	**.640**	-.371
6	The model uses many unfamiliar concepts.		**.767**
7	The model is complex.		**.757**
8	The model confused me.	.345	**.612**
9	I would not understand the text without the model.	.389	**.449**

Note. Factor loading < .2 are suppressed

The two extracted factors represent 54% of the variance. The first factor is determined by five items. Consequently, the first factor represents *clarity of the feedback model* (Cronbach's α = .756). Factor two represents *support through the feedback model* (Cronbach's α = .595) and is determined by four items (see Table 10.2). The two factors *clarity of feedback model* and *support of feedback model* were entered into a one-way ANOVA in order to test for differences between the three experimental groups (cutaway feedback, discrepancy feedback, and expert feedback). The ANOVA revealed a significant effect for the factor *support of feedback*, $F(2, 69) = 4.22$, $p = .019$, $\eta^2 = .11$. Accordingly, participants with *discrepancy feedback* ($M = 4.08$, $SD = .70$) rated the support of the feedback model highest (*cutaway feedback*: $M = 3.81$, $SD = .56$; *expert feedback*: $M = 3.55$, $SD = .59$). The ANOVA indicated no further significant effects.

Quality of re-representations (HIMATT measures)

The graphical re-representations of the participants were analyzed automatically with the HIMATT analysis feature. Hence, we computed the knowledge gain of the seven HIMATT measures by subtracting the pre- from the post measure. Table 10.3 shows the average gain of the HIMATT measures (surface, graphical, structural, gamma, concept, propositional, and balanced propositional matching) for the three experimental groups (cutaway feedback, discrepancy feedback, and expert feedback).

TABLE 10.3
Average gain of HIMATT measures for the three experimental groups ($N = 74$)

	Cutaway feedback ($n = 26$)	*SD*	Discrepancy feedback ($n = 24$)	*SD*	*Expert feedback (n = 24)*	*SD*
Surface Matching	1.731	3.779	3.375	2.871	4.826	4.579
Graphical Matching	-.192	1.497	.875	1.985	1.609	1.438
Structural Matching	1.231	3.766	2.583	1.213	3.087	2.353
Gamma Matching	.005	.099	-.001	.142	-.019	.155
Concept Matching	.052	.074	.020	.067	.-010	.109
Propositional Matching	.007	.027	.006	.026	-.001	.002
Balanced Propositional Matching	-.008	.091	.000	.044	-.009	.079

The results showed a significant effect between participants in the three experimental groups for the HIMATT measure *Surface Matching*, $F(2, 70) = 4.080$, $p = .021$, $\eta^2 = .10$, with participants of the *expert feedback group* increasing their number of vertices higher than the other experimental groups. The one-way ANOVA also revealed a significant effect for the HIMATT measure *Graphical Matching*, $F(2, 70) = 7.355$, $p = .001$, $\eta^2 = .17$. The increase of complexity of participants was higher in the *expert feedback group* than in the others. Between the experimental groups, the increase of the HIMATT measure *Structural Matching* was also significant, $F(2, 70) = 3.140$, $p = .049$, $\eta^2 = .08$. Again, the participants in the *expert feedback group* outperformed the other experimental groups. For the semantic HIMATT measure *Concept Matching* we found a final significant effect, $F(2, 70) = 3.243$, $p = .045$, $\eta^2 = .08$. Here, participants in the *cutaway feedback group* gained more correct concepts than the participants in the other two groups. However, we found no further effects for the HIMATT measures.

Discussion

The large body of theoretical and empirical studies on feedback provides very diverse insight into possible ways to support and regulate learning processes. Even meta-analyses (Azevedo & Bernard, 1995; Kluger & DeNisi, 1996; Schimmel, 1983) have provided contradictory results. However, feedback is considered to be an elementary component for facilitating learning outcomes. As feedback can take on many forms depending on the theoretical perspective, the role of feedback, and the methodological approach, it is important to consider which form of feedback is right for a specific learning environment.

The aim of our study was to examine different forms of model-based feedback for improving expertise. Hence, we introduced two new forms of model-based feedback, which we defined as (1) cutaway model-based feedback and (2) discrepancy model-based feedback. As we were able to generate the model-based feedback automatically and on the fly, the participants received the model-based feedback just after finishing their pre-test, which served to motivate them further. Additionally, our HIMATT analysis features enabled us to score the participants solution automatically within an instant. Not only do these automated process have very high objectivity, reliability, and validity (Pirnay-Dummer, et al., 2010), they are also very economical, especially when large sets of data need to be analyzed within a short period of time (Ifenthaler, 2010c).

An explorative factorial analysis of our newly developed instrument for identifying the quality of the model-based feedback found two factors. Our subsequent analysis of the factors *clarity of feedback* and *support of feedback* showed that learners rated the discrepancy feedback as being most supportive. Thus, by providing propositions which have no semantic similarity compared to an expert's representation we were able to bring about the intended cognitive conflict (accommodation processes) and induce a reorganization of the participants' cognitive structures (Piaget, 1976; Seel, 1991). From the participant's perspective, simply receiving an expert solution as feedback seemed less helpful.

With the help of our seven automatically calculated HIMATT measures, we were able to investigate changes in the participants' understanding of the subject domain "climate change" and re-represent them with causal diagrams. Participants who received the expert feedback added significantly more relations to their causal diagrams (*Surface Matching*) than did those in the other groups. Accordingly, the expert feedback provided them a broad spectrum of concepts and relations, which were then integrated into their own understanding of the phenomenon in question. This also explains the significant differences between the measures *Graphical* and *Structural Matching*. As the number of relations of a causal diagram increases, there is also a high probability that its complexity and complete structure will also increase.

However, an increase in these structural measures does not necessarily mean that the solutions of participants in the expert feedback group are better than these of the other participants. As a further analysis of the semantic HIMATT measures revealed, participants in the cutaway feedback group outperformed the other participants with regard to their semantic understanding of the phenomenon in question (*Concept Matching*). Accordingly, even if the structure increases, the semantic correctness of the learner will not automatically also increase. Hence, learners may integrate a huge amount of concepts into their understanding of the phenomenon which do not necessarily help them to come to a better and more correct solution to the problem.

Therefore, a further empirical investigation will focus on participants' misconceptions (e.g., Ifenthaler & Seel, 2005) and how they can be influenced by model-based feedback. Another study will investigate the similarities and differences between causal diagrams and natural language texts written on the same subject domain, “climate change.” Our hypothesis is that causal diagrams and texts do represent different forms of knowledge. However, this does not necessarily lead to the conclusion that one of these forms of assessment (causal diagram or text) is obsolete for identifying expertise and expert performance. Rather, we argue that both graphical and textual re-representations are needed to better understand the underlying cognitive processes of learning-dependent progression from novice to expert and, as a consequence, to provide more effective feedback and instructional materials.

As in a previous study (Ifenthaler, et al., 2007), intellectual abilities (verbal and spatial abilities) were not found to have an effect on the mental model building process. Only for spatial abilities did we find a positive correlation with the participants’ learning outcome. This result was also found in a study by Hilbert and Renkl (2008). Accordingly, when we train learners to become experts, we should not limit our focus to general abilities such as learning strategies and intellectual abilities. For expert performance it is far more important to train mental model building processes which enable persons to act and decide within complex domains. This strengthens our assumption that the mental model building process for experts and expert performance should be trained in a more direct way, such as with simulation environments (Dörner & Wearing, 1995; Ifenthaler, et al., 2007).

In further studies we will focus on the learning trajectories while providing forms of model-based feedback. This will give us more detailed insight into the effects of model-based feedback and how it helps to support and improve expertise and expert performance.

11

EPILOGUE

The epilogue will highlight some ongoing projects which are based on the so far acquired scientific knowledge on cognitive structure. Combining the theoretical and empirical knowledge on cognitive structure with new technological developments of the 21st century opens up new fields of research and instruction. First, AKOVIA (Automated Knowledge Visualization and Assessment) is presented as a consequent further development of the tools described above (e.g., SMD, HIMATT). Second, a new experimental research program is presented which addresses an extended longitudinal perspective. Third, a research program investigating emotions and the development of cognitive structures is introduced. Finally, two tools for an automated feedback generation (TASA and iGRAF) are highlighted.

