Education Tech Research Dev (2013) 61:443–463
DOI 10.1007/s11423-013-9297-2

DEVELOPMENT ARTICLE

Validation study of a method for assessing complex ill-structured problem solving by using causal representations

Deniz Eseryel · Dirk Ifenthaler · Xun Ge

Published online: 26 April 2013

Abstract The important but little understood problem that motivated this study was the lack of research on valid assessment methods to determine progress in higher-order learning in situations involving complex and ill-structured problems. Without a valid assessment method, little progress can occur in instructional design research with regard to designing effective learning environments to facilitate acquisition of expertise in complex, ill-structured knowledge domains. In this paper, we first present a method based on causal representations for assessing progress of learning in complex, ill-structured problem solving and discuss its theoretical framework. Then, we present an experimental study investigating its validity against adapted protocol analysis. This study explored the impact of a massively multiplayer online educational game, which was designed to support an interdisciplinary STEM education on ninth-grade students' complex, ill-structured problem solving skill acquisition. We identify conceptual similarities and differences between the two methods, present our comparative study and its results, and then discuss implications for diagnostics and applications. We conclude by determining how the two approaches could be used in conjunction for further research on complex and ill-structured problem solving.

Keywords Complex problem solving · Causal representation · HIMATT

D. Eseryel (✉) · D. Ifenthaler · X. Ge
Department of Educational Psychology, University of Oklahoma, 820 Van Vleet Oval Rm 321, Norman, OK 73019-2041, USA
e-mail: eseryel@ou.edu

D. Ifenthaler
e-mail: dirk@ifenthaler.info

X. Ge
e-mail: xge@ou.edu

Introduction

Emerging interactive learning technologies, such as massively multi-player digital games, offer new opportunities for facilitating desired higher-order learning outcomes in complex knowledge domains like science, technology, engineering, and mathematics (STEM). However, little is known about which design strategies work best to support complex problem solving skill acquisition required for sustained impact on learning and instruction in complex knowledge domains. Based on their review of the literature, Eseryel et al. (2011) argue that lack of validated methods for assessing complex problem solving processes and outcomes is what impedes progress in this regard. Without a valid assessment method, little progress can occur in instructional design research with regard to designing effective learning environments including digital game-based learning environments. This was the motivation behind this study.

Assessing simple, well-structured problem solving is straight-forward since there is usually a single correct answer to such problems (Funke 2012; Jonassen 2011; Seel et al. 2009). However, assessing complex, ill-structured problems is more challenging. Two issues present obstacles to assessing complex and ill-structured problem solving. First, complex and ill-structured problems do not have standard correct answers. This is well-documented in prior research on expert problem solving in complex knowledge domains, such as instructional design (Eseryel 2006), architecture (Akin 1978), medicine (Spector and Koszalka 2004), and computer science (Guindon 1988). Secondly, it is hard to evaluate individual students on the basis of their actual performances because complex and ill-structured problem solving often requires teams of experts and specialists interacting over a period of months to develop an actual solution (Bierhals et al. 2007; Kearney et al. 2009). There are, of course, established approaches researchers use in studying problem solving, such as the think-aloud protocol analysis (Ericsson and Simon 1993), cognitive task analysis (Means 1993), or the structure formation technique (Scheele and Groeben 1984). However, these approaches require extensive resources in time and people; therefore, scalability to a large number of students within the limited resources of an educational context is problematic.

Hence, approaches to assessment of learning in complex, ill-structured domains mainly concentrate on assessing domain knowledge, largely through standardized tests, especially in school settings (Day et al. 2001). However, recent studies show that domain knowledge, while being a necessary component, is not sufficient to effectively solve complex, ill-structured problems (e.g., Newble et al. 2000). Further research on complex, ill-structured problem solving suggests that there are important cognitive activities or skills that are required in order to be able to solve complex, ill-structured problems (see Funke 1991). Of these, the two most important kinds are *structural knowledge* and *causal reasoning* that do not lend themselves for assessment by standardized tests (Jonassen 2004; Robertson 1990; Scandura 1977; Seel 1999). Structural knowledge and causal reasoning allow the problem solver to view the complex and dynamic problem space in its entirety, choose the most appropriate solution approach by mental simulation, and justify the decision.

One of the semantic networking tools that is appropriate for eliciting solvers' structural knowledge and causal reasoning is causal representations (Dörner and Wearing 1995; Jonassen 2000, 2004; Seel 1999). When presented by a complex and ill-structured problem, a causal representation of the problem-solver illustrates the conditions under which problem variables are interdependent and the effects of that interdependence. Hence, there is a growing literature on the use of causal representation as an alternative form of assessment for complex, ill-structured problem solving outcomes (e.g., Al-Diban and

Ifenthaler 2011; Eseryel 2006; Funke 1985; Ifenthaler 2010a, 2011a; Scheele and Groeben 1984; Seel 1999; Seel et al. 2009; Spector and Koszalka 2004). However, there have not been many systematic studies on the validity of these measures (Ifenthaler and Pirnay-Dummer 2013; Johnson et al. 2006).

The purpose of this paper is to present a study investigating validity of an automated assessment method (Highly integrated model assessment technology and tools (HIMATT); Pirnay-Dummer et al. 2010) for causal representations against detailed qualitative assessment based problem-solving rubrics (PSR) (Ge and Land 2003; Ge et al. 2010; Kauffman et al. 2008). The validity of an assessment instrument is the extent to which it measures what it was designed to measure (Belland et al. 2009; Mason and Bramble 1989; Seel et al. 2009). Associated sub-criteria of validity include (1) face validity, (2) content validity, (3) criterion validity, and (4) construct validity. Within the context of our study, we refer to construct validity, assuming that two independent instruments measure the identical construct, i.e., representation of ill-structured problem solving.

First we provide a comprehensive review of cognitive processes underlying problem solving in complex, ill-structured domains in order to provide a framework and rationale for the proposed assessment method. Following the details of the assessment method based on causal representations, we discuss the findings of the validation study. The paper concludes with an in-depth discussion of the findings for advancing research on assessing complex and ill-structured problem solving outcomes.

Cognitive processes underlying complex and ill-structured problem solving

The information-processing view suggests that problem solving requires a mental representation of the situation, which is referred as the problem space (Newell and Simon 1972). The process of solving a problem requires active manipulation of the problem space to achieve the desired outcome. The quality of problem solving is highly affected by how well a problem solver's mental representation reflects the actual situation, in other words, how well in the solver's mental representation the variables of the problem space and the interrelationships among these variables resemble or fail to resemble those of the *actual* situations.

In the case of complex problem solving, building and actively manipulating a problem space may be highly challenging due to the cognitive demands of understanding the *actual* complex problem-solving situation and the limited capacity of human short-term memory. Funke (1991) identified six features of complex problem-solving in further explaining the challenges: (1) complexity of the situation due to a large number of problem variables; (2) connectivity of the problem variables; (3) intransparency of the problem variables; (4) time-delayed effects; (5) dynamic developments; and (6) multiple goals.

