Model based assessment of learning dependent change within a two semester class

Nadine Schlomske · Pablo Pirnay-Dummer

Abstract The following empirical study examines the acquisition of expertise. The model structures in the domain of empirical methodologies were examined in a time period of two semesters. At each measurement point (N = 5), the model structures of the reference groups were compared with those of the group of learners. The group of learners consisted of (N = 31) students who did not have any experience with research methods and their theoretical foundations. The results indicate that the external criteria enabled a precise assessment of learning dependent model change.

Keywords Acquisition of expertise · Learning dependent change · Learning and instruction · Mental models · Knowledge assessment

Knowledge and excellence are acquired by students as they learn. Assessing the acquisition of expertise requires valid and reliable instruments. Existing instruments already in use at schools and universities are often very time consuming and difficult to handle in procedure and analysis. The present study examines whether MITOCAR (Model Inspection Trace of Concepts and Relations) is an appropriate tool for predicting the acquisition of expertise. MITOCAR is based on the theory of mental models (c.f. Seel 1991) and is an automated instrument for the analysis of knowledge and expertise (Ifenthaler 2008; Pirnay-Dummer 2006, 2007; Pirnay-Dummer and Spector 2008; Pirnay-Dummer and Walter 2008). Moreover, the present study is interested in whether or not the models identified with the help of this instrument can be drawn on as external criteria in order to predict changes of this kind in learners' models. According to Johnson-Laird (1983), learners change the

N. Schlomske (✉)
Department of Educational Science, University of Jena, Am Planetarium 4, 07743 Jena, Germany
e-mail: nadine.schlomske@uni-jena.de

P. Pirnay-Dummer
Department of Educational Science, University of Freiburg, Rempartstraße 11, 79098 Freiburg, Germany
e-mail: pablo.pirnay-dummer@ezw.uni-freiburg.de

structures of their models as they progress from novices to experts. Within our theoretical framework we work with words as basic semantic references to the world. We do not exclude the possibility that there may be sub-symbolic levels of knowledge, but we do not need this level for our theorems at this time. Words, mainly nouns, are connected pairwise by a function (e.g., a causal relation) and a level of association (a strength). Although we are interested in both connections, we have made more coherent progress with the latter—especially for automated assessment. A model structure is a set of pairwise connections with a focus on either a specific subject domain, a problem solving or reasoning task, or both. It contains the strongest connections between higher symbols (words, nouns) as regards associations.

The group of learners in the present study consists of students who are at the beginning of their studies and do not have any prior experience in research methods and who can thus be seen as novices. Their model structures will be examined at five different points of measurement and will be compared with the model structures of two reference groups which have been empirically identified beforehand with the same instrument and have taken the same seminar.

The present paper is structured in two main parts: First, we discuss the theoretical background and the selection of MITOCAR as an appropriate instrument. Second, we explain how we conducted the study, the measurements, and the statistical analysis. Finally, we discuss the empirical results and embed them in the theoretical background.

Learning evokes models which change over time

Mental models

The theory of mental models (Johnson-Laird 1983; Seel 1991) posits a change in the learner's knowledge of the world and thus also a change in model structures after learning has taken place. In order to negotiate one's way in a changing environment, one needs the ability to permanently update one's knowledge (Pirnay-Dummer 2006). As a consequence, the internal states of knowledge need to be approximated to the external world (van der Meer 1996) because decisions are made on the basis of present knowledge. This implies that the internal states of knowledge be changed so that learning can take place. Individuals create mental models to explain the world on the basis of their knowledge (Ifenthaler 2006). Mental models contain only elements of knowledge which are relevant for explaining one's present situation (Stachowiak 1973). The statements represented in a person's knowledge must be plausible according to his or her knowledge of the world, regardless of whether they are true in reality or not (Seel 1991). According to Pirnay-Dummer (2006) mental models are only constructed if a person cannot explain a particular situation on the basis of his or her present knowledge of the world. In addition, he argues that there cannot be instruments that measure mental models reliably because of their instability. The models identified in the present study are regarded as novice model structures that change to become more advanced models during the course of learning.