Essentials of cognitive structures

Much effort was devoted to the development of a theoretical foundation of cognitive structures (e.g., Jonassen, 1987; Jonassen, et al., 1993), mental models (Dinter, 1993; Gentner & Stevens, 1983; Johnson-Laird, 1989; Norman, 1983; Seel, 1991), and schemata (Bransford, 1984; Rumelhart, 1980; Rumelhart, et al., 1986), as well as to their instructional application (e.g., Anzai & Yokoyama, 1984; Ifenthaler, et al., in press; Mayer, 1989; Seel, 1995, 2003). However, there are still a number of concerns as to their validity, i.e., which form of expression (visual or contextual - descriptive) better represents what one comprehends from a learning environment (Ifenthaler, 2008; Ifenthaler & Seel, 2005; Seel, 1999a).

One essential question concerning the assessment of cognitive structure is which methodology should be used, one that uses visual representation (i.e., concept map) or one that consists of a written text (i.e., a summary). Many authors consider concept maps to be an adequate format of externalization for analyzing complex knowledge structures (T. E. Johnson, et al., 2009; Novak, 1998). Concept maps seem preferable to classical knowledge tests, such as multiple-choice tests for the purpose of representing linked knowledge by means of network-like visualization. On the other hand, there are strong arguments indicating that natural language representations are a good method for assessing cognitive structures (Ifenthaler, 2008; Pirnay-Dummer & Ifenthaler, in press).

Various approaches and empirical studies enabling an insight into cognitive structure by addressing the above mentioned assessment and analysis issues have been presented (Al-Diban & Ifenthaler, in press; Ifenthaler, 2010c, 2010d, accepted; Ifenthaler, et al., in press; Pirnay-Dummer, et al., 2010). Further empirical studies investigated instructional innovations which may foster learning and therefore possibly changing underlying cognitive structures (Ifenthaler, 2009; Ifenthaler & Seel, in press; Pirnay-Dummer & Ifenthaler, in press). However, these empirical investigations do not mark the end of this challenging research program on cognitive structure. If anything, it is a first tiny step for an ongoing research endeavor in the 21st century.

Pursuing the insight into cognitive structure

The following sections highlight some ongoing projects which are based on the so far acquired scientific knowledge on cognitive structure. Combining the theoretical and empirical knowledge on cognitive structure with new technological developments of the 21st century opens up new fields of research and instruction (Ifenthaler, 2010b). First, AKOVIA (Automated Knowledge Visualization and Assessment) is presented as a consequent further development of the tools described above (e.g., SMD, HIMATT). Second, a new experimental research program is presented which addresses an extended longitudinal perspective. Third, a research program investigating emotions and the development of cognitive structures is introduced. Finally, two tools for an automated feedback generation (TASA and iGRAF) are highlighted.

AKOVIA

Although HIMATT (Highly Integrated Model Assessment Technology and Tools) has already been used by several researchers, it has two design problems worth mentioning. On the one hand, the user interface was accepted by researchers and subjects alike, and it even had a good usability (Pirnay-Dummer, et al., 2010). On the other hand, it was a web service which integrated both the data collection and the analysis. Researchers understandably wanted to integrate the data collection into their experiments and studies. However, subjects needed to log into HIMATT in order to input their data as text or draw graphs. They needed to enter another login, username, and password, which might have disturbed the experimental setting in some cases. The second design problem results from the first: We were often given raw data to upload into the HIMATT system so that the researchers could use the analysis facilities on their data. After following this procedure more often than the system had been used through the "front door," it was time for a complete redesign of the blended methods.

Based on our experience with the HIMATT framework, the diagnostic toolset is taken one step further and developed AKOVIA (Automated Knowledge Visualization and Assessment). Instead of limiting the framework to a narrow set of data collection procedures, the development focuses on the implementation of more interfaces to different methods. The core analysis in AKOVIA is a comprehensive blend of MITOCAR, T-MITOCAR, and the SMD Technology. Thus, it is also based strictly on mental model theory (Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Seel, 1991, 2003). The results of the analysis are unchanged. However, the input formats and outputs have been changed to better accommodate the needs of researches, thus allowing more applications as in the original technologies and HIMATT.

AKOVIA offers several different analysis tools which were initially developed for different purposes and integrates them into a single framework to obtain a more comprehensive perspective on the knowledge externalizations under analysis.

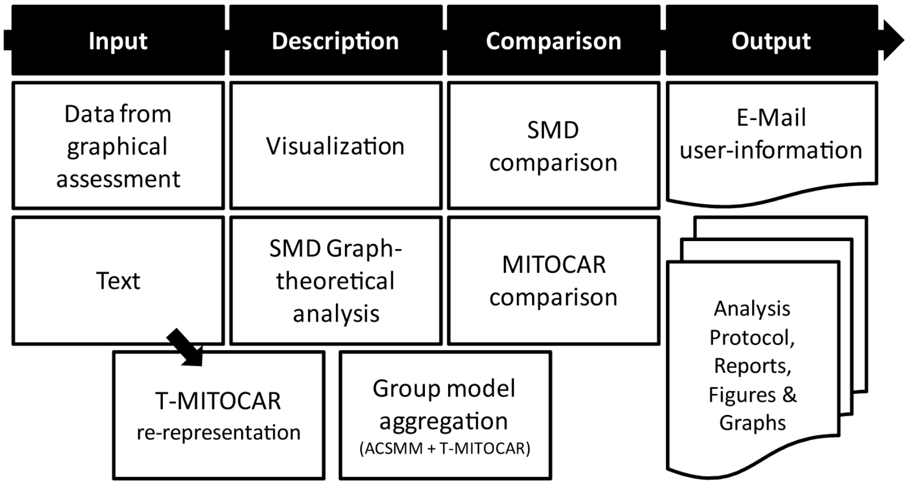

FIGURE 11.1. *AKOVIA framework (Pirnay-Dummer & Ifenthaler, 2010)*

Figure 11.1 provides an overview on the modules of AKOVIA. There are two general input formats (text and graph). Thus, the software can be used to analyze many currently available assessment methods. A standard interface may be used for graphical methods. This interface is derived from SMD and HIMATT and uses the list form. Specific interfaces are under construction. The software can visualize, aggregate, describe in detail, and compare the models. The measures from SMD and MITOCAR are embedded and available for use, as are the text to graph algorithms from T-MITOCAR. The availability of AKOVIA will provide researchers a simple to use toolset for a large set of research designs.

Longitudinal perspective

In previous research high fluctuations in the probability of change in solving logical reasoning tasks have been found (Ifenthaler, et al., in press; Ifenthaler & Seel, in press). This result corresponds largely with the theory of mental models (Johnson-Laird, 1989), where mental models are defined as ad hoc constructions which a person builds over and over again while solving new and unfamiliar problems. However, in this previous research not any evidence could be found for the emergence and consolidation of a cognitive schema during the time series measurements (Ifenthaler & Seel, in press). Based on the results of this study the investigation of model-based reasoning over an expanded period of time is extended, i.e. in total 20 measurements.

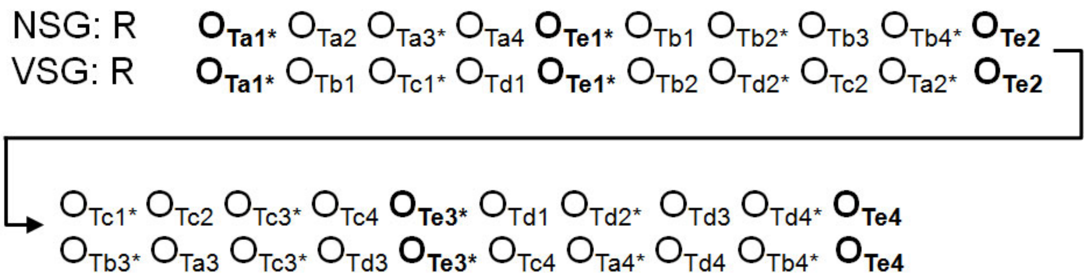

FIGURE 11.2. *Longitudinal research design*

Figure 11.2 shows the longitudinal design of the current study which enables a precise assessment across a total of 20 points of measurement. A computer-based multimedia learning environment has been created with a large set of different tasks. Participants were randomly assigned to two different experimental groups (NSG: *non-varying strategy* vs. VSG: *varying strategy*). The performance (applied strategy and solution) of the participants is measured for each of the 20 tasks (Ta, Tb, Tc, Td, Te being different task classes; * being feedback in form of a correct solution after solving the task). This extended research design may give a better insight into the development of cognitive structures in problem-solving situations.

