2. Using an Evidence-Based Approach to Assess Mental Models

Valerie J. Shute[1] **and Diego Zapata-Rivera**[2]

[1] Florida State University, Tallahassee, FL, USA

[2] Educational Testing Service, Princeton, NJ, USA

Abstract: This chapter describes a new idea for the design and development of assessments for mental models using "flexible belief networks" (FBNs). The idea involves joining and extending two assessment approaches—evidence-centered design (ECD) and concept mapping (CM). ECD will be extended beyond single, static proficiency models to dynamic models of learning over time. CM will be extended to include belief networks, which may be accomplished by overlaying concept maps with Bayesian networks. Our goal is to derive a methodology to better assess mental models as they evolve over time, with valid inferences regarding both syntactic (structural) and semantic (conceptual) similarities to reference models. This work leverages the seminal research conducted in the area of assessing mental models by Norbert M. Seel.

Keywords: Belief patterns; concept maps; mental models; formative assessment; evidence-centered design.

Introduction

One rich and enduring area of research in educational and cognitive psychology focuses on learners' construction and use of symbolic (or mental) models of knowledge. Mental models have been implicated in many phenomena that are fundamental parts of human cognition, such as the ability to reason—inductively and deductively—about complex physical and social systems, to generate predictions about the world, and to realize causal explanations for what happens around us (e.g., Gentner & Stevens, 1983).

In an increasingly technological society, understanding the nature of mental models for complex systems, and figuring out how to help people develop and hone good mental models are important goals with potentially large educational and economic benefits (e.g., Seel, 1999; Spector, Dennen, & Koszalka, 2006). In addition to knowledge and systems understanding, such constructed representations can also represent and communicate subjective experiences, ideas, thoughts, and feelings (e.g., Mayer et al., 1999; Seel, 2003).

Learners with access to good mental models demonstrate greater learning—outcomes and efficiency—compared to those with less adequate models in various domains (e.g., Mayer, 1989; DeKleer & Brown, 1981; White & Frederiksen, 1987), particularly mathematics and science. However, *assessing* these internal (hence invisible) mental models is a difficult task. Currently, to assess mental models, researchers often rely on learners' construction of external representations (e.g., concept maps) as a proxy for what resides inside the learner's head. And when the externalized maps are compared with experts' or other reference maps, structural similarities may be computed. But what about assessment of the quality or semantics of the underlying map? New methodologies in educational psychology and artificial intelligence are emerging which may help in this type of assessment effort. We will discuss this in more detail later in the chapter.

Besides difficulties associated with assessing mental models, instructing (or fostering) mental model construction is another large challenge. According to Seel (2003), there are three main instructional paradigms that have been used to promote model building: discovery learning, guided discovery learning, and the more common receptive learning that ensues from a teacher's explanation or an expert's demonstration. The basic premise underlying model-based instructional interventions (that are not purely discovery learning) is that providing learners with models—of tasks and/or representations of causal relations—facilitates knowledge and skill acquisition in the content area, particularly if the models are provided sufficiently early during the course of learning. But this premise is still largely unsubstantiated (see Johnson-Laird, 1989; and Seel, 2003 for more).

The glue that binds these ideas together is called evidence-centered design (ECD; e.g., Mislevy, Steinberg, & Almond, 2003) for assessment, which provides (a) a way of reasoning about assessment design, and (b) a way of reasoning about student understanding. For our purposes, ECD allows the assessment pieces to be joined together to form an informative profile of the learner, and provides the mechanism for specifying and linking concepts and propositions with appropriate evidence needed to demonstrate particular levels of proficiency (or belief). This will be discussed in the next section.

The organization of this chapter is as follows. We begin with some simple definitions to ground the ensuing discussion. This includes: (a) clarifying the distinction between mental models and concept maps; (b) specifying the underlying models and functionality of ECD (e.g., proficiency, evidence, and task models); (c) distinguishing between summative and formative assessment (see Black & Wliam, 1998a; 1998b; Shute, in press; Stiggins, 2002); and (d) distinguishing beliefs from knowledge. After defining key terms and concepts, we will summarize the important contributions of Norbert Seel to the field, and show how we plan to leverage this research for the purpose of assessing the structure and content of mental models using externalized representations. This will then pave the way for deriving innovative instructional interventions—using a formative assessment approach to assist learners in building better mental models.

Definitions

This section operationalizes and compares various terms and concepts including: (a) concept maps vs. mental models, (b) evidence-centered design models, (c) summative vs. formative assessment, and (d) beliefs vs. knowledge.

Concept Maps vs. Mental Models

Concept maps are *external* representations. They comprise the output or product emanating from the process of "concept mapping," which is a popular technique used for visualizing the relationships among different concepts. A concept map (or "causal influence diagram;" see Sterman, 1994; Spector, Dennen, & Koszalka, 2006) is usually a diagram depicting relationships among concepts. Concepts are connected to each other via labeled arrows, typically in a hierarchical structure. Some common links include: "is part of," "causes", "is required by," or "contributes to." Concept mapping began in the 1970s by Novak and colleagues to represent students' emerging knowledge of science (e.g., Novak, 1995; Novak & Gowin, 1984). It has subsequently been used as a tool to increase meaningful learning in the sciences and other subjects as well as to represent the expert knowledge of individuals and teams in education, government, and business.