Complex problem solving typically involves large number of variables with a high-degree of connectivity among these variables. This means that changes in one variable affect the status of many other related variables; therefore, it is very difficult to anticipate all possible consequences of any action. Secondly, in complex problem-solving situations, only some of the variables lend themselves to direct observation. Often, knowledge about the symptoms is available, from which one has to infer the underlying state. Other cases of intransparency arise if variables can be assessed in principle, but their huge number requires selection of a few relevant ones. Dealing with intransparency of the problem variables and the interrelationships among them is often difficult due to time-delayed effects; not every action shows immediate consequences in complex problem solving. In addition, complex problem solving situations often change decrementally and worsen, forcing a problem solver to act immediately, under considerable time pressure. Therefore, complex-problem-solving situations bear multiple goals, some of which could be contradictory requiring a reasonable trade-off. As a result, the problem space called forth by a complex problem could be likened to a complex and dynamic system with above six attributes described by Funke (1991).

In addition, a complex problem could be ill-structured. Ill-structured problems are ill-defined because one or more of the problem states (given or goal) are unknown or not known with any degree of confidence (Reitman 1965; Wood 1983). Ill-structured problems typically possess multiple representations and understandings of the problem. So, identifying an appropriate solution approach from among the competing options and justifying it make up the most important part of ill-structured problem-solving (Jonassen 1997; LeBlanc and Fogler 1995).

Hence, in order to successfully solve complex and ill-structured problems, the solver must be able to view and simulate in their "mind's eye" (Seel 2001, p. 407) the dynamic problem system in its entirety imagining the events that would take place if a particular action were to be performed. In this way, such mental simulation allows one to perform entire actions internally and to judge and interpret the consequences of actions and to draw appropriate conclusions. Accordingly, the solver makes a mental effort to understand a complex and ill-structured problem, and in doing so the solver constructs appropriate mental representations to model and comprehend the dynamic interrelationship among problem variables. The main purpose of these cognitive processes consists in the construction of causal explanations with the help of appropriate mental models so that the problem solver could understand how a change in one construct may affect another, thereby, leading to changes in the overall problem state (Ifenthaler and Seel 2011).

As such, cognitive processes underlying complex and ill-structured problem solving call for two important cognitive activities: *structural knowledge* and *causal reasoning*. *Structural knowledge* is the knowledge concerning how concepts within a domain are interrelated (Jonassen et al. 1993). In other words, structural knowledge refers to the integration and organization of concepts in an individual's memory. Scandura (1977) maintains that structural knowledge enables learners to form the complex interconnections required to make predictions and inferences necessary for solving problems. *Causal reasoning* is the other important cognitive component skill that is argued to be very important for solving complex, ill-structured problems. Causal reasoning, like structural knowledge, requires the attributions of one set of concepts to another. However, with causal reasoning, the attributions are causal. That is, one concept causes a change in state of another concept. Research on complex problem solving suggest that causal reasoning is perhaps the most important cognitive skill for complex, ill-structured problem solving because such problems usually involve changes in states brought on by causal factors (Chi and Glaser 1985; Dörner 1987; Dörner and Wearing 1995; Robertson 1990).

Together, structural knowledge and causal reasoning allow problem solver to view the complex and dynamic problem space in its entirety including all possible solution alternatives; choose the most appropriate solution approach by mental simulation; and justify the decision. Indeed, a number of research studies have shown that structural knowledge and causal reasoning are important for complex, ill-structured problem solving (Chi and Glaser 1985). For instance, Robertson (1990) used think-aloud protocols to assess structural knowledge and causal reasoning of students and found that the extent to which those think-aloud protocols contained relevant structural knowledge and causal reasoning was a strong predictor of how well learners would solve transfer problems in physics on a written exam. In fact, structural knowledge and causal reasoning were much stronger predictors than either aptitude (as measured by standardized test scores) or performance on a set of similar problems. Therefore, Robertson (1990) concluded that cognitive structures that causally connect the formula and important concepts in the knowledge base are important to understanding physics principles and it seems it is more important than aptitude.

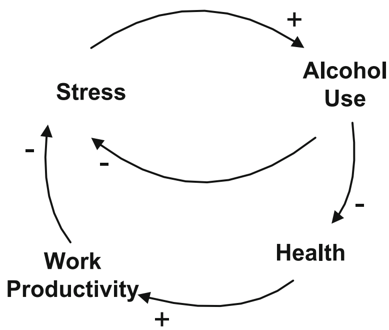
As seen in the study conducted by Robertson (1990), providing a solver a with problem scenario and asking the individual to think aloud while solving the problem has proved to be a valid and reliable method to assess structural knowledge and causal reasoning, and it is an established method for studying problem solving in research settings. However, collecting data using the think-aloud method and its subsequent analysis are very labor-intensive and takes considerable amount of time. Therefore, it is not a very cost-effective method in educational settings for assessment purposes. Additionally, the practicability for instructional settings (e.g., school, higher education) needs to be questioned, as theses assessment methods do not produce instant feedback to the problem solver (Ifenthaler 2011b). Hence, there is a need for other methods that can elicit and track changes in solvers' problem space (including their structural knowledge and causal reasoning), which can serve as a basis for assessing solvers' progress in complex and ill-structured problem solving.

Using causal representations to assess complex ill-structured problem solving

There is a growing literature on the use of causal representations as an alternative form of assessment for complex, ill-structured problem solving outcomes (e.g., Al-Diban 2008; Eseryel 2006; Funke 1985; Ifenthaler, 2010a, 2011a, 2012; Scheele and Groeben 1984; Seel 1999; Seel et al. 2009; Spector and Koszalka 2004). A causal representation is a semantic networking and knowledge elicitation approach that illustrates the interdependence of problem variables, specifically the conditions under which problem variables are interdependent and the causal effect of that interdependence. For instance, Fig. 1 describes the causal relationships in a family where a husband who is stressed at work drinks more alcohol, which in turn, causes more stress. Alcohol also causes a reduction in his health, and as his health crumbles, his work productivity decreases, which causes more stress at work, and so on.

In order to build causal representation of a complex and ill-structured problem, one must determine the problem variables affecting the problem state and the causal interrelationships among problem variables, and articulate the various causes that affect a problem state, the solution approaches that are available, and the trade-offs that result from each solution approach. As such, affordances of a causal representation make it an ideal candidate to elicit test persons' structural knowledge and causal reasoning presented by

Fig. 1 Sample causal influence diagram for stress



new, authentic, unencountered scenario-based questions that require them to make predictions about what will happen or draw an inference about what did happen (Dörner et al. 1983; Jonassen 2000, 2004; Jonassen and Cho 2008; Jonassen and Wang 1993). Therefore, causal representations are uniquely suited to elicit and continuously track a test person's structural knowledge and causal reasoning that make up their problem space during complex, ill-structured problem solving.

In the case of complex, ill-structured problem solving, assessment is challenging because domain experts are likely to exhibit a wide range of performances and many relevant activities are open-ended. Therefore, it is not possible to use a *correct solution* as a basis for the assessment (Berliner 2002; Feldon 2007). Since agreed-upon performance-measures do not exist, an important question is, despite the variances in the solution approaches of domain experts, whether it is possible to use an experts' problem space (including structural knowledge and causal reasoning) as a comparison point for assessment of complex, ill-structured problem-solving outcomes.

