DEVELOPMENT ARTICLE

Concept map engineering: methods and tools based on the semantic relation approach

Minkyu Kim

Abstract The purpose of this study is to develop a better understanding of technologies that use natural language as the basis for concept map construction. In particular, this study focuses on the semantic relation (SR) approach to drawing rich and authentic concept maps that reflect students' internal representations of a problem situation. The following discussions are included: (a) elaborate classifications of concept map approaches that use natural language responses (e.g., student essay); (b) the SR process of eliciting concept maps, established using studies on domain ontology; and (c) a more effective way to identify key concepts and relations from a concept map generated by the SR approach. By comparing the SR approach to other promising concept map technologies that constrain the analytical process in various ways, this study suggests that the SR approach is likely to draw richer and more authentic concept maps. In addition, this study suggests that a certain combination of graph-related metrics be used to filter key concepts from a SR concept map drawn from a written text of 350–400 words. The methods suggested in the study could be used to design an automated assessment technology for complex problem solving and to develop adaptive learning systems.

Keywords Assessment technology · Concept map · Mental model · Adaptive learning · Semantic relation

Introduction

Teachers need a precise but efficient way to assess student applications of understanding and complex problem solving. A simple knowledge-based (e.g., multiple choice or short answer) test is not sufficient to measure a learner's ability to solve complex problems (Champagne et al. 2000; Duschl 2003; Quellmalz and Haertel 2004); indeed, assessing

M. Kim (✉)
Center for Scholarly Technology, University of Southern California, 651 West 35 Street, Leavey Library (LVL) 301, Los Angeles, CA 90089-2571, USA
e-mail: kimmk@usc.edu

cognitive models of complex problem-solving knowledge and skills, such as scientific inquiry, can be challenging.

In educational research, the concept map technique has been employed as a way to represent and analyze a student's understanding of a complex problem situation (Clariana 2010; Novak and Cañas 2006; Spector and Koszalka 2004; Villalon and Calvo 2011; Zouaq et al. 2010) and used for educational applications that transform traditional instruction into student-centered adaptive learning (Azevedo 2011; Schwartz et al. 2009). However, the use of concept maps for educational assessment is relatively new, and the best approach for eliciting concept maps is still a matter of debate (Kim 2012a; Taricani and Clariana 2006; Zouaq et al. 2011a). Failure to use an appropriate and reliable concept map technique could disqualify a study or, more seriously, generate distorted information. To address this issue, the primary goal of this study is to build a better understanding of the nature of concept map technologies, particularly those that use natural language responses (e.g., student essay) as their sources.

Studies have explored various types of concept map techniques, including traditional and state-of-the-art technologies (Ifenthaler et al. 2009; Kim 2012a, b; Shute et al. 2009; Taricani and Clariana 2006). These techniques have been classified according to two criteria: graphical approach and natural language processing. The graphical approach (e.g., Cmaptools, Novak and Cañas 2006; Dynamic Evaluation of Enhanced Problem-Solving (DEEP), Spector and Koszalka 2004) allows students to draw their concept maps directly, adding link labels as annotated relations (e.g., "is associated" and "leads to") to complete the representation. Natural language processing features various methods (word association technique, Geeslin and Shavelson 1975; paired-comparison, Curti and Viator 2000; ordered recall, Frederick 1991; structural formation technique, Scheele and Groeben 1984; card sorting procedures, Frederick et al. 1994; and ordered tree technique, Naveh-Benjamin et al. 1986) and various technologies (KU-Mapper, Clariana et al. 2009; ALA-Mapper-Analysis of Lexical-Reader, Taricani and Clariana 2006; ALA-Reader, Clariana et al. 2009; and Text-Model Inspection Trace of Concepts and Relations (T-MITOCAR), Pirnay-Dummer and Ifenthaler 2010). Most of these tools enable students to judge directly the relation of a set of pre-defined concepts (e.g., KU-Mapper and ALA-Mapper) whereas some (e.g., Analysis of Lexical Aggregates-Reader (ALA-Reader) and T-MITOCAR) utilize written text, instead of pre-defined words, to create proximity data. The latter is called the open-ended approach (Clariana et al. 2009).