Mental models are the *internal* representations of reality that people use to understand specific phenomena. Gentner and Stevens (1983) note that these internal models provide predictive and explanatory power for understanding interactions with the world around us. Mental models have also played a prominent role in cognitive processing theories. For instance, Johnson-Laird (1983) proposed that mental models are the basic structure of cognition, "It is now plausible to suppose that mental models play a central and unifying role in representing objects, states of affairs, sequences of events, the way the world is, and the social and psychological actions of daily life." (p. 397). Some characteristics of mental models include: (a) they are incomplete and constantly evolving; (b) they may contain errors, misconceptions, and contradictions; (c) they may provide simplified explanations of complex phenomena; and (d) they often contain implicit measures of uncertainty about their validity that allow them to used even if incorrect.

Evidence-centered Design — Models and Framework

Evidence-centered assessment design (ECD; Mislevy, Steinberg, & Almond, 2003) is a methodology for designing assessments based around the central question of how to gather evidence about a student's knowledge, skills, and abilities. ECD is a knowledge elicitation and management process whereby the goal is a detailed blueprint of the assessment called the conceptual assessment framework (CAF). The CAF is comprised of five different types of models, and a typical CAF contains multiples of each type:

- *Proficiency Model*—Describes students' knowledge, skills, and abilities about which we want to make claims.
- *Evidence Model*—Describes the relationship between observable outcomes from tasks and the relevant proficiency variables.
- *Task Model*—Describes the kinds of situations in which we can observe evidence of proficiencies.
- *Assembly Model*—Describes the collection of proficiency, evidence, and task models that will constitute a given assessment. It contains the rules used to assemble the form of the assessment seen by a learner from a pool of potential tasks.
- *Presentation Model and Delivery System Model*—Describes characteristics of a particular delivery environment, including format, platform and security considerations.

Almond and Mislevy (1999) describe how to use this framework to track the state of an individual learner as more and more observations arrive. The proficiency model, often represented by a Bayesian network (Mislevy, 1994; Almond et al., in press; Shute, Hansen, & Almond, 2007), is instantiated with the prior distribution over the proficiencies for a particular learner. When a set of observations from a task arrives, the appropriate evidence model is attached to the proficiency model and the evidence is absorbed. The evidence model fragment is then discarded and the proficiency model remains, tracking our beliefs about the knowledge, skills, and abilities of the student posterior to the observations.

Summative vs. Formative Assessment

> If we think of our children as plants... summative assessment of the plants is the process of simply measuring them. The measurements might be interesting to compare and analyze, but, in themselves, they do not affect the growth of the plants. On the other hand, formative assessment is the garden equivalent of feeding and watering the plants - directly affecting their growth. Clarke (2001, p. 2).

Summative assessment reflects the traditional approach used to assess educational outcomes. This involves using assessment information for high-stakes, cumulative purposes, such as promotion, certification, and so on. It is usually administered after some major event, like the end of the school year or marking period. Benefits of this approach include the following: (a) it allows for comparing student performances across diverse populations on clearly defined educational objectives and standards; (b) it provides reliable data (e.g., scores) that can be used for accountability purposes at various levels (e.g., classroom, school, district, state, and national) and for various stakeholders (e.g., students, teachers, and administrators); and (c) it can inform educational policy (e.g., curriculum or funding decisions).

Formative assessment reflects a more progressive approach to education. This involves using assessments to support teaching and learning. Formative assessment is tied directly into the fabric of the classroom and uses results from students' activities as the basis on which to adjust instruction to promote learning in a timely manner. This type of assessment is administered much more frequently than summative assessment, and has shown great potential for harnessing the power of assessments to support learning in different content areas and for diverse audiences. When teachers or computer-based instructional systems know how students are progressing and where they are having problems, they can use that information to make real-time instructional adjustments such as re-teaching, trying alternative instructional approaches, altering the difficulty level of tasks or assignments, or offering more opportunities for practice. Such events are, broadly speaking, formative assessment (Black & Wiliam, 1998a). Formative assessment has been shown to improve student achievement (Black & Wiliam, 1998b; Shute, Hansen & Almond, 2007).

In addition to providing teachers with evidence about how their students are learning so that they can revise instruction appropriately, formative assessments (FAs) may directly involve students in the learning process, such as by providing feedback that will help students gain insight about how to improve. Feedback in FA should generally guide students toward obtaining their goal(s). The most helpful feedback provides specific comments to students about errors and suggestions for improvement. It also encourages students to focus their attention thoughtfully on the task rather than on simply getting the right answer (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Shute, 2007). This type of feedback may be particularly helpful to lower-achieving students because it emphasizes that students can improve as a result of effort rather than be doomed to low achievement due to some presumed lack of innate ability (e.g., Hoska, 1993).

An indirect way of helping students learn via FA includes instructional adjustments that are based on assessment results (Stiggins, 2002). Different types of FA data can be used by the teacher or instructional environment to support learning, such as diagnostic information relating to levels of student understanding, and readiness information indicating who is ready or not to begin a new lesson or unit. FAs can also provide teachers or computer-based learning environments with instructional support based on individual student (or classroom) data. Examples of instructional support include: (a) recommendations about how to use FA information to alter instruction (e.g., speed up, slow down, give concrete examples), and (b) prescriptions for what to do next, links to web-based lessons and other resources, and so on.

Black and Wiliam (1998a; 1998b) very clearly established the importance of formative assessment to both teaching and learning. They also originated the widely-used distinction between (a) assessment *for* learning, and (b) assessment *of* learning, which maps to formative and summative assessment, respectively.

Knowledge vs. Belief

Everybody is entitled to their own opinion, but they're not entitled to their own facts.

–Daniel Patrick Moynihan

Although actual philosophers and epistemologists may quibble with the following definitions, we characterize knowledge and belief as follows. Knowledge is the comprehension or awareness of a verifiable idea, proposition, or concept, and the representation thereof. Belief refers to what one accepts as true, rejects as false, or withholds judgment about its truth-value (probabilistic). Furthermore, belief is a representational mental state that could be part cognitive and part affective. Knowledge typically has no affective aspects.