The answer to this question was positive in recent studies investigating complex and ill-structured problem solving in ecology (Spector et al. 2001), engineering, medicine, and biology (Spector and Koszalka 2004), instructional design (Eseryel 2006; Perez et al. 1995; Spector 1998; York and Ertmer 2011), and teaching (Berliner 2002). For instance, in a study with instructional design experts (Eseryel 2006), despite the fact that each expert provided a different solution approach to the given problem, their causal representations showed that expert instructional designers understand and interpret the given complex, ill-structured problem similarly; identify similar factors that are important for the solution of the problem; and foresee similar relationships among these factors. In other words, there were recognizable patterns among the problem spaces of experts that could be used as a comparison point for the purposes of assessment. On the other hand, the causal representations of novice instructional designers, who possessed the theoretical knowledge but not real-life experience, showed recognizably different patterns from those of the experts. Therefore, it was concluded that, in complex and ill-structured knowledge domains, expert problem space, elicited via a causal representation, could be reliably used as a comparison point for continuously assessing novice problem space as they move towards expertise (Eseryel 2006). Figure 2 depicts this assessment framework.

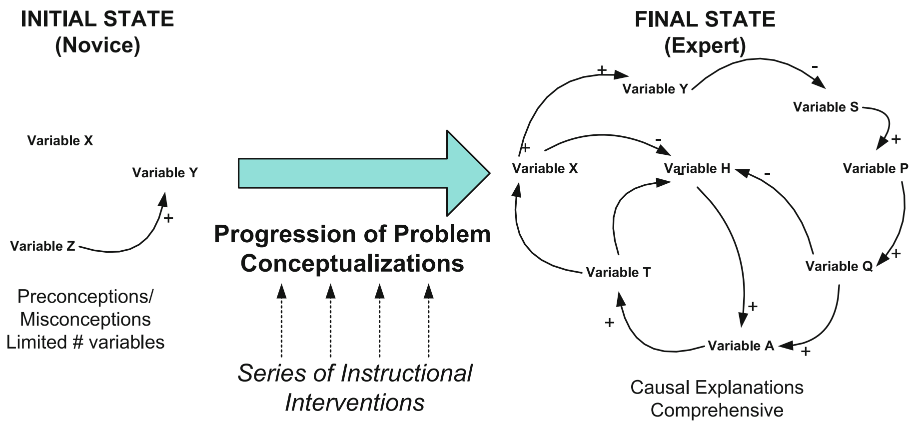


Fig. 2 An illustration of the assessment method based on causal representations

Based on previous research, this assessment framework assumes that problem space of novices initially exhibit preconceptions or misconceptions (Ifenthaler et al. 2011; Snow 1990). As novice's level of expertise increases, these preconceptions are gradually replaced with more comprehensive, causal explanations (Berliner 2002; Seel et al. 2000; Snow 1990). Thus, in the context of complex, ill-structured problem-solving, it is possible to describe effective learning process as one that facilitates transition of problem spaces of learners from the state of preconceptions or misconceptions to the state of comprehensive, causal explanations (Ifenthaler and Seel 2005, 2011) (see Fig. 2).

In order for this framework to be used as a robust assessment method, it is crucial to develop a similarity metric to help compare how problem spaces of novices change over time through a sequence of instructional interventions to resemble (or differ from) problem conceptualizations of advanced learners or experts so that it will be possible to measure the effectiveness of any particular instructional strategy (Spector et al. 2001). Recently, Pirnay-Dummer et al. (2010) devised a computer-based approach, called HIMATT, for comparing the similarity of sets of causal representations.

Research question and hypothesis

The purpose of this investigation is to extend the HIMATT analysis approach by describing construct-related validity of *HIMATT measures* with *problem-solving rubric* scores as an outside criterion. Over the past years, several possible solutions to the assessment and analysis of complex ill-structured problem solving skills and underlying mental models have been discussed (e.g., Clariana 2010; Clariana and Wallace 2007; Johnson et al. 2009; Spector 2006). However, most of these studies only focused on conceptual differences of the available methodologies and do not use adequate empirical data (e.g., Al-Diban and Ifenthaler 2011; Clariana and Wallace 2007; Eseryel et al. 2011; Johnson et al. 2009; Jonassen 2009; Jonassen and Cho 2008). Accordingly, the central reseach objective in this study was to identify strength and limitations of the two methodologies, i.e., HIMATT measures and the PSR, for assessing and analyzing progress of learning in complex and ill-structured problem solving using data from a longitudinal experimental study. We expected that both methodologies would generate similar results when applied on the same dataset.

Method

Following a design-based research framework, a longitudinal experimental study was carried out over two semesters to investigate the effects of an educational massively multiplayer online game (MMOG), called *McLarin's Adventures,* on facilitating the development of students' complex, ill-structured problem solving skills and to contribute to the literature with effective educational game design and implementation practices (Eseryel et al. 2011). Within this setting, we were able to conduct the validation study for comparing the two methodologies.

Participants

A rural high school in the Midwest of the United States was used as a testbed for this study. Three hundred and forty-nine 9th-grade students were randomly assigned to one of the nineteen classes. Out of these nineteen classes, ten were randomly assigned to the treatment condition (game group) and nine were randomly assigned to the control condition (no game group). We received both parental assent forms and student consent forms from 251 students. The data reported here are from 159 students out of those 251 students from whom we have complete responses. Of these 159 students, 85 were in the treatment group and 74 were in the control group. There were 44 % males and 56 % females.

Materials

The materials for this validation study included the McLarin's Adventures MMOG and the pretest and posttest data collection instruments (Eseryel et al. 2011).

McLarin's adventures

McLarin's Adventures is a MMOG designed for the 8th and the 9th grade students. When the students first enter the game-based learning environment, they are presented with the news reporting the eccentric trillionaire Jonathan McLarin's dream of interplanetary and interstellar travel. His company, McLarin International, finally produced a vehicle capable of traveling one light year in a single day. In this news video, McLarin announces the plans to send a team of experts to explore and survey Earth-like planets outside of our solar system.

Consequently, in *McLarin's Adventures* MMOG, the students are invited to play the role of researchers set in a survivor story where they would explore an uninhabited, uncharted island. Their earlier exploration indicated that the planet had similar characteristics to an island on earth. The students' task was to work with the teams to develop a living habitat and then report back to McLarin International. This task would test learners' complex problem solving and higher-order reasoning skills at finding necessary resources. Each complex task scenario requires students to apply their knowledge in mathematics, bioscience, geography, geology, social studies, and literacy to solve the problems, which are represented in two processes: problem representation (PRP) and generating solutions (GSO). The students were provided with a task completion protocol, which was intended to facilitate them to complete the problem-solving task.

Problem representation is a stage when solvers are required to represent the problem situation in the problem space (Newell and Simon 1972), and there could be multiple representations and understandings of an ill-structured problem (Jonassen 1997). During problem representation, the protocol prompted the students to analyze the situations and think about the essential factors for humans to survive, for example, ''what factors are important for your team to address in order to ensure the survival of humans in this new planet?'' The students were required to write each of the factors on a sticky note paper. Then the students were prompted to explain the relationships among the factors they had identified by arranging the notes (factors) and drawing the relationships. In addtion, the students were also asked to elaborate the relationship of the factors involved.

After the problem-representation activitiy, the students were asked to generate viable solutions from among more than one competing options and provide justifications about their selection (Jonassen 1997; LeBlanc and Fogler 1995) based on the analysis of problem situation. During the solution generation stage, students were prompted to make a recommendation to Mr. McLarin about their plans and the specific tasks their team would carry out to ensure human survival on this planet.

Within the game, the problem-solving task directed students to engage in activities such as locating water resources, locating food sources for colonization, determining the quality of water supplies and purification, settlement planning and building of shelters, creating an inventory of supplies and requirements for additional supplies, building a sanitation system, and so on. The goal of the game is to successfully complete the complex problem scenario embedded in the MMOG. Directions and hints for completing the complex problem scenario are embedded in the game.