Emotions

Besides the above discussed cognitive foundations of mental models and schemas, it is argued that emotional and motivational experiences have a major impact on the learning-dependent progression of cognitive structures due to the fact that whenever assimilation in a schema fails, this schema enters a state of disequilibrium which in turn evokes arousal. The term “motive” can be used to denote the presence of disequilibrium. Whenever an attempt at assimilation fails and corrective attempts are not immediately successful, a motive will be originated. This argumentation follows Berlyne’s (1971) views on the central role of arousal in curiosity motivation and active stimulus seeking. High levels of incongruity are innately aversive.

Indeed, emotions are mental states which arise spontaneously rather than through conscious effort. A growing body of empirical studies shows that information processing is highly related with emotional experiences (e.g., Gray, 2001; Isen, 1999; K. C. Klauer & von Hecker, 2009; Kuhl, 1983, 2000). According to Goetz, Preckel, Pekrun, and Hall (2007), emotions can be differentiated into present emotional experiences (state-emotions; e.g. "I am anxious at this moment") and emotional experiences that occur consistently in specific situations (trait-emotions; e.g. "I am generally anxious while taking math exams"). Kuhl (1983) introduced a model of emotional emergence where cognitive, emotional, and operational processes are reciprocal affect another. Accordingly, cognitive processes and the reciprocal interactions with emotional states are the basis for goal-directed actions (Gross, 1998). More specifically, positive emotions promote the activation of schemas and mental models, whereas negative emotions restrict these activating functions. Baumann and Kuhl (2002) showed that learners in sad mood performed worse while solving tasks than those who were able to regulate negative emotions. Alternatively, positive emotional experiences may increase the learner's optimism and confidence and thus facilitate the construction of mental models or application of alternative schemata.

In light of these observations, it is assumed that while measuring the learning-dependent progression of model-based reasoning and their associated emotional experiences will improve the understanding of these complex cognitive functions. As a result, instructional materials and instructor feedback that are most appropriate at various times during the learning process may be identified.

Intelligent feedback

Research studies have shown that it is very difficult but possible to influence cognitive structures by providing specific information (see Anzai & Yokoyama, 1984; Ifenthaler, et al., in press; Mayer, 1989; Pirnay-Dummer & Ifenthaler, in press; Seel, 1995; Seel & Dinter, 1995). Ifenthaler and Seel (2005) argue that it is important to consider how such information is provided to the learner at specific times during the learning process and how it is structured. In accordance with the general definition of feedback introduced above (Wagner & Wagner, 1985), an important aspect of *model-based feedback* is providing dynamic feedback generated purposively and individually to student-constructed models (Ifenthaler, 2009).

Intelligent model-based feedback helps students to monitor their individual learning process. Automated knowledge assessment tools provide the basis to produce instant feedback on semantic and structural aspects of a person's learning progression at all times during the learning process (Ifenthaler, 2009). Such dynamic and timely feedback can promote the learner's self-regulated learning (Zimmerman & Schunk, 2001). Based on these new technologies, two intelligent and automated model-based feedback tools have been developed and implemented: TASA (Text-Guided Automated Self Assessment), which generates automated feedback to learners based on natural language text input (Pirnay-Dummer & Ifenthaler, 2011). iGRAF (Instant Graphical Feedback) automatically generates graphical representations based on the prior knowledge of the learner (Ifenthaler, 2009, 2010a).

The main limitations for TASA so far are on the volitional level. Hence, future studies will concentrate on this aspect and also consider several covariates on the learners' side. With the additional data at hand, we should be able to make the tool more stimulating. TASA is applicable to any learning task which involves writing. It may be used for short writing assignments. However, its strength clearly unfolds in long-term writing assignments, in which the students may continuously monitor their own progress and make their own decisions when using the automated tool.

The graphical feedback produced with iGRAF proved to facilitate self-regulated learning. However, no systematic effect of the various forms of model-based feedback could be found yet. However, the overall effectiveness of feedback generated with iGRAF shows high potential. Already available empirical evidence on the facilitation of self-regulated learning processes through intelligent model-based feedback (TASA and iGRAF) provides high hopes for future developments and practical implications. Therefore, model-based feedback will guide a promising voyage towards the world of learning within Web 3.0 (Ifenthaler, 2010b; Ifenthaler & Seel, 2010b).

Technology, Instruction, Cognition, and Learning

In our digital age, technology, instruction, cognition, learning, and educational diagnostics are closely linked (Ifenthaler, 2010d; Ifenthaler, Isaias, Spector, Kinshuk, & Sampson, 2009; Ifenthaler & Seel, 2010a, 2010b). Researchers and engineers have always endeavoured to design and develop useful diagnostic systems to serve professional communities in the field of learning and instruction, and they will continue to do so (Ifenthaler, 2010b). Future work on automated computational diagnostics, including approaches such as graph theory, will provide more and more powerful dynamic systems for the comprehensive analysis of large amounts of data in a short space of time.

References

Abrams, M. H. (1993). *A glossary of literary terms*. Fort Worth, TX: Harcourt Brace College Publishers.

Acton, W. H., Johnson, P. J., & Goldsmith, T. E. (1994). Structural knowledge assessment: Comparison of referent structures. *Journal of Educational Psychology, 86*(2), 303-311.

Al-Diban, S. (2002). *Diagnose mentaler Modelle*. Hamburg: Verlag Dr. Kovac.

Al-Diban, S. (2008). Progress in the diagnosis of 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* (pp. 81-102). New York: Springer.

Al-Diban, S., & Ifenthaler, D. (in press). Comparison of two analysis approaches for measuring externalized mental models: Implications for diagnostics and applications. *Journal of Educational Technology & Society*.

Al-Diban, S., & Stark, A. (2007). *Pflichtenheft zur Graph to Context (GTC) Schnittstelle*. Dresden: Technische Universität.

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *I-S-T 2000 R*. Göttingen: Hogrefe.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anzai, Y., & Yokoyama, T. (1984). Internal models in physics problem solving. *Cognition and Instruction, 1*(4), 397-450.

Ausubel, D. P. (1963). Cognitive structure and the facilitation of meaningful verbal learning. *Journal of Teacher Education, 14*, 217-221.

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*(2), 111-127.

Baalmann, W. (1997). Schülervorstellungen zur Evolution. In H. E. Bayrhuber (Ed.), *Biologieunterricht und Lebenswirklichkeit* (pp. 163-167). Kiel: IPN.

Baird, J. R., & White, R. T. (1982). A case study of learning styles in biology. *International Journal of Science Education, 4*(3), 325-337.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction. An introduction to sequential analysis*. Cambridge, MA: Cambridge University Press.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events *Review of Educational Research, 61*(2), 213-238.

Bartholomew, D. J. (1967). *Stochastic models for social processes*. New York: Wiley.

Baumann, N., & Kuhl, J. (2002). Intuition, affect, and personality: Unconscious coherence judgements and self-regulation of negative affect. *Journal of Personality and Social Psychology, 83*, 1213-1223.

Bayrhuber, H. E. (2001). *Biowissenschaft in Schule und Öffentlichkeit*. Kiel: IPN.

Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts.

Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology, 57*(3), 195-203. doi: 10.1037/h0034701

Birkhoff, G. (1973). *Lattice theory*. Providence, RI: American Mathematical Society.

Bliss, J. (1996). Piaget und Vygotsky: Ihre Bedeutung für das Lehren und Lernen der Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften, 2*(3), 3-16.

Bollobàs, B. (1998). *Modern graph theory*. New York: Springer.

Bonato, M. (1990). *Wissenstrukturierung mittels Struktur-Lege-Techniken. Eine grapentheoretische Analyse von Wissensnetzen*. Frankfurt am Main: Lang.

Bonatti, L. (1994a). Propositional reasoning by model? *Psychological Review, 101*(4), 725-733.

Bonatti, L. (1994b). Why should we abandon the mental logic hypothesis? *Cognition, 50*(1-3), 17-39.

Borkenau, P., & Ostendorf, F. (2006). *NEO-Fünf-Faktoren-Inventar*. Göttingen: Hogrefe.