Sometimes the words 'know' and 'believe' are used interchangeably, but they are actually quite different. Belief typically applies to something that you are either unsure about or for which there is insufficient proof. For instance, one might say, "I believe that dogs make better pets than cats." This belief may (or may not) be true, and may be based on an overgeneralization or otherwise inadequate evidence. Knowledge, however, applies to things that are true (or that at least have a reasonable amount of supporting evidence). Therefore, it may be inappropriate to say, "I know that dogs make better pets than cats" because there is an element of doubt (i.e., disputable evidence) involved with this assertion. Knowledge implies belief.

How does knowledge relate to truth? Consider the following: until 1610, nobody *knew* that Jupiter had moons. Then there was a brief period of time when Galileo was the only person who knew Jupiter had moons. Eventually, larger numbers of people knew that Jupiter had moons. This shows that knowledge can change and be unevenly distributed, although the truth did not change in 1610. So, truth is something to be discovered while knowledge is something to be invented. In fact, much of scientific activity revolves around coming up with models that capture some aspects of the truth with some degree of fidelity. And that is just what we're attempting to accomplish with the ideas in this chapter. Now, going back to the example of Galileo's claim that Jupiter had moons, he had difficulty persuading others of this fact. Many simply did not want to believe that Jupiter has moons, and some people have a powerful ability to be blind to what they don't want to see.

So people hold all sorts of beliefs about the world around them. Some beliefs are more accurate than others—depending on the goodness of the evidence underlying the nodes in the belief network. As educators, we would like to be able to make valid inferences about what a person knows and believes, analyze how well that meshes with the body of knowledge and concepts to be learned, and then try to adjust errant or unfounded beliefs toward more reasonable and well-founded ones.

Having defined relevant terms, we now turn our attention to the current state of research in the area of mental models.

Current Research

Seel's contributions to the mental-model *researchscape* has direct relevance to our work in terms of the assessment of mental models. His approach opens up ways to capture important pieces of evidence relevant to aspects of knowing and learning that we have not done with ECD—namely modeling conceptual (or system) understanding. Heretofore, our assessment expertise and development efforts have focused on modeling declarative knowledge and procedural skills. However Seel et al.'s assessment tasks involve externalizing internal representations of conceptual and functional relatedness. Our tasks have tended to be more specific (defined) from an assessment point of view—capturing clear evidence directly from task performances (or from log files—see Shute, Ventura, Bauer, & Zapata-Rivera, in press). Representative tasks of this type include multiple-choice problems or constructed responses, where the key is a clear, known response. Cognitive models permit the analysis and comparison of responses to keys for diagnostic purposes.

Seel (2003) reported on the results from a long-term analysis of model-based teaching and learning. Among the important findings, the basic research on the development of mental models has shown that the models tend not be fixed structures of the mind, but are constructed by learners on an as-needed basis in response to a specific learning situation and associated cognitive demands. Seel thus concluded that mental models are situation-dependent constructions (or reconstructions) of previously generated models, are essential for problem solving, and may be captured via concept maps. Because concept maps are dynamic, adaptable, and interactive, they are well-suited for this purpose, and may be created and used by single persons or by small groups (Weinberger & Mandl 2003). Furthermore, the idea of using such flexible models to make inferences about what a learner knows and believes, to what degree, and the underlying reasons for these beliefs, comprises a great challenge to people who model how the mind works.

In previous assessment and learning research, the authors of this chapter have focused mostly on topics and tasks that (a) are typically well-defined, (b) have a correct solution (or constrained set of solutions), and (c) are free of controversial issues or indirect evidence. But leveraging Seel's research with mental models (and the progression thereof), provides us with an intriguing way to assess much richer mental representations, and to use that information to inform and update our proficiency models. These more comprehensive proficiency models can include information not only about procedural and declarative proficiencies, but also conceptual understanding and the underlying belief structures.

Assessing Concept Maps

In addition to providing a glimpse at internal mental models, concept maps help in organizing learners' knowledge by integrating information into a progressively more complex conceptual framework. When learners construct concept maps for representing their understanding in a domain, they reconceptualize the content domain by constantly using new propositions to elaborate and refine the concepts that they already know. More importantly, concept maps help in increasing the total quantity of formal content knowledge because they facilitate the skill of searching for patterns and relationships among concepts.

A variety of simple measures have been developed to measure completeness and structural complexity of concept maps. These indicators include the number of nodes, number of links, number of cross links, number of cycles, number of hierarchy structures, and number of examples (Vo, Poole, & Courtney, 2005; Novak & Gowing, 1984). Structural matching indicators, such as the deep structure measure from Seel's research, have also been used to determine how close a concept map is to a reference map (i.e., a concept map crafted by an expert) (Ifenthaler, Pirnay-Dummer, & Seel, 2007). Some of these simple indicators have been shown to be reliable and effective measures of the completeness and structural complexity of concept maps, and have been used to support research in the area (e.g., in relation to learning and intelligence). However, although such reliable and simple indicators play an important role in assessing certain characteristics of a concept model, they do not always provide enough information at the right granularity level to support instructional feedback (i.e., feedback that can be used by students to improve their learning).