Pre- and posttest data collection instruments and procedure

During the pre- and posttest, students in both experimental (game group (GG)) and control groups (control group (CG)) were provided with an account for the *McLarin's Adventures* game and an ultra-mobile PC (UMPC). After the students logged into the game-based learning environment they designed their avatars for the game, and watched the opening news video which announced the competition by McLarin International and invited potential applicants to complete the evaluation form, which would be used to select viable applicants. In order to collect the participant's understanding of the phenomenon in question, identical pre- and posttests were administered as follows.

During the pretest, all students were asked to respond to a complex and ill-structured problem scenario. The scenario represented a near-transfer problem solving task when compared with the problem scenario in *McLarin's Adventures* and it required students to play the role of lead scientist for the team of experts exploring Earth-like planets outside of our solar system (Refer to Eseryel et al. 2011) for further details of the study. Students were asked to construct an *annotated causal representation* of the problem constituents by using sticky notes to indicate each constituent and using one-way or two-way arrows between the constituents that have cause-and-effect relationship. They were then asked to provide a *written elaboration* explaining each problem constituent and each relationship. These assessments were later analyzed with HIMATT.

Following pretest data collection, the students in the treatment condition (GG) played *McLarin's Adventures* for 2 days per week (50 min per day) during class hour, while the students in the control condition had regular interdisciplinary class. After 16 weeks, posttest data was collected. The same problem solving scenario was used in the posttest. Accordingly, students constructed an *annotated causal representation* of the problem and provided a *written elaboration* explaining each problem constituent and each relationship.

Assessment and analysis methodologies

Basic questions of a reliable and valid assessment and analysis of complex ill-structured problem solving skills and underlying mental models are not yet solved completey (e.g., Ifenthaler 2008; Spector 2010). Accordingly, this validation study contributes to these unsolved questions by comparing two analysis approaches using an identical data set. In this section, we present the basic framework of both methods: (1) PSR, (2) HIMATT.

Method I: problem-solving rubrics

Adapted from the protocol analysis method (Ericsson and Simon 1993) PSR is an established research method for studying complex and ill-structured problem solving (Ge and Land 2003; Ge et al. 2010; Kauffman et al. 2008). In this study, the PSR were developed by the research team in collaboration with the group of experienced K-12 teachers, who designed the problem scenarios for the *McLarin's Adventures MMOG.* First, we asked these teachers to work through the same problem scenario that the students would go through and provide solutions to the problem-solving task. Their solution to the complex problem scenario was captured by the researchers. Second, the research team identified a list of key features important for solving the problem-solving task based on the information elicited from the teachers. These key features were categorized and organized into problem representation, justification, and solutions, as well as evaluation according to complex, ill-structured problem-solving literature (see Jonassen 1997, 2000). The identified problem-solving features served as critical areas for developing descriptors and criteria for the rubrics. Third, the scoring rubrics were tested with a smaller sample of students' responses, which were collected during the pilot study and excluded from this main study. This served as important feedback to enable the revision and improvement of the PSR. The final version of PSR, which assessed two ill-structured problem solving areas (i.e., *PRP* and *generate solutions*), were consolidated after the research team went through several iterative revision processes—discussion, meaning negotiation, and consensus reaching.

Problem representation (PRP) measured (a) the number of factors a student could list for survival on the planet, which were compared with the category list suggested by the subject matter experts (e.g., habitability, terrain, vegetation, and animal life), and (b) the degree a student described and elaborated the identified items related to survival. Based on the number of factors identified and the extent those factors were described and elaborated, a score was assigned on the overall depth of elaboration regarding the relationships among the factors influencing the problem state.

The rubrics for measuring *GSO* included four areas: (a) Has a recommendation been made by listing measures or actions? (b) Are solutions aligned with the problem representation completed in the previous step? (c) Has a justification been made about the recommendation? and (d) How confident does the individual feel about his or her recommendations?

Three raters coded students' problem-solving reports based on the PSR. As described earlier in detail, prior to the coding, they reached a conceptual consensus on how to interpret the scoring rubrics through discussion and examples. The scoring rubric was further refined while coding of pilot-test data that was not included in the reported study. Then, the coding of the main data started. Any discrepancies of assigned values were further discussed among the raters until 100 % agreement was reached and the adjudicated scores were used. Consequently, a high consensus was reached. We did not use Cohen's kappa for analyzing interrater agreement because it would not be applicable in this case because during the data coding any discrepancies of assigned values by the raters were further discussed until an agreement was reached; adjudicated scores were reported.

Method II: HIMATT

Highly integrated model assessment technology and tools (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 and Ifenthaler 2010; Pirnay-Dummer et al. 2010). Methodologically, the tools integrated into HIMATT use qualitative and quantitative research methods and provide bridges between them. First of all, text (e.g., student essay) can be analyzed very quickly without loosening the associative strength of natural language. Furthermore, graphical representations (e.g., causal representations) can be annotated by experts and compared to other solutions.

The automated analysis function produces measures which range from surface-oriented structural comparisons to integrated semantic similarity measures. There are four *structural* (surface, graphical, structural, and gamma matching) and three *semantic* (concept, propositional, and balanced propositional matching) measures available (Fig. 3).

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 for indepth qualitative analysis (Ifenthaler 2010a).

In order to analyze the learners' understanding of the problem scenario, the seven measures implemented in HIMATT can be applied. Accordingly, each of the learners' response to the phenomenon in question can be compared automatically against a reference solution (e.g., an expert solution, conceptual model, chapter from a text book, learning materials, etc.). In this study, the reference solution was provided by the same group of K-12 teachers that designed the problem scenarios in the game and helped developed the PSR. Table 1 describes the seven measures of HIMATT, which include four structural measures and three semantic measures (Ifenthaler 2010a, b; Pirnay-Dummer and Ifenthaler 2010; Pirnay-Dummer et al. 2010).

HIMATT uses specific automated comparison algorithms to calculate similarities between a given pair of frequencies. The similarity index s 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. In this case, the similarity introduced by Tversky (1977) applies for the given sets. This index also results in a measure of $0 \leq s \leq 1$, where $s = 0$ is complete exclusion and $s = 1$ is identity.

Reliability scores exist for the single measures integrated into HIMATT (see Pirnay-Dummer et al. 2010). They range from $r = .79$ to $r = .94$ and are tested for the semantic and structural measures separately and across different knowledge domains. Validity scores are also reported separately for the structural and semantic measures. Convergent