Bransford, J. D. (1984). Schema activation versus schema acquisition. In R. C. Anderson, J. Osborn & R. Tierney (Eds.), *Learning to read in American schools: Basal readers and content texts* (pp. 259-272). Hillsdale, NJ: Lawrence Erlbaum.

Brill, E. (1995). *Unsupervised learning of dismabiguation rules for part of speech tagging.* Paper presented at the Second Workshop on Very Large Corpora, WVLC-95, Boston.

Cañas, A. J., Hill, R., Carff, R., Suri, N., Lott, J., Eskridge, T., et al. (2004). CmapTools: A Knowledge Modeling and Sharing Environment. In A. J. Cañas, J. D. Novak & F. M. González (Eds.), *Concept Maps: Theory, Methodology, Technology, Proceedings of the First International Conference on Concept Mapping* (pp. 125-133). Pamplona: Universidad Pública de Navarra.

Chartrand, G. (1977). *Introductory graph theory*. New York: Dover.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55-81.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391-426.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121-152.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 1-75). Hillsdale, NJ: Lawrence Erlbaum.

Christmann, U., & Groeben, N. (1999). Psychologie des Lesens. In B. Franzmann, K. Hasemann, D. Löffler & E. Schön (Eds.), *Handbuch Lesen* (pp. 145-223). München: Saur.

Chung, G. K. W. K., & Baker, E. L. (2003). An exploratory study to examine the feasibility of measuring problem-solving processes using a click-through interface. *Journal of Technology, Learning and Assessment, 2*(2), Available from http://www.jtla.org.

Clariana, R. B. (1993). A review of multiple-try feedback in traditional and computer-based instruction. *Journal of Computer-Based Instruction, 20*(3), 67-74.

Clariana, R. B., & Wallace, P. E. (2007). A computer-based approach for deriving and measuring individual and team knowledge structure from essay questions. *Journal of Educational Computing Research, 37*(3), 211-227.

Clement, J. (1981). Student's preconceptions in introductory mechanics. *American Association of Physics Teachers, 50*(1), 66-71.

Coffey, J. W., Carnot, M. J., Feltovich, P. J., Feltovich, J., Hoffman, R. R., Cañas, A. J., et al. (2003). A summary of literature pertaining to the use of concept mapping techniques and technologies for education and performance support. Pensacola, FL: Chief of Naval Education and Training.

Collins, L. M., & Sayer, A. G. (Eds.). (2001). *New methods for the analysis of change*. Washington, DC: American Psychological AssocITation.

Couné, B., Hanke, U., Ifenthaler, D., & Seel, N. M. (2004). Modellkonstruktionen beim Problemlösen im Kontext entdeckenden Lernens: Eine Studie zur Implementierung theoretisch-begründeter Instruktionsprinzipien. Zweiter Bericht aus dem Forschungsprojekt „Modell-begründetes Lernen und Lehren. Multimediale Lernumgebungen als Gelegenheiten zum Nachdenken. Freiburg: Institut für Erziehungswissenschaft.

Courant, R., & Robbins, H. (2000). *Was ist Mathematik?* Berlin: Springer.

Craik, K. J. W. (1943). *The nature of explanation*. Cambridge, UK: Cambridge University Press.

Crinon, J., & Legros, D. (2002). The semantic effects of consulting a textual database on rewriting. *Learning and Instruction, 12*(6), 605-626.

Csapo, B. (1997). The development of inductive reasoning: Cross-sectional assessments in an educational context. *International Journal of Behavioral Development, 20*(4), 609-626.

Davis, E. (1990). *Representations of commonsense knowledge*. San Mateo, CA: Morgan Kaufmann.

de Corte, F., Greer, B., & Verschaffel, L. (1996). Mathematics teaching and learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 491-549). New York: Macmillan.

de Vries, E. (2006). Students' construction of external representations in design-based learning situations. *Learning and Instruction, 16*(3), 213-227. doi: 10.1016/j.learninstruc.2006.03.006

Derbentseva, N., Safayeni, F., & Cañas, A. J. (2004). Experiments on the Effects of Map Structure and Concept Quantification during Concept Map Construction. In A. J. Cañas, J. D. Novak & F. M. González (Eds.), *Concept Maps: Theory, Methodology, Technology, Proceedings of the First International Conference on Concept Mapping* (pp. 125-132). Pamplona: Universidad Pública de Navarra.

Diestel, R. (2000). *Graph theory*. New York: Springer.

Ding, Y. (2001). A review of ontologies with the semantic web in view. *Journal of Information Science, 27*(6), 377-384.

Dinter, F. R. (1993). *Mentale Modelle als Konstrukt der empirischen Erziehungswissenschaft*. Saarbrücken: Universität Dissertation.

Donovan, M. S., & Bransford, J. D. (Eds.). (2005). *How students learn. History, mathematics, and science in the classroom*. Washington, D.C.: The National Academic Press.

Dörner, D., & Wearing, A. (1995). Complex problem solving: Toward a (computersimulated) theory. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 65-99). Hillsdale, NJ: Lawrence Erlbaum.

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

Durso, F. T., & Coggins, K. A. (1990). Graphs in social and psychological sciences: Empirical contributions to Pathfinder. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 31-51). Norwood, NJ: Ablex Publishing Corportion.

Eckert, A. (2000). Die Netzwerk-Elaborierungs-Technik (NET)—Ein computerunterstütztes Verfahren zur Diagnose komplexer Wissensstrukturen. In H. Mandl & F. Fischer (Eds.), *Wissen sichtbar machen—Wissensmanagement mit Mapping-Techniken* (pp. 138-157). Göttingen: Hogrefe.

Eliaa, I., Gagatsisa, A., & Demetriou, A. (2007). The effects of different modes of representation on the solution of one-step additive problems. *Learning and Instruction, 17*(6), 658-672.

Ellson, J., Gansner, E. R., Koutsofios, E., North, S. C., & Woodhull, G. (2003). *GraphViz and Dynagraph. Static and dynamic graph drawing tools*. Florham Park, NJ: AT&T Labs.

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher, 18*(4), 4-10.

Ennis, R. H. (1990). The extent to which critical thinking is subject-specific: Further clarification. *Educational Researcher, 19*(13), 13-16.

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

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life. *Mind, Culture, and Activity, 5*(3), 178-186.

Eschenhagen, D., Kattmann, U., & Rodi, D. (2008). *Fachdidaktik Biologie*. Köln: Aulis Verlag Deubner.

Feeney, A., & Heit, E. (Eds.). (2007). *Indictive reasoning: Experimental, developmental, and computational approaches*. New York: Cambridge University Press.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378-382.

Frazier, L. (1999). *On sentence interpretation*. Dordrecht: Kluwer.

Funke, J. (1990). Systemmerkmale als Determinanten des Umgangs mit dynamischen Systemen. *Sprache & Kognition, 9*(3), 143,153.

Funke, J. (1991). Solving complex problems: Exploration and control of complex problems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185-222). Hillsdale, NJ: Lawrence Erlbaum.

Funke, J. (1992). *Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung*. Berlin: Springer.

Funke, J., & Frensch, P. A. (1995). Complex problem solving research in North America and Europe: An integrative review. *Foreign Psychology, 5*, 42-47.

Galbraith, D. (1999). Writing as a knowledge-constituting process. In M. Torrance & D. Galbraith (Eds.), *Knowing what to write. Conceptual processes in text production* (pp. 139-160). Amsterdam: University Press.

Ganter, B., & Wille, R. (1996). *Formale Begriffsanalyse. Mathematische Grundlagen*. Berlin: Springer.

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

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306-355.

Glaser, R. (1999). Expert knowledge and processes of thinking. In R. McCormick & C. Paechter (Eds.), *Learning and knowledge* (pp. 88-102). Thousand Oaks, CA: Sage Publications.

Goetz, T., Preckel, F., Pekrun, R., & Hall, N. C. (2007). Emotional experiences during test taking. Does cognitive ability make a difference? *Learning and Individual Differences, 17*, 3-16.

Gray, J. R. (2001). Emotional modulation of cognitive control. *Journal of Experimental Psychology: General, 130*, 436-452.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrica, 25*, 95-112.

Greeno, J. G. (1989). Situations, mental models and generative knowledge. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing* (pp. 285-318). Hillsdale, NJ: Lawrence Erlbaum.

Groeben, N. (1992). *Leserpsychologie: Textverständnis - Textverständlichkeit*. Münster: Aschendorff.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*(3), 271-299.