Expertise

Novices are characterized as having no prior experience in a particular domain (Gruber 1994). Experts differ from novices not only in the amount of knowledge available to them,

but also in their knowledge structures (Pirnay-Dummer 2006). Gruber (1994) makes a distinction between four different activities in which high achievement can be recognized by experts: (1) manual activities, (2) mental and academic activities, (3) complex activities, and (4) artistic activities (Gruber 1994). The second and third parts are of special interest in the present study.

The learning and instruction dependent acquisition of expertise

Learning and instruction at schools or universities have the aim of bringing learners to a higher level of competence in a particular domain (Gruber and Mandl 1996). Learners change their models by means of constructing mental models. This implies that they have to be influenced by teaching in such a way that they approximate expert models (Seel 1991). Otherwise, the novices hold on to their "incorrect" models as long as they can still explain the world on the basis of their invalid knowledge. Only when their knowledge no longer suffices to explain the world will they change their invalid knowledge (Ifenthaler 2006). Finally, novice models change to become expert models through learning (Johnson-Laird 1989). Learning—and finally the change of learners' models—takes place on three levels: first, on a declarative level, second, on a procedural level, and third, on a semiotic level (Ifenthaler 2006). The amount of epistemic statements available to the cognitive system in a particular domain is permanently updated on the declarative level. Moreover, models constructed by experts are more elaborated than those made by novices. Heuristics are improved on the procedural level, while symbolic signs are updated on the semiotic level depending on how the knowledge can be represented (Ifenthaler 2006). Johnson-Laird (1983) calls this change of models and the processes involved in it "fleshing-out," while Seel (1991) speaks of it as "successive model change." An external criterion is needed in order to demonstrate a learning dependent change and, ultimately, to reveal the acquisition of expertise. In the present study, we draw on novice models in order to represent the present state of knowledge of the target group and advanced learners' models, which may be considered expert models, in order to represent the target state of knowledge.

Re-representation of knowledge

Bruner (1964) distinguishes between three different formats of knowledge representations (cf. Ifenthaler 2006). (1) the enactive format of knowledge representation, (2) the iconic format of knowledge representation, and (3) the symbolic format of knowledge representation. The first indicates activities that have a particular aim, such as "knowing what to do in order to send an email." The iconic format of knowledge representation indicates activities by means of pictures, for instance having a map in order to get from point A to point B. The last format of knowledge representation indicates symbols, such as signs that are important for re-representing one's knowledge.

According to Seel (1991), a symbolic characteristic is realized in internal representations (Seel 1991) because a symbolic system is necessary in order to represent one's knowledge. For assessment and observation, the internal structure has to be externalized again, which can be described as a re-representation. In the present study, we examine re-representations of knowledge. In the following section, we will discuss the examination of such external structures.

Research question

The question driving this study is: Can the acquisition of expertise be assessed with already proven instruments?

The central issues in the present study were, first, whether MITOCAR is able to identify the acquisition of expertise, and second, what kind of learning dependent model change can be detected by learners of empirical methodologies using MITOCAR.

Hypotheses

H_{1a}: The model structures of the learners correlate with the model structures of the reference groups (novices and advanced learners).

H_{0a}: The model structures of the learners do not correlate with the model structures of the reference groups (novices and advanced learners).

While a decrease is expected in the correlation coefficients between the novices, a learning dependent increase is expected between the advanced learners.

H_{1b}: There is a correlation between the model structures of the novices/advanced learners and the measurement points.

H_{0b}: There is no correlation between the model structures of the novices/advanced learners and the measurement points.

Model guided assessment

Instruments are required that reliably and validly assess the construction of 'knowledge' (Eckert 1998). According to the theory of mental models (Johnson-Laird 1983; Seel 1991), learners change the structures of their models as they progress from novice to expert. In order to examine this process, it is thus necessary to employ an instrument that identifies novices and experts and distinguishes them from one another. As already pointed out, internal representations can never be assessed in a direct way, but rather only their re-represented realization. This implies that they can be re-represented by a system. Thus, the instrument must be able to re-represent the model structures of novices and experts (Eckert 1998). Janetzko and Strube (2000) points out how important it is to consider the relational connections of model structures when modeling knowledge procedures (Janetzko and Strube 2000). This can be done when novices and experts relate concepts to each other graphically. This allows one to infer their cognitive structures (Mandl and Fischer 2000). Existing instruments that consider this aspect of analyzing knowledge are already being used in pedagogical diagnostics (Eckert 1998). Ifenthaler (2006) provides an overview of available instruments for assessing model structures.