Understanding the semantics or meaning of a concept map is a very challenging endeavor. The complexity of this problem can be handled by employing approaches that limit the scope of the concepts and relationships that can be represented and require the user to participate in the process to some extent, such as collaborative diagnosis (e.g., Cimolino, Kay, & Miller, 2004). Some of these approaches include: (a) asking students to select from a list of predefined organizational templates (organizers) representing various reasoning patterns (e.g., Ifenthaler & Seel, 2005; Jonassen, Beissner, & Yacci, 1993; Zapata-Rivera, Greer, & Cooke, 2000); (b) using a logic representation of the concept map, dialogue games, and sentence openers (e.g., Dimitrova, 2003; Jeong & Juong, 2007); and (c) using ontologies and teacher feedback to create a knowledge representation middle layer of the concept map that can be used to provide customized feedback to students (Cimolino, Kay, & Miller, 2004).

Both structural and semantic information can be combined in an evidence-based assessment framework (i.e., ECD). Computer-based learning tools developed on top of this framework can then use the information embedded in student concept maps to adapt their interaction. Monitoring the progress of concept maps over time (Seel, 1999; Ifenthaler & Seel, 2005) is an important goal to be achieved. But while the methods employed by Seel et al. are useful for tracking macro-level (or summative) changes in models over time, we also need a more micro-analysis approach to examine the factors that promote and inhibit specific concept mapping behaviors. We now present an extension of ECD to illustrate our plan for modeling student belief structures and their change over time.

Flexible Belief Networks

The basic idea we want to communicate herein concerns our approach to representing a learner's current set of beliefs about a topic as Bayesian networks (Pearl, 1988) that have been overlaid on top of concept maps. By overlaying a probabilistic network (i.e., a Bayesian network) on top of a concept map structure, we can model and question the degree to which relationships among concepts/nodes hold as well as the strength of the relationships. In addition, prior probabilities can be used to represent preconceived beliefs. A probabilistic network provides us with a richer set of modeling tools that we can use to represent the degree to which people ascribe to a particular belief pattern.

Accomplishing this goal would involve incorporating an assessment layer on top of the concept maps to flesh out the maps more fully. This approach would result in a collection of evidence from students in terms of their evolving mental models as indicated by their relationship to the strength and relevance of associations, directionality of the stated relations, and the specified type or nature of the relationship. The result should be a set of flexible belief networks (or FBNs).

To derive these FBNs, we would need to conduct a domain analysis on the topic in question, and use ECD to (a) model belief structures, and (b) design embedded assessments to gather evidence on learners' concepts, misconceptions, and beliefs. By employing embedded assessments, we will be able to infer a learner's current belief structure (via Bayesian networks) based on performance data (evidence) for a variety of purposes—e.g., to modify thinking, or increase cognitive flexibility and perspective taking. The benefits of such an approach are that it would render tacit (unobservable) knowledge and beliefs visible, and permit, if not actively encourage examination. Models (one's own and alternatives) may be displayed via "lenses" to enhance communication and understanding. Each lens would correspond to a particular belief "pattern" that was representative of, and fairly common in the population. The patterns, as will be discussed later, will be derived from both top-down (e.g., interviews with experts) and bottom-up (e.g., data mining) methods. This approach is expected to enable the modeling of changes in beliefs over time.

Figure 1 illustrates a simplified example of the progression from concepts to concept maps to belief nets when Bayesian networks are overlaid to specify structure, node size, and links (i.e., type, directionality, and strength of association). Evidence is attached to each node-relationship which either supports or counters a given claim. The example used here, for illustrative purposes only, represents some of the concepts and relations among variables related to the war in Iraq.

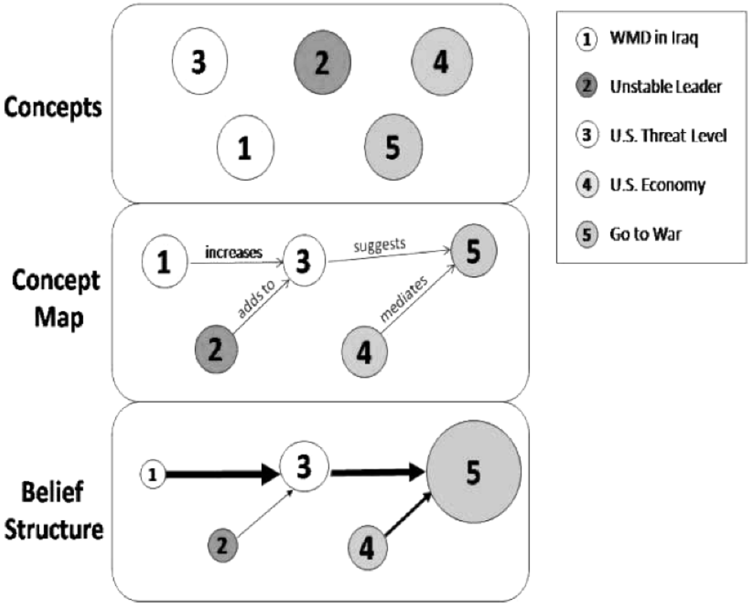

Fig. 1. Progression from concepts to concept map to belief structure.

Note that the *size* of the node in the belief structure indicates a given node's marginal probability (e.g., *p*(node 1 <weapons-of-mass-destruction> = True) = 0.1—a tiny node with a low probability of being true). *Links* illustrate the perceived relationships among the nodes in terms of *type, direction,* and *strength*. *Type* refers to the probabilistic or deterministic representation—defining the nature of the relationship. The s*trength* of the relationship is shown by the thickness of the link, and the *direction* indicates that the relationship has an origin and a destination. The belief structure in Figure 1 models the beliefs of a person (or group of people) that, for example: (a) nodes 1 and 3 exist, (b) the current probabilities of node 1 and node 3 are fairly low (0.1 and 0.3 respectively), and (c) there is a positive and strong relationship between nodes 1 and node 3 (represented by a thick line). So, if the low probability of node 1 (existence of weapons of mass destruction in Iraq) turned out to be true, then the effect on node 3 (U.S. threat level) would be a substantial elevation of the threat level.