Gruber, H. (1994). *Expertise*. Opladen: Westdeutscher Verlag.

Gruber, H., & Ziegler, A. (1996). *Expertiseforschung. Theoretische und methodische Grundlagen*. Opladen: Westdeutscher Verlag.

Gunstone, R. F. (1980). Word association and the description of cognitive structure. *Research in Science Education, 10*, 45-53.

Harary, F. (1974). *Graphentheorie*. München: Oldenbourg.

Hardy, I., & Stadelhofer, B. (2006). Concept Maps wirkungsvoll als Strukturierungshilfen einsetzen. Welche Rolle spielt die Selbstkonstruktion? *Zeitschrift für Pädagogische Psychologie, 20*(3), 175-187.

Harris, C. W. (Ed.). (1963). *Problems in measuring change*. Madison, WI: The University of Wisconsin Press.

Hasberg, W. (2001). *Empirische Forschung in der Geschichtsdidaktik*. Neuried: ars una.

Hayes, B. K., & Thompson, S. P. (2007). Causal relations and feature similarity in children's inductive reasoning. *Journal of Experimental Psychology: General, 136*(3), 470-484. doi: 10.1037/0096-3445.136.3.470

Heit, E. (1998). A bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 248-274). Oxford: Oxford University Press.

Herl, H. E., Baker, E. L., & Niemi, D. (1996). Construct validation of an approach to modeling cognitive structure of U.S. history knowledge. *Journal of Educational Research, 89*(4), 206-218.

Hietaniemi, J. (2008). Graph-0.84 Retrieved 06-05-2008, from http://search.cpan.org/~jhi/Graph-0.84/lib/Graph.pod

Hilbert, T. S., & Renkl, A. (2008). Concept mapping as a follow-up strategy to learning from texts: what characterizes good and poor mappers? *Instructional Science, 36*, 53-73.

Holland, J., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Holley, K. (2009). The challenge of an interdisciplinary curriculum: a cultural analysis of a doctoral-degree program in neuroscience. *Higher Education, 58*(2), 241-255. doi: 10.1007/s10734-008-9193-6

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps. Analogy in creative thought*. Cambridge, MA: MIT Press.

Hox, J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.

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

Ifenthaler, D. (2008). Practical solutions for the diagnosis of progressing 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* (pp. 43-61). New York: Springer.

Ifenthaler, D. (2009). Model-based feedback for improving expertise and expert performance. *Technology, Instruction, Cognition and Learning, 7*(2), 83-101.

Ifenthaler, D. (2010a). Bridging the gap between expert-novice differences: The model-based feedback approach. *Journal of Research on Technology in Education, 43*(2), 103-117.

Ifenthaler, D. (2010b). Learning and instruction in the digital age. In J. M. Spector, D. Ifenthaler, P. Isaías, Kinshuk & D. G. Sampson (Eds.), *Learning and instruction in the digital age: Making a difference through cognitive approaches, technology-facilitated collaboration and assessment, and personalized communications* (pp. 3-10). New York: Springer.

Ifenthaler, D. (2010c). 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. (2010d). Scope of graphical indices in educational diagnostics. In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 213-234). New York: Springer.

Ifenthaler, D. (accepted). Identifying cross-domain distinguishing features of cognitive structures. *Educational Technology Research and Development*.

Ifenthaler, D., Isaias, P., Spector, J. M., Kinshuk, & Sampson, D. G. (2009). Editors' introduction to the special issue on cognition & learning technology. *Educational Technology Research and Development, 57*(6), 721-723. doi: 10.1007/s11423-009-9127-8

Ifenthaler, D., Masduki, I., & Seel, N. M. (in press). 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. (2009). Assessment of knowledge: Do graphical notes and texts represent different things? In M. R. Simonson (Ed.), *Annual proceedings of selected research and development papers presented at the national convention of the Association for Educational Communications and Technology (32nd, Louisville, KY, 2009). Volume 2* (pp. 86-93). Bloomington, IN: AECT.

Ifenthaler, D., & Pirnay-Dummer, P. (2010a). Artefacts of thought: Properties and kinds of re-representations. In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 75-76). New York: Springer.

Ifenthaler, D., & Pirnay-Dummer, P. (2010b). Using knowlege to support knowing. In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-based* *diagnostics and systematic analysis of knowledge* (pp. 259-260). New York: Springer.

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*(4), 353-366.

Ifenthaler, D., Pirnay-Dummer, P., & Spector, J. M. (Eds.). (2008). *Understanding models for learning and instruction. Essays in honor of Norbert M. Seel*. New York: Springer.

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.

Ifenthaler, D., & Seel, N. M. (2010a). Online-Lernen im Unterricht. *Schulmagazin 5-10, 12*, 11-14.

Ifenthaler, D., & Seel, N. M. (2010b). Online-Lernen in der Schule. *Schulmagazin 5-10, 12*, 7-10.

Ifenthaler, D., & Seel, N. M. (in press). A longitudinal perspective on inductive reasoning tasks. Illuminating the probability of change. *Learning and Instruction*. doi: 10.1016/j.learninstruc.2010.08.004

Iggers, G. G. (1996). *Geschichtswissenschaft im 20. Jahrhundert*. Göttingen: Vandenhoeck und Ruprecht.

Isen, A. M. (1999). Positive affect. In T. Dalgleish & M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 521-539). John Wiley & Sons: Chichester.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

Jacobs, B. (1998). Aufgaben stellen und Feedback geben Retrieved 06-10, 2008, from http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/index.htm

Jacobson, M. J., & Archodidou, A. (2000). The design of hypermedia tools for learning: Fostering conceptual change and transfer of complex scientific knowledge. *Journal of the Learning Sciences, 9*(2), 145-199.

Jech, T. (2007). *Set theory*. New York: Springer.

Jensen, F. V. (2001). *Bayesian networks and decision graphs*. New York: Springer.

Johnson, J., McKee, S., & Vella, A. (Eds.). (1994). *Artificial intelligence in mathematics*. New York: Oxford University Press.

Johnson, T. E., Ifenthaler, D., Pirnay-Dummer, P., & Spector, J. M. (2009). Using concept maps to assess individuals and team in collaborative learning environments. In P. L. Torres & R. C. V. Marriott (Eds.), *Handbook of research on collaborative learning using concept mapping* (pp. 358-381). Hershey, PA: Information Science Publishing.

Johnson, T. E., O'Connor, D. L., Spector, J. M., Ifenthaler, D., & Pirnay-Dummer, P. (2006). Comparative study of mental model research methods: Relationships among ACSMM, SMD, MITOCAR & DEEP methodologies. In A. J. Cañas & J. D. Novak (Eds.), *Concept maps: Thery, methodology, technology. Procedings of the Second International Conference on Concept Mapping, Voume 1* (pp. 87-94). San José: Universidad de Costa Rica.

Johnson-Laird, P. N. (1983). *Mental models. Towards a cognitive science of language, inference, and consciousness*. Cambridge, UK: Cambridge University Press.

Johnson-Laird, P. N. (1989). Mental models. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 469-499). Cambridge, MA: MIT Press.

Johnson-Laird, P. N., & Byrne, R. (1991). *Deduction*. Hove: Lawrence Erlbaum.

Jonassen, D. H. (1987). Assessing cognitive structure: Verifying a method using pattern notes. *Journal of Research and Development in Education, 20*(3), 1-14.

Jonassen, D. H. (1988). Designing structured hypertext and structuring access to hypertext. *Educational Technology, 28*(11), 13-16.

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research & Development, 48*(4), 63-85. doi: 10.1007/BF02300500

Jonassen, D. H. (2009). Externally modeling mental models. In L. Moller, J. B. Huett & D. Harvey (Eds.), *Learning and instructional technologies for the 21st century. Visions of the future* (pp. 49-74). New York: Springer.

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

Jonassen, D. H., & Cho, Y. H. (2008). Externalizing mental models with mindtools. In D. Ifenthaler, P. Pirnay-Dummer & J. M. Spector (Eds.), *Understanding models for learning and instruction. Essays in honor of Norbert M. Seel* (pp. 145-160). New York: Springer.

Jonassen, D. H., Reeves, T. C., Hong, N., Harvey, D., & Peters, K. (1997). Concept mapping as cognitive learning and assessment tools. *Journal of Interactive Learning Research, 8*(3/4), 289-308.