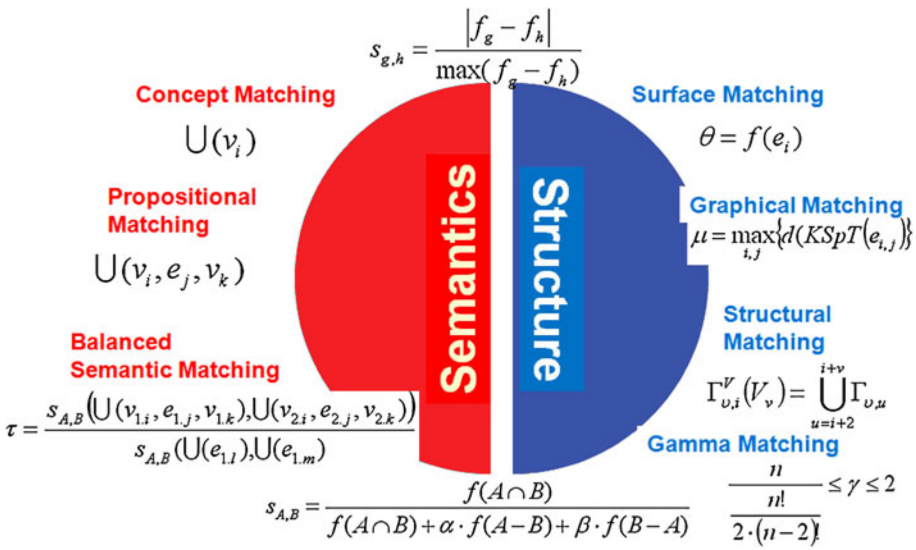


Fig. 3 HIMATT measures

Table 1 Description of the seven HIMATT measures

Measure [abbreviation] and type	Short description
Surface matching [SFM] *Structural indicator*	The surface matching (Ifenthaler 2010a) 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 2010a) 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 and 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 and 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 and 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 2010a) value compares only fully identical propositions between two graphs. It is a good measure for quantifying semantic similarity between two graphs.
Balanced semantic matching [BSM] *Semantic indicator*	The balanced semantic matching (Pirnay-Dummer and Ifenthaler 2010) is the quotient of propositional matching and concept matching. In specific cases (e.g., when focusing on complex causal relationships), balanced propositional matching could be preferred over propositional matching.

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).

HIMATT and its tools have been successfully applied within classical experimental settings (Ifenthaler 2011a, 2012; Ifenthaler and Lehmann 2012; Spector 2010) and fields of applications within the domains of instructional planning, medical diagnosis, and geology (Kim 2008; Lachner and Pirnay-Dummer 2010; Lee 2009; McKeown 2009). Further, empirical studies showed that HIMATT measures (e.g., GAM, CCM) helped to distinguish inexperienced from highly experienced problem solvers in all domains examined so far (Lee 2009; McKeown 2009). Finally, a usability test showed that HIMATT is widely accepted among the users and the usage is easy to learn (Pirnay-Dummer et al. 2010).

Results

In order to test our hypothesis, we analyzed the pre- and posttest data of the causal representations from the longitudinal experiment with both methods (1) PSR and (2) HIMATT.

Normal Q–Q plots indicated that the dependent variables used in the comparison study are meeting the requirements for the analysis. All effects were assessed at the .05 level. As effect size measures, we used partial η^2 (small effect: $\eta^2 < .06$, medium effect $.06 \leq \eta^2 \leq .13$, strong effect $\eta^2 > .13$). First, the results of an in-depth investigation of complex problem solving performance are presented. Next, the comparison of both methods is reported.

Results by PSR (method I) and HIMATT (method II)

We computed a repeated-measure MANOVA with the *problem representation measures* (PRP, GSO, SFM, GRM, STM, GAM, CCM, PPM, BSM) at two measurement points as a within-subjects factor, and *experimental groups* (GG and CG) as a between-subjects factor. MANOVA revealed a significant main effect of experimental group on problem representation measures, Wilks' Lamda = .841, $F(9, 146) = 3.060$, $p = .002$, $\eta^2 = .159$, for time on problem representation measures, Wilks' Lamda = .656, $F(9, 146) = 8.493$, $p < .001$, $\eta^2 = .344$, and for time x group, Wilks' Lamda = .879, $F(9, 146) = 2.238$, $p = .023$, $\eta^2 = .121$. This suggests that the problem representation measures across the measuremt points and experimental groups have at least one mean vector pairing which produced a significant difference. Table 2 shows the descriptive results (pre- and post-test) for the problem representation measures for the experimental groups as well as the differences between the two measurement points. In order to determine the significance of these differences a series of univariate pairwise comparisons were conducted.

First, results indicate significant differences between the measurement points for the problem representation measures as follows: PRP, $F(1, 154) = 41.34$, $p < .001$, $\eta^2 = .212$; GSP, $F(1, 154) = 23.79$, $p < .001$, $\eta^2 = .134$; SFM, $F(1, 154) = 11.33$, $p = .001$, $\eta^2 = .069$; GRM, $F(1, 154) = 7.59$, $p = .007$, $\eta^2 = .047$; STM, $F(1, 154) = 4.97$, $p = .027$, $\eta^2 = .031$; CCM, $F(1, 154) = 4.39$, $p = .038$, $\eta^2 = .028$.

Second, results indicate significant differences between the experimental groups for the problem representation as follows: GRM, $F(1, 154) = 8.72$, $p = .004$, $\eta^2 = .054$; STM,

Table 2 Means (standard deviations in parenthesis) of PSR and HIMATT measures ($n = 159$)

	Pre test		Post test		Difference	
	Treatment group ($n = 84$)	Control group ($n = 72$)	Treatment group ($n = 85$)	Control group ($n = 72$)	Treatment group ($n = 84$)	Control group ($n = 72$)
PRP	.71 (.77)	.85 (94)	.35 (.55)	.28 (.54)	−.37 (.22)	−.57 (.41)
GSO	1.37 (1.04)	1.21 (.84)	.98 (.58)	.85 (.64)	−.39 (.46)	−.36 (.20)
SFM	.36 (.15)	.33 (15)	.27 (.14)	.32 (.19)	−.09 (.19)	−.01 (.21)
GRM	.50 (.18)	.39 (.18)	.42 (.16)	.39 (.17)	−.08 (.20)	−.01 (.19)
STM	.65 (.14)	.56 (.19)	.59 (.14)	.56 (.16)	−.06 (.19)	−.01 (.19)
GAM	.70 (.23)	.58 (.22)	.66 (.23)	.60 (.26)	−.04 (.34)	.02 (.31)
CCM	.41 (.13)	.40 (.11)	.38 (.11)	.38 (.14)	−.03 (.15)	−.01 (.15)
PPM	.04 (.06)	.01 (.11)	.03 (.05)	.02 (.03)	−.02 (.07)	.01 (.04)
BSM	.09 (.13)	.02 (.05)	.06 (.11)	.05 (.09)	−.031 (.162)	.029 (.098)

PSR measures: *PRP* problem representation, *GSO* generating solutions; HIMATT similarity measures: *0* no similarity, *1* total similarity; SFM, GRM, STM, and GAM are structural measures; CCM, PPM, and BSM are semantic measures

$F(1, 154) = 6.10$, $p = .015$, $\eta^2 = .038$; GAM, $F(1, 154) = 9.89$, $p = .002$, $\eta^2 = .060$; PPM, $F(1, 154) = 10.88$, $p = .001$, $\eta^2 = .066$; BSM, $F(1, 154) = 11.78$, $p = .001$, $\eta^2 = .071$.

Third, we found significant interaction effects of time and group on the problem representation measures as follows: SFM, $F(1, 154) = 5.12$, $p = .025$, $\eta^2 = .032$; GRM, $F(1, 154) = 5.14$, $p = .025$, $\eta^2 = .032$; PPM, $F(1, 154) = 10.78$, $p = .001$, $\eta^2 = .065$; BSM, $F(1, 154) = 7.89$, $p = .006$, $\eta^2 = .049$.

To sum up, there were significant changes in participants' problem representations among the variables affecting the problem state (i.e., their causal reasoning) both structurally and semantically from pretest to posttest.

However, the changes were opposite to the desired direction—away from resembling the reference model. Descriptive statistics (Table 2) showed that most of the problem representation measures of the treatment and control group decreased over time.