A *belief pattern* (BP) is our term for a representative set of nodes and relations. lations. Continuing with the illustrative war in Iraq theme, following are two hypothetical BPs through the eyes of two fictitious persons who differ quite a bit in their respective beliefs about the war (see Figures 2 and 3).

When comparing the two BPs, they contain basically all of the same concepts, but the size of the respective nodes, the directionality of relations, and the strength

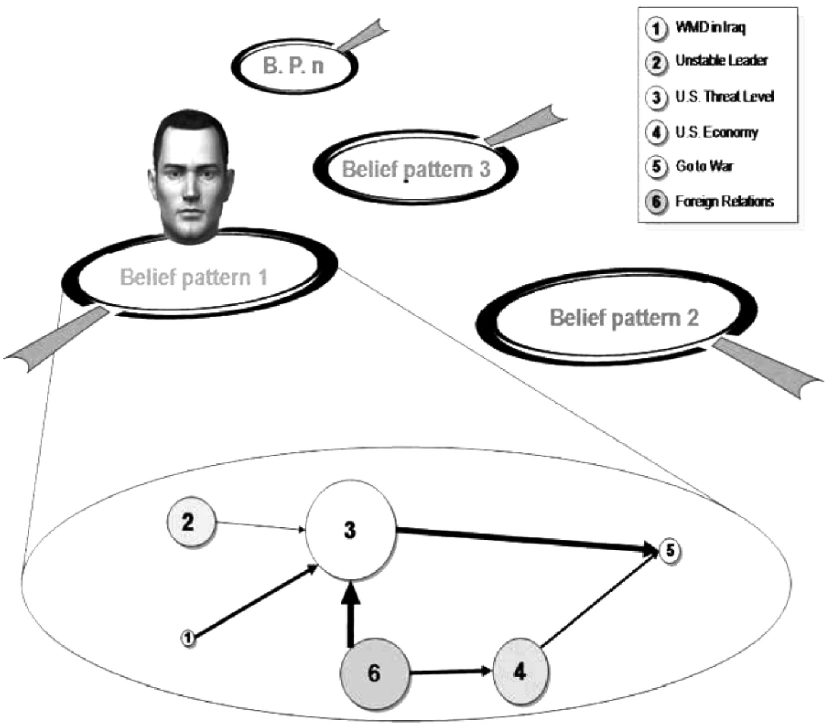

Fig. 2. BP through the lens of Person 1.

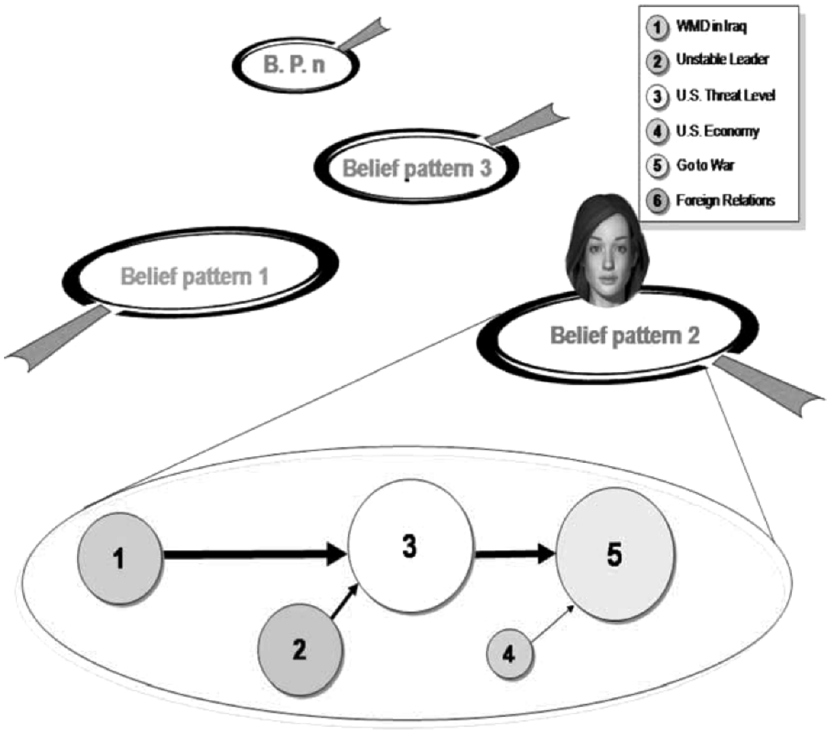

Fig. 3. BP through the lens of Person 2.

of the links are very different. Because we have chosen to use Bayesian networks to represent belief structures, this enables us to examine not only (a) the structure of the map, but also (b) the content (nodes and links), as well as (c) the underlying evidence that exists per structure (and per node). That is, as part of creating a current belief structure, the student arranges concepts and establishes links, and he or she includes specific evidence (sources) per claim (i.e., arguments and documentation in support of, or in opposition to a given claim). The credibility of the evidence, then, should match the strength of the links established in the structure. For instance, if a student made a strong claim about the existence of WMD in Iraq, and cited a dubious source as the only evidence, then that would not count as being credible evidence—and would imply that the student needed some assistance in his critical thinking/analysis skills. In short, we not only want to model the structures, but also the supporting evidence that lives underneath. Figure 4 shows a generic model with its supporting evidence attached.