In order to assess learning dependent changes in model structures, methodological requirements have to be aligned precisely to the theoretical assumptions (Ifenthaler 2006). Several measurements are desirable to reveal changes from novice models to advanced learning models. Therefore, a design is required that identifies learning dependent model changes between the individual time points on a quantitative level. In addition, it is desirable for the parameter value of the dependent variable to remain constant at each time point to enable a comparison of the values (Ifenthaler 2006). Finally, it is desirable to set external criteria that describe both the beginning state of knowledge and the target state of

knowledge. The target state of knowledge enables an especially precise assessment of how far away the learners are from the learning goals. The present study is particularly interested in assessing changes in models after learning has taken place.

Methods

Experiment

The design of the study is an *intra-subject design* which shows the learners' progress. The group of learners consisted of 31 students who were at the beginning of their studies and did not have any experience in empirical research methods and theoretical fundamentals and who could thus be seen as novices. During their first semester they were familiarized with the descriptive statistics as well as the basics of scientific hypotheses. This involved having them run little projects in which they investigated everyday problems. In the following semester they learned about inferential statistics by conducting little empirical studies with educational relevance and learning to choose the appropriate statistical analysis for their research questions. In addition to the seminar, they participated in tutorials (Table 1).

As external criteria, we compared the group of learners to two further groups which had been assessed before by the same instrument: a group of novices (N = 26) (equivalent to the beginning state of knowledge of the group of learners) and a group of advanced learners (N = 32) (equivalent to the target state of knowledge of the group of learners). Both reference groups had participated in the same seminar in the past but with different teachers. Their models had been assessed during another study (Pirnay-Dummer 2006). Within our study, their models were used only to compare with the models constructed by the learner groups.

The group of learners was measured at five different points within a period of 7 months: first, at the beginning of their first semester, second, in the middle of their first semester right before Christmas break and, third, at the end of their first semester. Finally, they were measured at the beginning of their second semester and in the middle of their second semester right before Pentecost break. At each measurement point, the model structures of the reference groups were compared with those of the group of learners. We expected a change in the structure of the learner group's models in the course of the learning process away from the models of the first reference group and toward those of the second reference group.

The experiment took place in the computer rooms of the Department of Education in Freiburg. The data was collected anonymously. After the subjects had logged in with a web browser, they entered their number, the group and block number, and their gender and age. Finally, they read the instructions and worked through the experiment on their own. This prevented investigator effects and thus guaranteed a high objectivity. The group and block

Table 1 Intra-subject design

	Winter semester			Summer semester	
2004	Novices (November)			Advanced learner (July)	
2006/07	lst MP	2nd MP	3rd MP	4th MP	5th MP
	November	December	January	April	May

numbers identified the particular experiment. The former number referred to the group of learners, whereas the latter referred to the stable sequences of the experiment.

The verification mode from MITOCAR was taken to test the subjects' models against the comparison models. Already existing models of the reference groups were drawn on as external criteria. The comparison models consisted of 30 pairs of concepts each. The paired concepts were rated by the subjects for their closeness, contrast, and for how sure the subjects were about their rating.

Instrument (MITOCAR)

MITOCAR is a software toolset developed and introduced by Pirnay-Dummer (2006, 2007). It is based strictly on mental model theory (cf. Seel 1991) and has proven to deliver valid, homogeneous, and reliable results. MITOCAR is an acronym for "Model Inspection Trace of Concepts and Relations." It measures properties of language re-representations of a realization by a group. The re-representation is called the group consensus model: MITOCAR also measures whether there is sufficient agreement within the group (homogeneity).

To produce the consensus model of the graph, all the subjects need to do is go through a two phase web-based assessment procedure which takes approximately 1.5 h for a whole group. Afterwards, MITOCAR generates automated reports which not only display the knowledge structure in a concept map-like format but also calculate and interpret several tests, e.g., multidimensional scaling and homogeneity (within a group), and provide additional descriptive measures and graphs which help the subjects to find answers within the knowledge structure (cf. Pirnay-Dummer 2007).