Kalyuga, S. (2006a). Assessment of learners' organised knowledge structures in adaptive learning environments. *Applied Cognitive Psychology, 20*, 333-342.

Kalyuga, S. (2006b). Rapid assessment of learners' proficiency: A cognitive load approach. *Educational Psychology, 26*(6), 735-749.

Kalyuga, S. (2006c). Rapid cognitive assessment of learners' knowledge structures. *Learning and Instruction, 16*(1), 1-11. doi: 10.1016/j.learninstruc.2005.12.002

Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigeluth (Ed.), *Instructional-design theories and models. An overview of their current status* (pp. 383-434). Hillsdale, NJ: Lawrence Erlbaum.

Kirwan, B., & Ainsworth, L. K. (1992). *A Guide to task analysis*. London: Taylor & Francis Group.

Kitcher, P. (1983). *The nature of mathematical knowledge*. Oxford: Oxford University Press.

Klauer, K. C., & von Hecker, U. (2009). Gedächtnis und Emotion. In V. Brandstätter & J. H. Otto (Eds.), *Handbuch der Allgemeinen Psychologie: Motivation und Emotion* (pp. 661-667). Göttingen: Hogrefe.

Klauer, K. J. (1996). Teaching inductive reasoning: some theory and three experimental studies. *Learning and Instruction, 6*(1), 37-57. doi: 10.1016/S0959-4752(96)80003-X

Kleinert, E. (2005). Drei Studien zur Struktur der Mathematik. *Hamburger Beiträge zur Mathematik, 229*, 1-66.

Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.

Koubek, R. J., Clarkston, T. P., & Calvez, V. (1994). The training of knowledge structures for manufacturing tasks: An empirical study. *Ergonomics, 37*(4), 765-780.

Koubek, R. J., & Mountjoy, D. N. (1991). *Toward a model of knowledge structure and a comparative analysis of knowledge structure measurement technique*. West Lafayette, IN: Purdue University.

Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61*(2), 179-211.

Kruskal, J. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometric Monographes, 29*, 115-129.

Ku, W. A. (2007). *Using concept maps to explore the conceptual knowledge of technology students: an exploratory study*. doctoral dissertation. Ohio State University. Columbus, OH.

Kuhl, J. (1983). Emotion, Kognition und Motivation. I: Auf dem Weg zu einer systemtheoretischen Betrachtung der Eomtionsgenese. *Sprache und Kognition, 2*, 1-27.

Kuhl, J. (2000). A functional-design approach to motivation and self-regulation: The dynamics of personality systems interaction. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 111-169). San Diego, CA: Academic Press.

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction, 9*(4), 285-327.

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*(2), 211-232.

Langer, I., Schulz v. Thun, F., & Tausch, R. (1974). *Verständlichkeit in der Schule, Verwaltung, Politik und Wissenschaft*. München: Reinhardt.

Le Ny, J.-F. (1993). Wie kann man mentale Repräsentationen repräsentieren? In J. Engelkamp & T. Pechmann (Eds.), *Mentale Repräsentation* (pp. 31-39). Bern: Huber.

Lee, Y., & Nelson, D. (2004). *Instructional Use of Visual Representations of Knowledge.* Paper presented at the Society for Information Technology and Teacher Education International Conference 2004, Atlanta, GA, USA.

Lehrer, R., & Romberg, T. (1996). Exploring children´s data modeling. *Cognition and Instruction, 14*(1), 69-108.

Lesh, R., & Doerr, H. M. (2000). Symbolizing, communicating, and mathematizing: Key components of models and modeling. In P. Cobb, E. Yackel & K. McClain (Eds.), *Symbolizing and communicating in mathematics classrooms. Perspectives on discourse, tools, and instructional design* (pp. 361-383). Mahwah, NJ: Lawrence Erlbaum Associates.

Lewin, K. (1922). Das Problem der Wissensmessung und das Grundgesetz der Assoziation. Teil 1. *Psychologische Forschung, 1*(191-302).

Lienert, G. A., & Raatz, U. (1994). *Testaufbau und Testanalyse*. Weinheim: Beltz.

Lin, D. (1998). An information-theoretic definition of similarity. In J. W. Shavlik (Ed.), *Proceedings of the fifteenth international conference on machine learning* (pp. 96 - 304). San Francisco, CA: Morgan Kaufmann Publishers Inc.

Magnani, L., & Nersessian, N. (Eds.). (2002). *Model-based reasoning: Science, technology, values*. Dordrecht: Kluwer.

Mandl, H., Gruber, H., & Renkl, A. (1995). Mental models of complex systems: When veridicality decreases functionality. In C. Zucchermaglio, S. Bagnara & S. U. Stucky (Eds.), *Organizational learning and technological change* (pp. 102-111). Berlin: Springer.

Mansfield, H., & Happs, J. (1991). Concept maps. *Australian Mathematics Teacher, 47*(3), 30-33.

Mayer, R. E. (1989). Models for understanding. *Review of Educational Research, 59*(1), 43-64.

Mayer, R. E., & Greeno, J. G. (1972). Structural differences between learning outcomes produced by different instructional methods. *Journal of Educational Psychology, 63*(2), 165-173.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communication by minimizing cognitive load. *Journal of Educational Psychology, 91*(4), 638-643.

McCoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*(3), 440-466.

McNamara, T. P. (1992). Priming and constraints it places on theories of memory and retrieval. *Psychological Review, 99*(4), 650-662.

McNamara, T. P. (1994). Priming and theories of memory: A reply to Ratcliff and McCoon. *Psychological Review, 101*(1), 185-187.

McPeck, J. E. (1990). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher, 19*(10), 10-12.

Mikkilä-Erdmann, M., Penttinen, M., Anto, E., & Olkinuora, E. (2008). Constructing mental models during learning from science text. Eye tracking methodology meets conceptual change. In D. Ifenthaler, P. Pirnay-Dummer & J. M.

Spector (Eds.), *Understanding models for learning and instruction. Essays in honor of Norbert M. Seel* (pp. 63-79). New York: Springer.

Minsky, M. (1981). A framework for representing knowledge in mind design. In R. J. Brachmann & H. J. Levesque (Eds.), *Readings in knowledge representation* (pp. 245-262). Los Altos, CA: Morgan Kaufmann.

Mintzes, J. J., Yen, C., & Barney, E. C. (2008). Assessing knowledge, attitudes, and behavior towards charismatic megafauna. The case of dolphins. *Journal of Environmental Education, 36*(2), 41-55.

Mirow, J. (1991). Geschichtswissen durch Geschichtsunterricht? Historische Kenntnisse und ihr Erwerb innerhalb und außerhalb der Schule. In B. von Borries, H. Pandel & J. Rüsen (Eds.), *Geschichtsbewußtsein empirisch* (pp. 53-109). Pfaffenweiler: Centaurus-Verlagsgesellschaft.

Moeira, M. A. (1983). Assessment of content and cognitive structures in physics at college level. *Assessment & Evaluation in Higher Education, 8*(3), 234-245.

Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 745-783). Mahwah, NJ: Lawrence Erlbaum.

Moskowitz, D. S., & Hershberger, S. L. (Eds.). (2002). *Modelling intraindividual variability with repeated measures data*. Mahwah, NJ: Lawrence Erlbaum.

Nägler, G., & Stopp, F. (1996). *Mathematik für Ingenieure und Naturwissenschaftler. Graphen und Anwendungen*. Stuttgart: Teubner.

Narciss, S. (2006). *Informatives tutorielles Feedback. Entwicklung- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse*. Münster: Waxmann.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. van Merrienboer & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 125-143). New York: Taylor & Francis Group.

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, D. Leutner & R. Brünken (Eds.), *Instructional design for multimedia learning* (pp. 181-195). Münster: Waxmann.

Nason, A., & Goldstein, P. (1969). *Biology; introduction to life*. Menlo Park, CA: Addison-Wesley.

Navicon. (2000). Cernato 2.1. Begriffliche Wissensverarbeitung. Frankfurt: Navicon GmbH.

Nikitina, S. (2005). Pathways of interdisciplinary cognition. *Cognition and Instruction, 23*(3), 389-425. doi: 10.1207/s1532690xci2303_3

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning *Psychological Review, 90*(4), 339-363. doi: 10.1037/0033-295X.90.4.339

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 7-14). Hilsdale, NJ: Lawrence Erlbaum Associates.