So how do we accomplish this kind of modeling? There are five main parts to our proposed BP modeling approach:

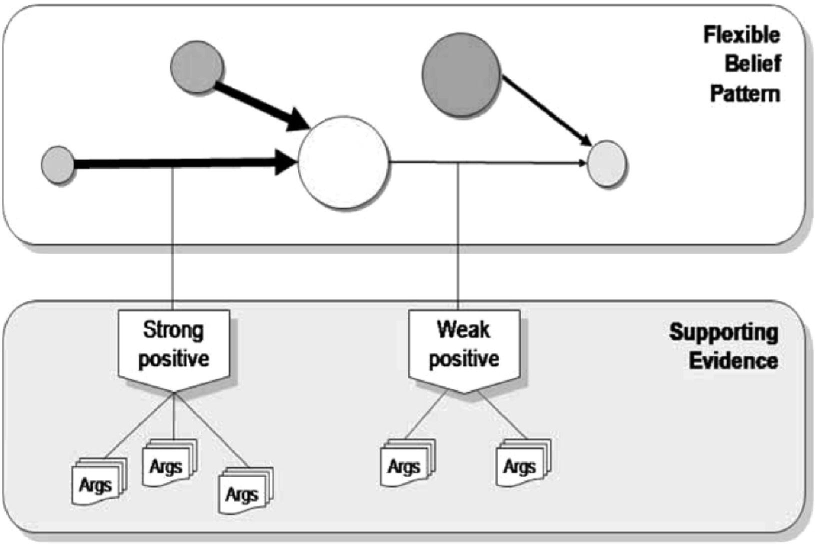

Fig. 4. Supporting evidence underlying an example BP.

1. Analyze the domain and integrate belief nets from human experts and data mining efforts.
2. Create an initial set of belief patterns (BPs).
3. Model beliefs, evidence, and assessments via extended ECD.
4. Infer individual BPs via assessment.
5. Model changing BPs over time via Dynamic Belief Networks (DBN).

The first step is to analyze the domain. This involves defining and structuring information about the topic area. It is instantiated as an FBN. Data sources for domain analysis include: (a) top-down creation of FBNs via ECD (e.g., subject-matter experts, research papers), and (b) bottom-up data mining to yield a large collection of variables relating to the topic, their relations from different perspectives, supporting arguments, claims, and so on. Data to be mined include: journal articles, blogs, listservs, newspapers, public documents, and data from surveys and tasks that students complete to further feed the models. This analysis phase is analogous to conducting a factor analysis on data to discern patterns.

The second step is to generate BPs. This may also be accomplished via top-down and bottom-up processes to effectively merge data from the analysis step – from data mining activities and subject-matter experts. This step informs the creation of the FBPs – both initial and alternative belief patterns.

The third step entails modeling using the proposed extended-ECD approach, and it has two main foci: designing valid assessments and diagnosing knowledge and beliefs. The assessment design process begins with defining three main models: (1) the belief model (BM)—What do you want to say about the person—what does she know and what does she believe is true?, (2) the evidence model (EM)—What observations would provide the best evidence for what you want to say?, and (3) the task model (TM)—What kinds of tasks or scenarios would allow you to make the necessary observations?. ECD thus provides a systematic way of laying out assessments, complete with evidentiary arguments that explicitly link performance data to claims about underlying knowledge and beliefs. Figure 5 shows three ECD models for a particular Belief Pattern (in this case, BP 1). Flowing from left-to-right, we depict the assessment design process from the Belief Model (labeled 'Current BP' in the Figure) to Evidence Model to Task Model. The Current BP model represents the initial *organization of concepts* (including preconceptions and misconceptions), beliefs, and relationships. Tasks will ultimately be designed to impose structure. Next, the Evidence Model specifies the criteria or rubrics needed for *evidence* of the current BP (i.e., specific student performance data, or observables). Finally, the Task Model contains a range of templates and parameters for *task development* to elicit data needed for the evidence model.

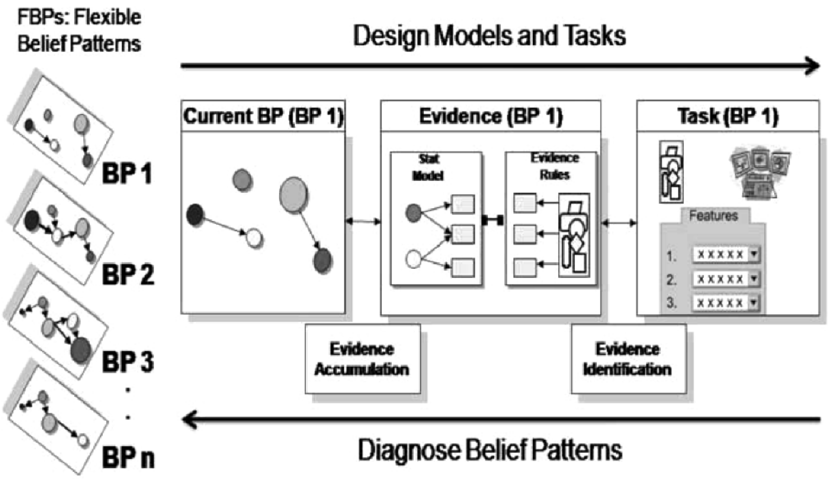

Fig. 5. Designing models and tasks based on extended ECD and diagnosing belief patterns based on students' performance data.

Reversing the flow (from right-to-left) permits diagnosis of what the learner knows/believes, and to what degree as related to each of the BPs. In Figure 5, "evidence identification" refers to the collection and scoring of data to analyze how the student performed while the "evidence accumulation" process refers to the derivation of inferences about what the student knows/believes, and how strongly it is known or believed.