Identification, Review, Construction, Verification, and Confrontation are the modules which are presented separately to and used by the subjects. While Identification and Verification are mandatory for the functioning of MITOCAR, all of the other modules can be used to improve the quality of the knowledge assessment. All of the other steps are calculated automatically by the software and handled and stored on a database. The *identification* mode is the first phase of MITOCAR and is a simple collection of statements on a given subject domain. Between the first and the second phase, a *concept parser* filters nouns (with and without attached adjectives) and compiles a list of the most frequent concepts from the "mini-corpus." The second phase consists of the review, the construction, the verification, and the confrontation. In the *review* mode, every group member rates all expressions of the group for plausibility and for their relatedness to the subject domain. In the *construction* mode, the subjects categorize concepts into groups which can be processed into model information using Markov chains (Pirnay-Dummer 2006). *Verification* and *confrontation* are both modes for a pairwise comparison of concepts. Table 2 has been translated to be better understandable. The original instrument was applied in the German language (Table 2).

Paired concepts are rated by the subjects in the second phase of MITOCAR for their closeness and contrast. Additionally, the subjects rate how sure they are about their rating. The three basic measures and meaningful combinations of them can be used for the graphical reconstruction of the model later on. All items are rated on a 5 point Likert scale on screen by the subjects.

1. *Closeness*: The item of closeness describes how closely related two concepts are rated as being by the subjects.

Table 2 Ratings of the concepts

Please rate the following concept pair		
Statistics empirical research		
1. The concepts are closely related to each other		
Not at all	O O O O O	Completely
2. The concepts are mutually exclusive		
Not at all	O O O O O	Completely
3. In my rating I'm		
Unconfident	O O O O O	Confident

2. *Contrast*: For the item of contrast the subjects rate how different two concepts are or to what extent they exclude each other (e.g., fire and water).
3. *Combined*: This measure combines the items of closeness s and contrast k. It is calculated by $|(s - 1) - (k - 1)| + 1 = |s - k| + 1$. High contrast with low closeness and low contrast with high closeness both generate high combined values. The closer contrast and closeness become the lower the combined value will be. The scale remains the same as for closeness and contrast.
4. *Confidence*: The confidence rating ς measures how sure the subjects are of their ratings of contrast and closeness. To save space in titles and headers, all measures which are weighted by the confidence rating are indicated by a $(+)$ sign.

The MITOCAR software takes six pairwise related model representation measures into account:

1. *Closeness*: The model is constructed only on the basis of the closeness rating s.
2. *Contrast*: The model is constructed only on the basis of the contrast rating k.
3. *Closeness+*: The model is constructed on the basis of closeness and weighted by confidence: $k \cdot \varsigma$. If the subjects rate the relation of concepts with more confidence, they will also be more likely to become a part of the model.
4. *Contrast+*: The model is constructed on the basis of contrast and weighted by confidence: $s \cdot \varsigma$. If the subjects rate the relation of concepts with more confidence, they will also be more likely to become a part of the model.
5. *Combined*: The model is constructed on the basis of the combined measure $|(s - 1) - (k - 1)|$.
6. *Combined+*: The model is constructed on the basis of the combined measure and weighted by confidence: $|(s - 1) - (k - 1)| \cdot \varsigma$. If the subjects rate the relation of concepts with more confidence, they will also be more likely to become a part of the model.

Depending on the quality of the data (which is tested before re-representation), different measures may be used. E.g., if the combined item has too much deviance or is inhomogeneous within a group it will be excluded from re-representation. This is automatically tested and reported by the MITOCAR software. In this study the data quality sufficed for the combined measure (all measures had a good quality). *Verification* and *confrontation* modes differ only in the pairs of terms which are rated. In the verification mode subjects rate the terms which come from their own group (utilizing their own power of language), while in the confrontation mode they rate pairs from another group (typically from a group which they are being compared to). This information is used to build (re-represent) the knowledge structure in the form of a concept map.

In this study the *combined+* measure was taken to compare the learners' models to the reference models.