Comparison of the two analysis methods

Using the same set of data, we were able to conduct an in-depth investigation of complex problem solving performance using both analysis methods—(1) PSR, and (2) HIMATT. Overall, the results of PSR and HIMATT indicated identical results: The students' problem space changed significantly from pretest to posttest as captured by *PRP* and GSO variables analyzed by the PSR method and the structural and semantic variables analyzed by the HIMATT method. In addition, both methods captured that the direction of the change was not towards resembling the subject matter expert's problem space; rather it changed in the opposite direction differing significantly from subject matter expert's problem space. Hence, the findings of both methods suggested that while the complexity of students' representations in the treatment group increased over time, those changes did not involve the desired conceptual changes. Rather, those changes implied newly-formed misconceptions, which could be attributed to cognitive overload during game play due to the interrelationships among the large number of variables affecting complex, ill-structured problems in the game.

Although both methodologies worked quite well and produced a lot of indicators, there were several strengths and limitations to report. First, the indicators of both methodologies measured on different conceptual levels. However, we were able to identify indicators for both methods, which reported on similar constructs. A correlational analysis of these measures showed the following results for the post-test measures. First, the measure PRP correlated significantly with the HIMATT measure GRM ($r = .176$, $p = .027$). Accordingly, both methods were able to capture the overall complexity of the problem representation with these measures.

Second, the measure GSO correlated significantly with the HIMATT measures STM ($r = .236$, $p = .003$), GAM ($r = .180$, $p = .024$), PPM ($r = .221$, $p = .005$), and BSM ($r = .263$, $p = .001$). Hence, the correlations among these measures in both methods speak to their validity and utility in capturing the students' deeper, conceptual understanding of the problem domain.

Discussion

In the light of recent research, educational researchers argued that current instructional design theories and models are inadequate in their ability to address complex, ill-structured problem-solving skills required for effective learning in complex domains such as STEM (e.g., Achtenhagen 2000; Dijkstra and van Merriënboer 1997; Gordon and Zemke 2000; Jonassen 1997; Spector et al. 2001). Several different instructional approaches have been proposed, however, no conclusive evidence exists regarding their effectiveness (see Spector and Anderson 2000). An important barrier that impedes progress in this regard is the lack of established assessment methodologies to determine progress of learning in complex knowledge domains (e.g., Baker and Schacter 1996), which require cognitive complexity (e.g., Andrews and Halford 2002), and complex ill-structured problem-solving skills (e.g., Funke 1991).

Accordingly, questions of a reliable and valid assessment and analysis of complex ill-structured problem solving are not yet completely solved. Especially the development of fast and economic methodologies still remains problematic (e.g., Clariana 2010; Eseryel et al. 2011; Spector 2010). In this paper, we first presented an assessment framework based on causal representations that shows a promise towards this goal. This assessment framework posits that in complex and ill-structured knowledge domains, expert problem space, elicited via a causal representation, could be reliably used as a comparison point for continuously assessing novice problem space as they move towards expertise (Eseryel 2006). In order for this framework to be used as a robust assessment method, it is crucial to develop a valid similarity metric to help compare how problem spaces of novices change over time through a sequence of instructional interventions to resemble (or differ from) problem conceptualizations of experts so that it will be possible to measure the effectiveness of any particular instructional strategy (Spector et al. 2001). The various measures in the HIMATT analysis method (Ifenthaler 2010a) provides the much-needed similarity metric for this assessment framework to serve as an assessment method that can be automated, and thus, highly scalable.

Comparison of the two methods

The study presented in this paper aimed at investigating the validity of the HIMATT analysis method against an established research and analysis method for complex problem solving, namely the problem-solving rubric analysis, that was driven from the think-aloud protocol analysis method (Ericsson and Simon 1984). The HIMATT analysis method provides automated analysis of students' causal representations of complex problem domain. On the other hand, protocol analysis method is an established method, but validity may suffer if a strong interrater reliability is not obtained. However, it requires extensive resources in both time and people, therefore, it is not scalable to serve as an assessment method that can be used in educational contexts. Table 3 provides an overview on the scientific quality, strength, limitations, and exploratory power of both methods.

Limitations and implications for future research

This study is limited in several aspects that must be addressed in future research. First, despite the reported significant correlations found in this study, it is important to point out that the indicators of both methods do not measure problem solving on an identical conceptual level. An established, valid and reliable method for assessing complex problem solving does not exist in the traditional sense of a measurement instrument (Johnson et al. 2006; Shute et al. 2009). Therefore, it is not possible to conduct a more elaborated validation study. Further, the small magnitudes of the correlation coefficients need to be considered regarding the overall significance of the results. On a positive note, the

Table 3 Comparison of indicators, scientific quality, and exploratory power of both analysis approaches

	PSR	HIMATT
Quantitative measures	Problem representation (including number of factors identified, elaboration of relationships among factors) Generating solutions (make a recommendation, justifying solutions, and evaluate solutions)	Structural measures Semantic measures Various graph theory measures (e.g., GRM, GAM, STM)
Qualitative measures	Qualitative and holistic analysis of the protocols, using the criteria specified in the rubrics as coding scheme Free coding providing additional information	Standardized re-representations for qualitative analysis
Objectivity	Achieved through refining the descriptors and criteria of the rubrics Achieved through verifying and discussing scores among the raters Rater agreement	Automated analysis of predefined raw data structure When assessed within HIMATT, no human coding required
Reliability	Achieved through interrater reliability	Tested (Pirnay-Dummer et al. 2010)
Validity	Grounded on ill-structured problem-solving literature Validated by panel experts	Tested (Pirnay-Dummer et al. 2010)
Practicability	Limited comparisons Single case analysis Small group analysis	Unlimited comparisons Single case analysis Large group analysis Stochastic analysis
Advantages	Acquire holistic interpretation of individuals' performance Customized to individual cases	Automated analysis Structural decomposition into key categories Recomposition into "re-representations"
Limitations	Time consuming	Output not interpretable by teachers or instructors without training

significant correlations found in this study among the relevant measures in both analysis methods provide assurances in the potential for developing a single, overall indicator in the HIMATT method that represents the quality of complex, ill-structured problem solving. An important future step for the development of a valid and reliable assessment method for complex problem solving would be an empirically tested overall indicator in the HIMATT tools that represents the quality of complex, ill-structured problem solving. The findings of this study suggest that such an overall indicator would consist of structural and semantic measures, which assesses the problem space of a specific subject domain. Therefore, we believe that the findings of this study are an important contribution to the current efforts in the literature towards establishing a valid and reliable method for assessing complex problem solving that is also highly scalable.

Second, this validation study only uses a single testbed and a limited number of assessment methods as well as a moderate sample size. Accordingly, future studies should aim on large sample sizes and different learning environments, Moreover, the focus should be on the best fit of available assessment and analysis methods (Al-Diban and Ifenthaler 2011; Shute et al. 2009; Spector 2009) and on the observation of learning-dependent changes during problem solving processes (Ifenthaler 2011a; Ifenthaler et al. 2011; Ifenthaler and Seel 2011). An essential requirement for future validation studies for complex ill-structured problem solving methodologies is the development of outside criteria which represent a high accuracy of fit with regard to the applied measures.