The fourth step concerns the inference of belief patterns. That is, after links are inferred (direction, strength, and type), each student is associated with a particular

BP, at a particular point in time. Here, we propose to use embedded assessments to infer BPs from users' performance data (observables). These BPs may be mapped to initial BPs derived from the Domain Analysis part of the process. Assessment of BPs will include knowledge, concepts (preconceptions, misconceptions), links, argument structures, biases, and so forth. Environments (i.e., embedded tasks and interventions) may include virtual reality, simulations, and tasks like IAT (implicit association tasks; see Greenwald & Banaji, 1995) to reduce "faking" and get at deeply hidden beliefs. We expect that entrenched beliefs will be relatively easy to assess given strong and consistent response patterns. However, research has suggested that entrenched beliefs are harder to modify than existing knowledge (Griffin & Ohlsson, 2001).

The final step involves modeling BPs over time. To accomplish this goal, we plan to (a) use the extended ECD approach to track changes in BPs over time; (b) use ECD models to provide parameters to create different interventions (e.g., VR, simulations, etc.), and (c) assess each user at the beginning and end of a given "session" to see the effects of the intervention(s) on the students' BPs. Figure 6 depicts the modeling over time.

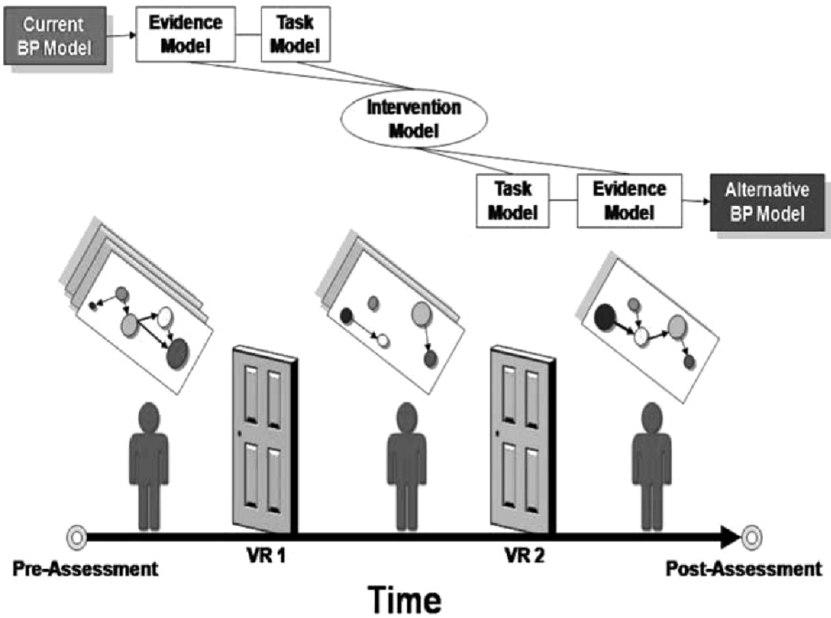

Fig. 6. Modeling BPs over time.

After the modeling part is accomplished, the next challenge will be to design effective and integrated interventions (e.g., making belief nets visible, showing others' nets, highlighting misconceptions, and so on) which must be coordinated because assessments will be embedded directly within the interventions.

Instructional Interventions

Once we know how mental models develop, and we can assess them syntactically and semantically, we will be able to set up interventions to foster their growth and development. That is, determining a learner's current BP is the critical first step in designing and delivering appropriate interventions. These interventions could include exposing the learner to an external representation of the current BP (e.g., an active belief map) and letting the learner react to it. For example, the learner could explain whether the current representation truly reflects what he/she believes, or if it does not, then why not. This can be done by allowing the learner to directly manipulate and annotate the current BP (Zapata-Rivera & Greer, 2004). We can also show the learner someone else's BP and ask her to compare it to her own BP. In fact, our intervention model, which serves to link ECD-based models (see Figure 6) can leverage Seel's model-based learning and instruction framework (Seel, 2003). That is, we can employ Seel's framework to design a variety of interventions that will help the learner analyze and reflect on his/her beliefs by using BP maps that change over time (e.g. Seel's "progression of mental models" concept). In short, we plan to combine Seel's framework with ideas underlying formative assessment as part of the instructional interventions.

Conclusion

Norbert Seel's foundational contributions to the areas of assessment and instructional use of mental models has informed and inspired many of our current ideas. There are still many challenges that lie ahead including: testing our FBN ideas across several "wicked" (i.e., ill-structured) topics, identifying conditions or factors that encourage or inhibit the processes of creating complex links between concepts/arguments, and creating effective interventions that make use of the rich mental model information.

We have described our idea for creating and using evidence-based flexible belief networks and their potential for serving as valid models for instructional intervention as well as communication tools that can be used to enhance learning, argument structures, and cognitive flexibility (e.g., Spiro et al., 1991). Some remaining research questions in this area include the following: If the ultimate goal is to diagnose entrenched BPs in order to help people acquire new knowledge and/or well-founded beliefs, how can we best exploit the information from the various models to create appropriate interventions? Also, how should the belief nets integrate knowledge and possibly affective aspects into the BPs? How broad and/or flexible should these FBNs be in relation to the scope of link types, node types, and so forth to be included in our BPs? Obviously, much more research is needed in this area, but we are very grateful for the firm foundation laid by

Norbert Seel and colleagues with their important research on mental models throughout the years.

Acknowledgments

In addition to Norbert Seel, we would like to acknowledge the following persons for their sage comments on an earlier draft of this paper: Allan Jeong, J. Michael Spector, Dirk Ifenthaler, Eric Hansen, and Jody Underwood.

References

Almond, R. G. (2007). *Cognitive modeling to represent growth (learning) using Markov Decision Processes*. Princeton, NJ: ETS Research Report.