Results

Learning dependent change

For the combined measures, a decrease was revealed in the model structures of the novices, whereas an increase was observed in those of the advanced learners. In addition, a stabilization was observed as compared to the advanced learner model, whereas no such stabilization was observed as compared to the novice model. Moreover, the table above illustrates that the model structures may approximate each other (Tables 3 and 4).

All correlations with an asterisk (*) are statistically significant ($t_{MP2Nov} = 2.90$, $df_{MP2Nov} = 49$, $p_{MP2Nov} = 0.016 < 0.05$, $d_{MP2Nov} = 0.85$; $t_{MP5Nov} = 2.01$, $df_{MP5Nov} = 31$, $p_{MP5Nov} = 0.030 < 0.05$, $d_{MP5Nov} = 0.75$; $t_{MP3AdvL} = 2.59$, $df_{MP3AdvL} = 47$, $p_{MP3AdvL} = 0.025 < 0.05$, $d_{MP2Nov} = 0.77$; $t_{MP4AdvL} = 2.65$, $df_{MP4AdvL} = 33$, $p_{MP4AdvL} = 0.009 < 0.05$, $d_{MP4AdvL} = 0.95$). The correlation coefficients reveal a decrease in similarity in comparison to the novice models (except for at the second and the last measurements points) and an increase as compared to the advanced learners' models, except for at the last point.

Statistical analyses

In order to examine whether or not the models of the present study were homogenous, we tested the variances within each item for all measurement points. Only the first measurement point had homogeneity problems, which shows that the subjects may have had

Table 3 Combined measures during the measurements of change

MP	M		SD	
	Novices	Advanced learner	Novices	Advanced learner
1	0.78	0.56	0.09	0.17
2	0.70	0.56	0.08	0.08
3	0.70	0.59	0.07	0.07
4	0.68	0.62	0.08	0.07
5	0.67	0.63	0.08	0.07

Table 4 Correlations (between the learners and the reference groups)

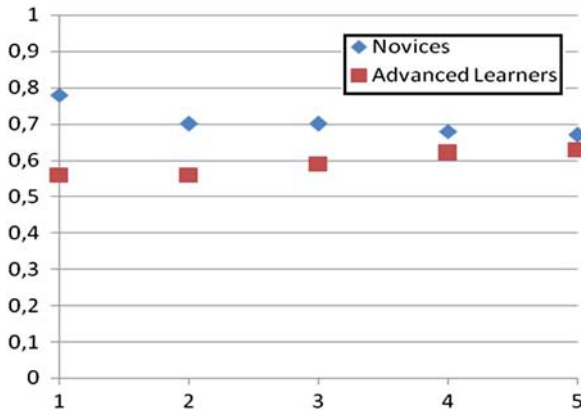| r_s		
MP	Novices	Advanced learners
1	0.12	0.25
2	0.39(*)	0.26
3	0.30	0.36(*)
4	0.25	0.43(*)
5	0.35(*)	0.27

Fig. 1 Measurement point as predictor

some problems with the rating environment as complete novices. However, all groups of novices in three previous studies had a sufficiently high homogeneity. The homogeneities were tested with an intra-item ANOVA (see Appendix).

The correlations (illustrated above) revealed that the novice models correlated significantly with the reference model at the second and at the fifth measurement point, whereas the advanced learner models correlated significantly with the reference models at the third and at the fourth measurement point (Fig. 1).

The figure above illustrates that the measurement points predict the means of the combined measures with a high correlation. A correlation of -0.99 was identified for the novices and a correlation of 1.00 for the advanced learners. Both correlations are statistically significant ($p < 0.05$).

Discussion

The results indicate that the external criteria may help to predict learning dependent model change beyond curricular variations. The measurements identified a correlation coefficient (Spearman) of -0.99 for the novices and a coefficient of 1.00 for the advanced learners. The learners in the present study showed a traceable change in their model structures as they progressed from novices to advanced learners. This indicates that the reference models provide appropriate indicators for predicting the development of expertise.