Third, the assessment method based on causal representation adopts the view of learning as placing learners on a trajectory towards expertise and tracks the development of problem representations of learners (Eseryel 2006; Ifenthaler 2008). Hence, this method only assesses cognitive aspects of complex problem solving. However, the findings of our previous studies suggest that affective factors such as motivation, metacognition, self-regulation, and epistemic beliefs also influence the development of complex problem-solving competencies (Ifenthaler and Eseryel 2013). For instance, we found that motivation during game play is a strong predictor of engagement, which in turn is a strong predictor for successful learning and complex problem solving game-based learning environments (Eseryel et al. 2013). Hence, there is a need for comprehensive methods that can validly and reliably assess the development of complex problem-solving competencies.

Conclusions

Due to the HIMATT method, which provides automated analysis of students' causal representations and compares each of them to the referent representation, it is very easy and relatively faster to feed the causal representations into HIMATT for analysis. However, since HIMATT was originally developed as a research tool, the output that HIMATT provides is not suitable for teachers' to interpret. Therefore, our future plans include developing a next-generation HIMATT technology as an assessment tool that teachers in K-12 or STEM faculty in undergraduate and graduate education could easily use in their classrooms. In this way, we are also compelled to merge the assessment processes within a model-facilitated learning framework to seek empirical evidence into its effectiveness and to further our collective understanding on facilitating learning in complex, ill-structured problem solving in STEM domains.

References

Achtenhagen, F. (2000). Reality, models, and complex teaching-learning environments. In J. M. Spector & T. M. Anderson (Eds.), *Integrated and holistic perspectives on learning, instruction, and technology: Understanding complexity* (pp. 159–174). Dordrecht: Kluwer Academic Publishers.

Akin, O. (1978). How do architectures design? In J. C. Latombe (Ed.), *Artificial intelligence and pattern recognition in computer-aided design* (pp. 65–119). Amsterdam: North-Holland.

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. (2011). Comparison of two analysis approaches for measuring externalized mental models: Implications for diagnostics and applications. *Journal of Educational Technology & Society, 14*(2), 16–30.

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

Baker, E. L., & Schacter, J. (1996). Expert benchmarks for student academic performance: The case for gifted children. *Gifted Child Quarterly, 40*, 61–65.

Belland, B. R., French, B. F., & Ertmer, P. A. (2009). Validity and problem-based learning research: A review of instruments used to assess intended learning outcomes. *The Interdisciplinary Journal of Problem-based Learning, 3*(1), 59–89.

Berliner, D. C. (2002). Learning about and learning from expert teachers. *International Journal of Educational Research, 35*(5), 463–482. doi:10.1016/S0883-0355(02)00004-6.

Bierhals, R., Schuster, I., Kohler, P., & Badke-Schaub, P. (2007). Shared mental models—linking team cognition and performance. *CoDesign, 3*(1), 75–94.
Chi, M. T. H., & Glaser, R. (1985). Problem solving ability. In R. J. Sternberg (Ed.), *Human abilities: An information processing approach* (pp. 227–257). San Francisco: W. H. Freeman & Co.
Clariana, R. B. (2010). Deriving individual and group knowledge structure from network diagrams and from essays. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 117–130). New York: Springer.
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.
Day, E. A., Arthur, W, Jr, & Gettman, D. (2001). Knowledge structures and the acquisition of complex skill. *Journal of Applied Psychology, 86*, 1022–1033.
Dijkstra, S., & van Merriënboer, J. J. G. (1997). Plans, procedures, and theories to solve instructional design problems. In S. Dijkstra, N. M. Seel, F. Schott, & R. D. Tennyson (Eds.), *Instructional design: International perspectives* (Vol. 2, pp. 23–43). Mahwah, NJ: Lawrence Erlbaum.
Dörner, D. (1987). On the difficulties people have in dealing with complexity. In J. Rasmussen, K. Duncker, & J. Leplat (Eds.), *New technology and human error* (pp. 97–109). Chichester, NY: Wiley.
Dörner, D., Kreuzig, H. W., Reither, F., & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität. [Lohhausen. On dealing with uncertainty and complexity]*. Bern: Huber.
Dörner, D., & Wearing, A. (1995). Complex problem solving: Toward a (computer-simulated) theory. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 65–99). Hillsdale, NJ: Lawrence Erlbaum.
Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: Bradford Books/MIT Press.
Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (revised ed.). Cambridge, MA: The MIT Press.
Eseryel, D. (2006). *Expert conceptualizations of the domain of instructional design: An investigative study on the DEEP assessment methodology for complex problem-solving outcomes*. Ph.D. Doctoral Dissertation, Syracuse University, Syracuse, NY.
Eseryel, D., Ge, X., Ifenthaler, D., & Law, V. (2011a). Dynamic modeling as cognitive regulation scaffold for complex problem solving skill acquisition in an educational massively multiplayer online game environment. *Journal of Educational Computing Research, 45*(3), 265–287.
Eseryel, D., Ifenthaler, D., & Ge, X. (2011b). Alternative assessment strategies for complex problem solving in game-based learning environments. 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. 159–178). New York: Springer.
Eseryel, D., Law, V., Ifenthaler, D., Ge, X., & Miller, R. B. (2013). An investigation of the interrelationships between motivation, engagement, and complex problem solving in game-based learning. *Educational Technology & Society* (under review).
Feldon, D. F. (2007). The implications of research on expertise for curriculum and pedagogy. *Educational Psychology Review, 19*(2), 91–110. doi:10.1007/s10648-006-9009-0.
Funke, J. (1985). Steuerung dynamischer Prozesse durch Aufbau und Anwendung subjektiver Kausalmodelle. *Zeitschrift für Psychologie, 193*(4), 443–465.
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. (2012). Complex problem solving. In N. M. Seel (Ed.), *The encyclopedia of the sciences of learning* (Vol. 3, pp. 682–685). New York: Springer.
Ge, X., & Land, S. M. (2003). Scaffolding students' problem-solving processes in an ill-structured task using question prompts and peer interactions. *Educational Technology Research and Development, 51*(1), 21–38.
Ge, X., Planas, L. G., & Er, N. (2010). A cognitive support system to scaffold students' problem-based learning in a Web-based learning environment. *Interdisciplinary Journal of Problem-based Learning, 4*(1), 30–56.
Gordon, I., & Zemke, R. (2000). The attack on ISD: Have we got instructional design all wrong? *Training, 37*, 43–53.
Guindon, R. (1988). *Software design tasks as ill-structured problems, software design as an opportunistic process*. Austin, TX: Microelectronics and Computer Technology Corporation.

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. (2010a). 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. (2010b). 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. (2011a). Identifying cross-domain distinguishing features of cognitive structures. *Educational Technology Research and Development, 59*(6), 817–840. doi:10.1007/s11423-011-9207-4.

Ifenthaler, D. (2011b). Intelligent model-based feedback. Helping students to monitor their individual learning progress. In S. Graf, F. Lin, Kinshuk, & R. McGreal (Eds.), *Intelligent and adaptive systems: Technology enhanced support for learners and teachers* (pp. 88–100). Hershey, PA: IGI Global.

Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Journal of Educational Technology & Society, 15*(1), 38–52.

Ifenthaler, D., & Eseryel, D. (2013). Facilitating complex learning by mobile augmented reality learning environments. In R. Huang, J. M. Spector, & Kinshuk (Eds.), *Reshaping learning: The frontiers of learning technologies in a global context* (pp. 415–438). New York: Springer.

Ifenthaler, D., & Lehmann, T. (2012). Preactional self-regulation as a tool for successful problem solving and learning. *Technology, Instruction, Cognition and Learning, 9*(1–2), 97–110.