Norman, D. A., Gentner, D. R., & Stevens, A. L. (1976). Comments on learning schemata and memory representation. In D. Klahr (Ed.), *Cognition and Instruction* (pp. 177-196). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Nückles, M., Gurlitt, J., Pabst, T., & Renkl, A. (2004). *Mind Maps und Concept Maps. Visualisieren - Organisieren - Kommunizieren*. München: DTV.

O'Donnell, A. M., Dansereau, D. F., & Hall, R. H. (2002). Knowledge maps as scaffolds for cognitive processing. *Educational Psychology Review, 14*, 71-86.

Pandel, H. (1987). Dimensionen des Geschichtsbewusstseins. Ein Versuch, seine Struktur für Empirie und Pragmatik diskutierbar zu machen. *Geschichtsdidaktik, 12*(2), 130-142.

Pape, M. (2006). Methodische Zugangsweisen zur Erfassung von Geschichtsbewusstsein im Kindesalter: Gruppendiskussionen und Kinderzeichnungen. In G. Hilke & M. Sauer (Eds.), *Geschichtsdidaktik empirisch - Untersuchungen zum historischen Denken und Lernen* (pp. 85-110). München: LIT Verlag.

Penner, D. E. (2001). Cognition, computers, and synthetic science: Building knowledge and meaning through modeling. *Review of Research in Education, 25*, 1-35.

Piaget, J. (1943). *Le developpement mental de l'enfant*. Zürich: Rascher.

Piaget, J. (1950). *La construction du réel chez l'enfant*. Neuchatel: Delachaux et Niestlé S.A.

Piaget, J. (1972). *Das mathematische Denken*. Stuttgart: Klett.

Piaget, J. (1976). *Die Äquilibration der kognitiven Strukturen*. Stuttgart: Klett.

Pirnay-Dummer, P. (2006). *Expertise und Modellbildung: MITOCAR*. Freiburg: FreiDok.

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. (2011). Text-guided automated self assessment. A graph-based approach to help learners with ongoing writing. In D. Ifenthaler, Kinshuk, P. Isaias, D. G. Sampson & J. M. Spector (Eds.), *Multiple perspectives on problem solving and learning in the digital age* (pp. 217-225). New York: Springer.

Pirnay-Dummer, P., & Ifenthaler, D. (in press). Reading guided by automated graphical representations: How model-based text visualizations facilitate learning in reading comprehension tasks. *Instructional Science*. doi: 10.1007/s11251-010-9153-2

Pirnay-Dummer, P., Ifenthaler, D., & Rohde, J. (2009). Text-guided automated self-assessment. In Kinshuk, D. G. Sampson, J. M. Spector, P. Isaias & D. Ifenthaler (Eds.), *Proceedings of the IADIS international conference on cognition and exploratory learning in the digital age* (pp. 311-316). Rome: IADIS Press.

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. doi: 10.1007/s11423-009-9119-8

Pollio, H. R. (1966). *The structural basis of word association behavior*. The Hague: Mouton.

Preece, P. F. W. (1976). Mapping cognitive structure: A comparison of models. *Journal of Educational Psychology, 68*(1), 1-8.

Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 216-270). Cambridge, MA: MIT Press.

Rasch, T., & Schnotz, W. (2009). Interactive and non-interactive pictures in multimedia learning environments: Effects on learning outcomes and learning efficiency*Learning and Instruction* (Vol. 19, pp. 411-422). doi: 10.1016/j.learninstruc.2009.02.008

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models. Applications and data analysis methods*. Thousand Oaks, CA: SAGE Publications.

Reh, H. (2007). MaNET (Mannheimer Netzwerk Elaborations Technik) Version 1.6.4. Mannheim: MaResCom GmbH.

Renkl, A., & Gruber, H. (1995). Erfasung von Veränderung: Wie und wieso? *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 27*(2), 173-190.

Rips, L. J. (1994). *The psychology of proof: Deductive reasoning in human thinking*. Cambridge, MA: MIT Press.

Rost, D. H. (2005). *Interpretation und Bewertung pädagogisch-psychologischer Studien*. Weinheim: Beltz.

Rothmaler, P. (2000). *Introduction to model theory*. Amsterdam: Gordon & Breach Science Publishers.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. Bruce & W. F. Brewer (Eds.), *Theoretical issues in reading and comprehension* (pp. 33-58). Hillsdale, NJ: Lawrence Erlbaum.

Rumelhart, D. E., & Norman, D. A. (1978). Accretion, tuning and restructuring: Three model of learning. In R. L. Klatzky & J. W. Cotton (Eds.), *Semantic factors in cognition* (pp. 37-53). Hillsdale, NJ: Lawrence Erlbaum.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing. Explorations in the microstructure of cognition. Volume 2: Psychological and biological models* (pp. 7-57). Cambridge, MA: MIT Press.

Rüsen, J., Fröhlich, K., Horstkötter, H., & Schmidt, H. G. (1991). Untersuchungen zum Geschichtsbewußtsein von Abiturienten im Ruhrgebiet. In B. von Borries, H. Pandel & J. Rüsen (Eds.), *Geschichtsbewußtsein empirisch* (pp. 221-344). Pfaffenweiler: Centaurus-Verlagsgesellschaft.

Russel, W. A., & Jenkins, J. J. (1954). The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test: University of Minnesota.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson.

Scaife, M., & Rogers, Y. (1996). External cognition: how do graphical representations work? *International Journal of Human - Computer Studies, 45*(2), 185-213.

Scandura, J. M. (1988). Role of relativistic knowledge in intelligent tutoring. *Computers in Human Behavior, 4*(1), 53-64.

Scandura, J. M. (2007). Introduction to knowledge representation, construction methods, associated theories and implications for advanced tutoring/learning systems. *Technology, Instruction, Cognition and Learning, 5*(2), 91-97.

Schaeken, W., Vandierendonck, A., Schroyens, W., d'Ydewalle, G., & Klauer, K. C. (Eds.). (2006). *The mental models theory of reasoning. Refinement and extensions*. Mahwah, NJ: Lawrence Erlbaum.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32*(1), 102-119.

Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Student´s transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching, 28*(859-882).

Scheele, B., & Groeben, N. (1984). *Die Heidelberger Struktur-Lege-Technik (SLT). Eine Dialog-Konsens-Methode zur Erhebung subjektiver Theorien mittlerer Reichweite*. Weinheim: Beltz.

Schimmel, B. J. (1983). *A meta-analysis of feedback to learners in computerized and programmed instruction.* Paper presented at the AREA 1983, Montreal.

Schnotz, W. (2001). Kognitive Prozesse bei der sprach- und bildgestützten Konstruktion mentaler Modelle. In L. Sichelschmidt & H. Strohner (Eds.), *Sprache, Sinn und Situation* (pp. 43-57). Wiesbaden: Deutscher Universitätsverlag.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*(2), 141-156. doi: 10.1016/S0959-4752(02)00017-8

Schönwiese, C.-D. (2005). Klimawandel - Tatsache oder Fiktion? *Energiewirt, 104*, 26-29.

Schuler, H., & Prochaska, M. (2001). *Leistungsmotivationsinventar*. Göttingen: Hogrefe.

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

Schwarzer, R., & Jerusalem, M. (Eds.). (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin: Freie Universität Berlin.

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

Seel, N. M. (1995). Mental models, knowledge transfer, and teaching strategies. *Journal of Structural Learning and Intelligent Systems, 12*(3), 197-213.

Seel, N. M. (1999a). 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. (1999b). Educational semiotics: School learning reconsidered. *Journal of Structural Learning and Intelligent Systems, 14*(1), 11-28.

Seel, N. M. (2001). Epistemology, situated cognition, and mental models: 'Like a bridge over troubled water'. *Instructional Science, 29*(4-5), 403–427.

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

Seel, N. M. (2008). Empirical perspectives on memory and motivation. In J. M. Spector, M. D. Merrill, J. van Merrienboer & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 39-54). New York: Routledge.

Seel, N. M., Darabi, A. A., & Nelson, D. W. (2006). A dynamic mental model approach to examine schema development in performing a complex troubleshooting task: Retention of mental models. *Technology, Instruction, Cognition and Learning, 4*(3-4), 303-329.