An examination of the learning dependent changes in the correlations revealed that the novices changed in accordance to the theoretical expectations, except for the change between the first and second measurements and at the last measurement point. The correlations of the advanced learners changed in line with the theoretical expectations, except for at the last measurement point. Ifenthaler (2006) has found a similar unexpected development of learning dependent change at the last measurement point. His empirical findings revealed the same phenomena, and his empirical data thus did not confirm his theoretical expectations. It may be assumed that knowledge which was plausible to the learner at a former point in time cannot easily be revised through learning and instruction. Obviously, the knowledge that is represented before learning has occurred and instructions have been given is still available to the learner even though it has been changed. This

finding underlines Seel's (1991) and Pirnay-Dummer's (2006) hypothesis that invalid models are not easy to revise. Regarding the standard deviations, a stabilization was revealed in the advanced learner models. Obviously, the instructions evoked learning, thus stabilizing the target knowledge about empirical research methods. Something which is still unclear is the unusually high correlation of the learner progression trajectory. This would even be surprising if the testing had been applied to the very same group, e.g., to account for retest reliability. Of course, we would have expected a correlation worth mentioning in this case. However, the dataset correlates with both models almost functionally and with precise linearity. At the moment, we can not explain such a high effect on a proper theoretical basis. And as much as researchers may like to find high effects, we will certainly need at least a replication to uncover the mechanisms which led to such results.

Conclusion

Even with a completely new group and a new instructor, the reference models of the previous group could be used to predict the learners' progress over time. In some cases this study shows that is possible to predict a group's learning behavior and progress. This further strengthens the position that—with a proper analysis—instruction can be planned on a systematic basis. The knowledge-based approach used in this study may provide a basis for continuing research planning and design. Undoubtedly, a good analysis will also require components other than knowledge assessment, but knowledge assessment can help extensively in planning curricula for groups (also see Pirnay-Dummer and Nußbickel 2008). The second conclusion is that automated knowledge assessment tools are in the process of becoming a methodologically sound means of tracking changes over time. Thus, the results from the study can be interpreted as another validation of the tools used in our study. Our future studies will therefore concentrate on two key aspects: How to utilize the predictability and validity for performance testing (particularly grading) and how to design and develop tools for automated self-assessment for individual learners. If it turns out that performance testing and self-assessment for individual learners can be assessed in an automated way and measured objectively, reliably, and validly with theory-based instruments, this research may lead to innovations in the educational field.

Appendix

Homogeneity in the reference model of the novices

The number of compared single ratings for the novices is n = 3518 (Tables 5 and 6).

- The group produced sufficiently homogeneous ratings. A model derived from the closeness rating (and combinations with it) can be interpreted as a model of the group.
- The group produced sufficiently homogeneous ratings. A model derived from the contrast rating (and combinations with it) can be interpreted as a model of the group.
- The group produced sufficiently homogeneous ratings. A model derived from the combined measure (and combinations with it) can be interpreted as a model of the group.

Table 5 Description of the measures

Measures	M	SD	Variance
Closeness	3.75	0.88	0.7762
Contrast	1.07	0.92	0.8390
Combined	3.32	1.24	1.5304
Confidence	3.83	0.85	0.7190

Table 6 ANOVA

Measures	df	F	ρ
Between subject			
Closeness	3045		
Contrast	3045		
Combined	3045		
Confidence	3045		
Within subject			
Closeness	434	0.2520	1.000
Contrast	434	0.4055	1.000
Combined	434	0.2632	1.000
Confidence	434	0.3640	1.000

- The group produced sufficiently homogeneous ratings. A model weighted by the confidence rating can be interpreted as a model of the group.

The derivation within the ratings of the novices is small enough to interpret their model using all ratings and combinations thereof. We can understand their model to be a consensus model.

Homogeneity in the reference model of the advanced learners

The number of compared single ratings for the advanced learners is n = 4208 (Tables 7 and 8).

- The group produced sufficiently homogeneous ratings. A model derived from the closeness rating (and combinations with it) can be interpreted as a model of the group.
- The group produced sufficiently homogeneous ratings. A model derived from the contrast rating (and combinations with it) can be interpreted as a model of the group.
- The group produced sufficiently homogeneous ratings. A model derived from the combined measure (and combinations with it) can be interpreted as a model of the group.