Ifenthaler, D., Masduki, I., & Seel, N. M. (2011). The mystery of cognitive structure and how we can detect it. Tracking the development of cognitive structures over time. *Instructional Science, 39*(1), 41–61. doi:10.1007/s11251-009-9097-6.

Ifenthaler, D., & Pirnay-Dummer, P. (2013). Model-based tools for knowledge assessment. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed.). 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. (2011). A longitudinal perspective on inductive reasoning tasks. Illuminating the probability of change. *Learning and Instruction, 21*(4), 538–549. doi:10.1016/j.learninstruc.2010.08.004.

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: Theory, methodology, technology. Proceedings of the Second International Conference on Concept Mapping, Volume 1* (pp. 87–94). San José: Universidad de Costa Rica.

Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development, 45*(1), 65–94.

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development, 48*(4), 63–85. doi:10.1007/BF02300500.

Jonassen, D. H. (2004). *Learning to solve problems: An instructional design guide*. San Francisco: Pfeiffer.

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. (2011). *Learning to solve problems. A handbook for designing problem-solving learning environments*. New York: Routledge.

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 mind tools. 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., & Wang, S. (1993). Acquiring structural knowledge from semantically structured hypertext. *Journal of Computer-Based Instruction, 20*(1), 1–8.

Kauffman, D., Ge, X., Xie, K., & Chen, C. (2008). Prompting in web-based environments: Supporting self-monitoring and problem solving skills in college students. *Journal of Educational Computing Research, 38*(2), 115–137.

Kearney, E., Gebert, D., & Voelpel, S. C. (2009). When and how diversity benefits teams: The importance of team members' need for cognition. *Academy of Management Journal, 52*(3), 581–598.

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

Lachner, A., & Pirnay-Dummer, P. (2010). Model-based knowledge mapping. In J. M. Spector, D. Ifenthaler, P. Isaias, Kinshuk, & D. G. Sampson (Eds.), *Learning and instruction in the digital age* (pp. 69–86). New York: Springer.

LeBlanc, S. E., & Fogler, H. S. (1995). *Strategies for creative problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Lee, J. (2009). *Effects of model-centered instruction and levels of learner expertise on effectiveness, efficiency, and engagement with ill-structured problem solving: An exploratory study of ethical decision making in program evaluation*. Tallahassee, FL: Florida State University.

Mason, E. J., & Bramble, W. J. (1989). *Understanding and conducting research: Applications in education and the behavioral sciences*. New York: McGraw-Hill.

McKeown, J. O. (2009). *Using annotated concept map assessments as predictors of performance and understanding of complex problems for teacher technology integration*. Tallahassee, FL: Florida State University.

Means, B. (1993). Cognitive task analysis as a basis for instructional design. In M. Rabinowitz (Ed.), *Cognitive science foundations of instruction* (pp. 97–118). Hillsdale, NJ: Lawrence Erlbaum.

Newble, D., Norman, G., & Vleuten, C. (2000). Assessing clinical reasoning. In J. H. M. Jones (Ed.), *Clinical reasoning in the health professions* (pp. 156–168). Oxford, UK: Butterworth-Heinemann.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Perez, R. S., Fleming Johnson, J., & Emery, C. D. (1995). Instructional design expertise: A cognitive model of design. *Instructional Science, 23*(5–6), 21–349.

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

Pirnay-Dummer, P., Ifenthaler, D., & Spector, J. M. (2010). Highly integrated model assessment technology and tools. *Educational Technology Research and Development, 58*(1), 3–18. doi:10.1007/s11423-009-9119-8.

Reitman, W. R. (1965). *Cognition and thought: An information-processing approach*. New York: Wiley.

Robertson, W. C. (1990). Detection of cognitive structure with protocol data: Predicting performance on physics transfer problems. *Cognitive Science, 14*, 253–280.

Scandura, J. M. (1977). *Problem solving: A structural/process approach with instructional implications*. New York: Academic Press.

Scheele, B., & Groeben, N. (1984). *Die Heidelberger Struktur-Lege-Technik (SLT). Eine Dialog-Konsens-Methode zur Erhebung subjektiver Theorien mittlerer Reichweite*. Weinheim: Beltz.

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

Seel, N. M. (2001). Epistemology, situated cognition, and mental models: Like a bridge over troubled water. *Instructional Science, 29*(4–5), 403–427.

Seel, N. M., Al-Diban, S., & Blumschein, P. (2000). Mental models and instructional planning. In J. M. Spector & T. M. Anderson (Eds.), *Integrated and holistic perspectives on learning, instruction and technology: Understanding complexity* (pp. 129–158). Dordrecht: Kluwer Academic Publishers.

Seel, N. M., Ifenthaler, D., & Pirnay-Dummer, P. (2009a). 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., Pirnay-Dummer, P., & Ifenthaler, D. (2009b). Quantitative Bildungsforschung. In R. Tippelt & B. Schmidt (Eds.), *Handbuch Bildungsforschung* (pp. 551–570). Wiesbaden: VS Verlag für Sozialwissenschaften.

Shute, V. J., Jeong, A. C., Spector, J. M., Seel, N. M., & Johnson, T. E. (2009). Model-based methods for assessment, learning, and instruction: Innovative educational technology at Florida State University. In M. Orey (Ed.), *Educational media and technology yearbook* (pp. 61–79). New York: Springer.

Snow, R. E. (1990). New approaches to cognitive and conative assessment in education. *International Journal of Educational Research, 14*(5), 455–473.

Spector, J. M. (1998). The role of epistemology in instructional design. *Instructional Science, 26*(3–4), 193–203.

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. (2009). Adventures and advances in instructional design theory and practice. In L. Moller, J. B. Huett, & D. M. Harvey (Eds.), *Learning and instructional technologies for the 21st Century* (pp. 1–14). New York: Springer.

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., & Anderson, T. M. (Eds.). (2000). *Integrated and holistic perspectives on learning, instruction, and technology: Understanding complexity*. Dordrecht: Kluwer Academic Publishers.

Spector, J. M., Christensen, D. L., Siotine, A. V., & McCormack, D. (2001). Models and simulations for learning in complex domains: Using causal loop diagrams for assessment and evaluation. *Computers in Human Behavior, 17*(5–6), 517–545.

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.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Wood, P. K. (1983). Inquiring systems and problem structures: Implications for cognitive development. *Human Development, 26*, 249–265.

York, C. S., & Ertmer, P. A. (2011). Towards an understanding of instructional design heuristics: An exploratory Delphi study. *Educational Technology Research and Development, 59*(6), 841–863.

Deniz Eseryel is an Assistant Professor in the Instructional Psychology and Technology Program of the University of Oklahoma. Her research centers on issues related to promoting and assessing learning and development of expertise in complex knowledge domains such as STEM. Towards this direction, she investigates effective instructional approaches for designing advanced computer-based learning environments, such as dynamic modelling, simulations, and digital games to facilitate development of complex learning and ill-structured problem-solving skills.

Dirk Ifenthaler is currently the Fulbright Scholar-in-Residence at the Department of Educational Psychology at the University of Oklahoma. His research interest focuses on the learning-dependent progression of mental models, problem solving, decision-making, game-based learning, automated assessment procedures, and learning management systems.

Xun Ge is an Associate Professor with the Program of Instructional Psychology and Technology, Department of Educational Psychology at the University of Oklahoma. Her research interest includes designing learning technologies and open learning environments to support students' ill-structured problem solving and self-regulated learning, computer-supported collaborative learning, and virtual learning communities.