Almond, R. G., & Mislevy, R. J. (1999). Graphical models and computerized adaptive testing, *Applied Psychological Measurement, 23,* 223-238.

Almond, R. G., Mislevy, R. J., Steinberg, L. S., Williamson, D. M., & Yan, D. (in press). *Bayesian networks in educational assessment,* manuscript in preparation to be published by Springer.

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

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-71.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.

Cimolino, L., Kay, J., & Miller, A. (2004). Concept mapping for eliciting verified personal ontologies, Special Issue on concepts and ontologies in WBES of the *International Journal of Continuing Engineering Education and Lifelong Learning, 14*(3), 212-228.

Clarke, S. (2001). *Unlocking formative assessment—Practical strategies for enhancing pupils' learning in the primary classroom*. London: Hodder and Stoughton.

deKleer, J. & Brown, J. S. (1981). Mental models of physical mechanisms and their acquisition. In J. R. Anderson (Ed.), *Cognitive Skills and their Acquisition.* Hillsdale, NJ: Erlbaum.

Dimitrova, V. (2003). StyLE-OLM: Interactive open learner modelling, *International Journal of Artificial Intelligence in Education, 13*(1), 35-78.

Forbus, K., Carney, K., Harris, R., & Sherin, B. (2001). *Modeling environment for middle-school students: A progress report*, in Biswas, G. (Ed.): Papers from the 2001 *Qualitative Reasoning Workshop*, The Printing House, Stoughton, WI.

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

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4-27.

Griffin, T., & Ohlsson, S. (2001). Beliefs versus knowledge: A necessary distinction for explaining, predicting, and assessing conceptual change. In J. D. Moore and K. Stenning (Eds.), *proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society* (pp. 364-369). Mahwah, NJ: Lawrence Erlbaum.

Hoska, D. M. (1993). Motivating learners through CBI feedback: Developing a positive learner perspective. In V. Dempsey & G. C. Sales (Eds.), *Interactive instruction and feedback* (pp. 105-132). Englewood Cliffs, N.J.: Educational Technology Publications.

Ifenthaler, D., Pirnay-Dummer, P., & Seel, N. M. (2007). The role of cognitive learning strategies and intellectual abilities in mental model building processes. *Technology, Instruction, Cognition and Learning,* Special Issue on Applying Cognitive Modeling.

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.

Jeong, A. & Juong, S. (2007). Scaffolding collaborative argumentation in asynchronous discussions with message constraints and message labels. *Computers & Education, 48*, 427-445.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge: Cambridge University Press.

Jonassen, D., Beissner, K., & Yacci, M. (1993). *Structural knowledge techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers Hove & London.

Mayer, R. E. (1989). Models for understanding. *Review of Educational Research, 59*(1), 43-64.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communication by minimizing cognitive load. *Journal of Educational Psychology, 91*(4), 638-643.

Mislevy, R. J. (1994). Evidence and inference in educational assessment, *Psychometrika, 12,* pp. 341-369.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspective, 1* (1), 3-62.

Novak, J. D. (1995). Concept mapping: A strategy for organizing knowledge. In S. M. Glynn, & R. Duit (Eds.), *Learning Science in the Schools: Research Reforming Practice* (pp. 229-245), Mahwah, NJ: Erlbaum Associates.

Novak, J. D., & Gowin, D. R. (1984). *Learning how to learn*. New York: Cambridge Press.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Kaufmann.

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

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

Shute, V. J. (in press). Tensions, trends, tools, and technologies: Time for an educational sea change. To appear in C. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning*. Mahwah, NJ: Erlbaum Associates.

Shute, V. J. (2007). *Focus on formative feedback.* ETS Research Report, RR-07-11 (pp. 1-55), Princeton, NJ.

Shute, V. J., Hansen, E. G., & Almond, R. G. (2007). An assessment for learning system called ACED: *Designing for learning effectiveness and accessibility*, ETS Research Report, RR-07-26 (pp. 1-54), Princeton, NJ.

Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (in press). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. To appear in U. Ritterfeld, M. J. Cody, & P. Vorderer (Eds.), *The social science of serious games: Theories and applications.* Philadelphia, PA: Routledge/LEA.

Spector, J. M., Dennen, V. P., & Koszalka, T. (2006). Causal maps, mental models and assessing acquisition of expertise. *Technology, Instruction, Cognition and Learning, 3*, 167-183.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology*, 24-33.

Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review, 10*(2-3), 291-330.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning, *Phi Delta Kappan Professional Journal, 83*(10), pp. 758-765, 2002.

Vo, H. V., Poole, M. S., & Courtney, J. F. (2005). An empirical comparison of causal mapping approaches. In D. J. Armstrong (Ed.), *Causal Mapping for Information Systems and Technology Research: Approaches, Advances and Illustrations* (pp. 142-173), Battle Creek, MI: Idea Group Publishing.

Weinberger, A. & Mandl, H. (2003). Computer-mediated knowledge communication, *Studies in Communication Sciences*, 81-105.

White, B. & Frederiksen, J. (1985). Qualitative models and intelligent learning environments. In R. Lawler & M. Yazdani (Eds.), *Artifical Intelligence and Education*. Norwood, NJ: Ablex.

Zapata-Rivera, J. D., & Greer, J. (2004). Interacting with Bayesian student models. *International Journal of Artificial Intelligence in Education, 14*(2), 127-163.

Zapata-Rivera, J. D., Greer, J., & Cooke, J. (2000). An XML-based tool for building and using concept maps in education and training environments. *Proceedings of the 8th Conference on Computers in Education*, ICCE 2000, 755-762.