Seel, N. M., & Dinter, F. R. (1995). Instruction and mental model progression: Learner-dependent effects of teaching strategies on knowledge acquisition and analogical transfer. *Educational Research and Evaluation, 1*(1), 4-35.

Seel, N. M., Ifenthaler, D., & Pirnay-Dummer, P. (2009). Mental models and problem solving: Technological solutions for measurement and assessment of the development of expertise. In P. Blumschein, W. Hung, D. H. Jonassen &

J. Strobel (Eds.), *Model-based approaches to learning: Using systems models and simulations to improve understanding and problem solving in complex domains* (pp. 17-40). Rotterdam: Sense Publishers.

Seel, N. M., & Schenk, K. (2003). Multimedia environments as cognitive tools for enhancing model-based learning and problem solving. An evaluation report. *Evaluation and Program Planning, 26*(2), 215-224.

Shavelson, R. J. (1972). Some aspects of the correspondence between content structure and cognitive structure in Physics education. *Journal of Educational Psychology, 63*(3), 225-234.

Shavelson, R. J. (1974). Methods for examining representations of a subject-matter structure in student memory. *Journal of Research in Science Teaching, 11*(3), 231-249.

Shavelson, R. J., & Stanton, G. C. (1975). Construct validation: Methodology and application to three measures of cognitive structure. *Journal of Educational Measurement, 12*(2), 67-85.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.

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* (pp. 23-42). New York: Springer.

Simons, P. R. J., & de Jong, F. P. C. M. (1992). Self-regulation and computer-aided instruction. *Applied Psychology: An International Review, 41*(4), 333-346.

Smith, W. G. (1894). Mediate association. *Mind, 3*(11), 289-304.

Smith, W. G. (1918). Methods for studying controlled word associations. *Psychobiology, 1*(6), 369-428.

Snow, R. E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Researcher, 18*(9), 8-14.

Snow, R. E. (1990). New approaches to cognitive and conative assessment in education. *International Journal of Educational Research, 14*(5), 455-473.

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (pp. 263-331). New York: ACE/Macmillan.

Sowa, J. F. (1984). *Conceptual structures: Information processing in mind and machine*. Reading, MA: Addison-Wesley.

Spada, H. (1983). Die Analyse von Veränderungen im Rahmen unterschiedlicher testtheoretischer Modelle. In W.-R. Minsel & R. Scheller (Eds.), *Brennpunkte der Klinischen Psychologie* (pp. 83-105). München: Kösel-Verlag.

Spector, J. M. (2006). A methodology for assessing learning in complex and ill-structured task domains. *Innovations in Education and Teaching International, 43*(2), 109-120.

Spector, J. M. (2010). Mental representations and their analysis: An epestimological perspective. In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 27-40). New York: Springer.

Spector, J. M., Dennen, V. P., & Koszalka, T. A. (2006). Causal maps, mental models and assessing acquisition of expertise. *Technology, Instruction, Cognition and Learning, 3*(2), 167-183.

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

Stachowiak, F. J. (1979). *Zur semantischen Struktur des subjektiven Lexikons*. München: Wilhelm Fink Verlag.

Sternberg, R. J. (1993). Giftedness as developing expertise. In K. A. Heller, F. J. Mönks, R. J. Sternberg & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (pp. 55-66). Oxford: Pergamon.

Sternberg, R. J., & Gardner, M. K. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General, 112*(1), 80-116. doi: 10.1037/0096-3445.112.1.80

Stoyanova, N., & Kommers, P. (2002). Concept mapping as a medium of shared cognition in computer- supported collaborative problem solving. *Journal of Interactive Learning Research, 13*(1/2), 111-133.

Stracke, I. (2004). *Einsatz computerbasierter Concept Maps zur Wissensdiagnose in der Chemie. Empiri-*

sche Untersuchungen am Beispiel des Chemischen Gleichgewichts. Münster: Waxmann.

Strasser, A. (2010). A functional view toward mental representations. In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 15-26). New York: Springer.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Taber, K. S. (1995). Development of student understanding: a case study of stability and lability in cognitive structure. *Research in Science & Technological Education, 13*(1), 89-99.

Taber, K. S. (2000). Multiple frameworks?: Evidence of manifold conceptions in individual cognitive structure. *International Journal of Science Education & Training, 22*(4), 399-417.

Tamir, P., & Jungwirth, E. (1972). Teaching objectives in biology: Priorities and expectations. *Science Education, 56*(1), 31-39.

Taricani, E. M., & Clariana, R. B. (2006). A technique for automatically scoring open-ended concept maps. *Educational Technology Research and Development, 54*(1), 65-82.

Tennyson, R. D., & Cocchiarella, M. J. (1986). An empirically based instructional design theory for teaching concepts. *Review of Educational Research, 56*(1), 40-71.

Tergan, S.-O. (2003). Managing knowledge with computer-based mapping tools. In D. Lassner & C. McNaught (Eds.), *Proceedings of the ED-media 2003 world conference on educational multimedia, hypermedia & telecommunication* (pp. 2514-2517). Honolulu, HI: University of Honolulu.

Thompson, T. L., & Mintzes, J. J. (2002). Cognitive structure and the affective domain: On knowing and feeling in biology. *Journal of Science Education, 24*(6), 645-660.

Tittmann, P. (2003). *Graphentheorie. Eine anwendungsorientierte Einführung*. München: Carl Hanser Verlag.

Tittmann, P. (2010). Graphs and networks. In D. Ifenthaler, P. Pirnay-Dummer & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 177-188). New York: Springer.

Trumper, R. (2006). Factors affecting junior high school students' interest in biology. *Science Education International, 17*(1), 31-48.

Turner, R. M. (1994). *Adaptive reasoning for real-world problems: A schema-based approach*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tutte, W. T. (2001). *Graph theory*. Cambridge, UK: Cambridge University Press.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

van der Meer, E., & Schmidt, B. (1992). Finale, kausale und temporale Inferenzen. Analyse ihres kognitiven Hintergrundes. *Zeitschrift für Psychologie, 200*, 303-320.

von Borries, B. (2001). Lehr- und Lernforschung im Fach Geschichte. In W. Gerhard (Ed.), *Lehren und Lernen im Kontext empirischer Forschung und Fachdidaktik* (pp. 399-438). Donau-Wörth: Auer.

Voss, J. F., Greece, T. R., Post, T. A., & Penner, B. C. (1983). Problem-solving skill in the social sciences. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press.

Vye, N. J., Goldman, S. R., Voss, J. F., Hmelo, C., & Williams, S. (1997). Complex mathematical problem solving by individuals and dyads. *Cognition and Instruction, 15*(4), 435-484.

Wagner, W., & Wagner, S. U. (1985). Presenting questions, processing responses, and providing feedback in CAI. *Journal of Instructional Development, 8*(4), 2-8.

Watts, M. (1988). From concept maps to curriculum signposts. *Physics Education, 23*, 74-79.

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 Revision*. Göttingen: Hogrefe.

Wells, F. L. (1911). Some properties of the free association time. *Psychological Review, 18*, 1-24.

Wild, K. P. (2000). *Lernstrategien im Studium. Strukturen und Bedingungen*. Münster: Waxmann.

Wilhelm, P., & Beishuizen, J. J. (2003). Content effects in self-directed inductive learning. *Learning and Instruction, 13*(4), 381-402. doi: 10.1016/S0959-4752(02)00013-0

Willett, J. B. (1988). Questions and answers in the measurement of change. *Review of Research in Education, 15*, 345-422.

Winter, H. (1975). Allgemeine Lehrziele im Mathematikunterricht. *Zentralblatt für Didaktik der Mathematik, 3*, 106-116.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. New York: Harcourt Brace & Company.

Wolfe, M. B. W., & Goldman, S. R. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction, 23*(4), 467-502.

Woods, C. (2007). Researching and developing interdisciplinary teaching: towards a conceptual framework for classroom communication. *Higher Education, 54*(6), 853-866. doi: 10.1007/s10734-006-9027-3

Young, M. J. (1993). Instructional design for situated learning. *Educational Technology Research and Development, 41*(1), 43-58.

Young, M. J. (1998). Quantifying the characteristics of knowledge structure representations: A lattice-theoretic framework. Los Angeles, CA: CRESST.

Zimmerman, B. J., & Schunk, D. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement. Theoretical perspectives* (pp. 1-37). Mahawah, NJ: Lawrence Erlbaum Associates.