Table 7 Descriptions of the measures

Measures	M	SD	Variance
Closeness	3.71	0.94	0.8819
Contrast	2.01	0.96	0.9135
Combined	3.36	1.34	1.7893
Confidence	3.64	1.05	1.0971

Table 8 ANOVA

Measures	df	F	ρ
Between subject			
Closeness	4350		
Contrast	4350		
Combined	4350		
Confidence	4350		
Within subject			
Closeness	434	0.1646	1.000
Contrast	434	0.2465	1.000
Combined	434	0.2242	1.000
Confidence	434	0.4516	1.000

- The group produced sufficiently homogeneous ratings. A model weighted by the confidence rating can be interpreted as a model of the group.

The derivation within the ratings of the novices is small enough to interpret their model using all ratings.

References

Bruner, J. S. (1964). The course of cognitive growth. *American Psychologist, 19*, 1–16.

Eckert, A. (1998). *Kognition und Wissensdiagnose. Die Entwicklung und empirische Überprüfung des computerunterstützten wissensdiagnostischen Instrumentariums Netzwerk-Elaborierungs- Technik (NET)*. Lengerich: Pabst.

Gruber, H. (1994). *Expertise. Modelle und empirische Untersuchungen* (Vol. 34). Opladen: Westdeutscher Verlag.

Gruber, H., & Mandl, H. (1996). Das entstehen von expertise. In J. Hoffmann & W. Kintsch (Eds.), *Enzyklopädie der Psychologie, Themenbereich C, Theorie und Forschung, Serie II, Kognition* (Vol. 7, pp. 583–615). Göttingen: Hogrefe.

Ifenthaler, D. (2006). Diagnose lernabhängiger Veränderung mentaler Modelle. Veränderungsmessungen als Verfahren der empirischen Lehr-Lern-Forschung (University dissertation, Freiburg: FreiDok).

Ifenthaler, D. (2008). Relational, structural, and semantic analysis of graphical representations and concept maps. *Educational Technology Research and Development*. doi:10.1007/s11423-008-9087-4.

Janetzko, D., & Strube, G. (2000). Knowledge tracking–Eine neue Methode zur Diagnose von Wissensstrukturen. In H. Mandl & F. Fischer (Eds.), *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken* (pp. 199–217). Göttingen: Hogrefe Verlag.

Johnson-Laird, P. (1983). *Mental models toward a cognitive science of language. Inference and language*. Cambridge: Cambridge University Press.

Johnson-Laird, P. N. (1989). Mental models. In M. I. Posner (Ed.), *Foundations of cognitve science* (pp. 469–499). Cambridge, MA: MIT Press.

Mandl, H., & Fischer, F. (2000). *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken*. Göttingen: Hogrefe.

Pirnay-Dummer, P. N. (2006). Expertise und Modellbildung. MITOCAR (University dissertation, Freiburg: FreiDok).

Pirnay-Dummer, P. (2007). Model inspection trace of concepts and relations. A heuristic approach to language-oriented model assessment. *Paper presented at the AERA 2007, Division C, TICL SIG, Chicago, April 2007*.

Pirnay-Dummer, P., & Nußbickel, M. (2008). New ways to find out what is needed to know. Using the latest tools for knowledge elicitation in the processes of needs assessment. *Paper presented to the AERA 2008, Devision I, Education in the Professions*.

Pirnay-Dummer, P., & Spector, J. M. (2008). Language, association, and model re-representation. How features of language and human association can be utilized for automated knowledge assessment. *Paper presented to the AERA 2008, TICL SIG*.

Pirnay-Dummer, P., & Walter, S. (2008). Bridging the world's knowledge to individual knowledge. Using latent semantic analysis and web ontologies to complement classical and new knowledge assessment technologies. *Paper presented to the AERA 2008, TICL SIG*.

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

Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Wien: Springer.

van der Meer, E. (1996). Gesetzmäßigkeiten und Steuerungsmöglichkeiten des Wissenserwerbs. In Enzyklopädie der Psychologie (Vol. 2, pp. 208–248). Göttingen: Hogrefe.

Nadine Schlomske is PhD-student at the Department of Educational Science at the University of Jena, Germany. Her research interests are learning and the automated assessment of performance.

Pablo Pirnay-Dummer is assistant professor at the University of Freiburg, Germany. His main research interests are in learning, cognition and instruction with a focus on mental models, knowledge and knowledge assessment as well in technology enhanced learning environments and knowledge management.