Kim (2012b) focused on the open-ended approach, arguing that language plays a critical role in building and mediating an individual's understanding and, furthermore, that using language as the basis for constructing a concept map is likely to represent more accurately the meaning and structure of the targeted internal knowledge (Kim 2012a; Pirnay-Dummer et al. 2010). Kim (2012b) selected two prominent technologies (i.e., ALA-Reader and T-MITO-CAR) and compared the technologies to an alternative semi-automated method devised to distill the semantic relations that are the underlying relations between two concepts expressed by words or phrases. The approach involved diverse types of relations of concepts beyond the typical noun–verb–noun relation form, including genitives (e.g., *teachers' participation*), prepositional phrases attached to nouns (e.g., *technology* in *school classrooms*), or sentences (e.g., *Emerging new media* have always led to *instructional changes*). However, the theoretical and methodological foundations underlying the two technologies and the alternative were not clarified, and the alternative method was not described in enough detail to portray how its attributes have the potential to create a better concept map technology.

The goals of this study were to extend attention to this alternative. First, this study established elaborate classifications of the "open-ended approach," detailing the

underlying assumptions and technical characteristics of a particular technology. Previous studies have rarely illustrated or compared the underlying linguistic assumptions and analytical methods of "open-ended" tools; in fact, no such studies were found in our literature review, but comparison of different tools facilitates progress in this area.

Second, we framed the approach known as "Semantic Relation" (SR) to guide the development of an automated concept map technology. This frame is based on the domain ontology in which concept map technologies are used to extract relevant knowledge structures from plain text documents (Concept Map Modeling (CMM), Villalon et al. 2010; Villalon and Calvo 2011; TEXCOMON, Zouaq and Nkambou 2008, OntoCmap, Zou-aqet al. 2011a).

Third, by comparing the SR approach to other promising concept map technologies that constrain the analytical process in various ways, this study argues that deep semantic structures (key concepts and relations) can be identified from naturalistic and rich knowl-edge representations from corpora. Various ways to identify deep structure were explored.

This study can enhance the use of concept maps in adaptive learning environments, such as technology-enhanced formative assessment and intelligent tutoring systems. Detecting the qualities of learner understanding as precisely as possible is critical to providing meaningful and productive instructional feedback suited to individual learning needs (Phelan et al. 2009; Shute and Zapata-Rivera 2007; Yorke 2003).

Mental models, natural language representations, and concept maps

The theory of mental models explains that a person builds his/her understanding by men-tally representing certain aspects of external situations that correspond to his/her precon-ceptions (Johnson-Laird 2005a, b; Norman 1986; Seel 2001, 2003). In that sense, the progress of mental models within an individual can be considered changes in knowledge structure toward an expected or desired state (Anzai and Yokoyama 1984; Collins and Gentner 1987; Seel 2001, 2003, 2004; Seel and Dinter 1995; Smith et al. 1993; Snow 1990). Problem solving includes conceptualizing a problem space as a more structured under-standing and integration of various ideas and concepts related to a problem (Dochy et al. 2003; Jonassen et al. 1993; Newell and Simon 1972; Spector and Koszalka 2004). Indeed, conceptualization of a problem space is a kind of mental model of a problem situation.

As a structural knowledge representation that consists of concepts and relations (Cla-riana 2010; Narayanan 2005; Novak and Cañas 2006; Spector and Koszalka 2004), a concept map can effectively assess a student's conceptualization of a problem space. Concept maps have been used to elicit cognitive representations of an individual's knowledge of a domain in which concepts are interrelated (Funke 1985; Narayanan 2005; Novak and Cañas 2006; Schvaneveldt 1990). The data used for concept maps are generally collected from interviews or texts. Text-based data collection is economical in terms of time and effort (Brown 1992) and is based on techniques that avoid recall bias and potentially leading or misleading questions (Axelrod 1976; Pirnay-Dummer et al. 2010).

Language is a symbol system, and mental models result from both perceptual and linguistic comprehension (Garnham 1987, 2001; Greeno 1989; Johnson-Laird 2005a, b; Seel 1999, 2001). As Fig. 1 illustrates, a mental model (1. Conceptual structure) is probably not the same as an internal linguistic semantic structure (2. Internal semantic structure) but embeds the properties of semantic representations (Bierwisch and Schreuder 1992; Kamp 1981; Kintsch 1994; Kintsch and van Dijk 1978; Levelt 1989). Natural language responses provide an individual's verbalized descriptions about a problem
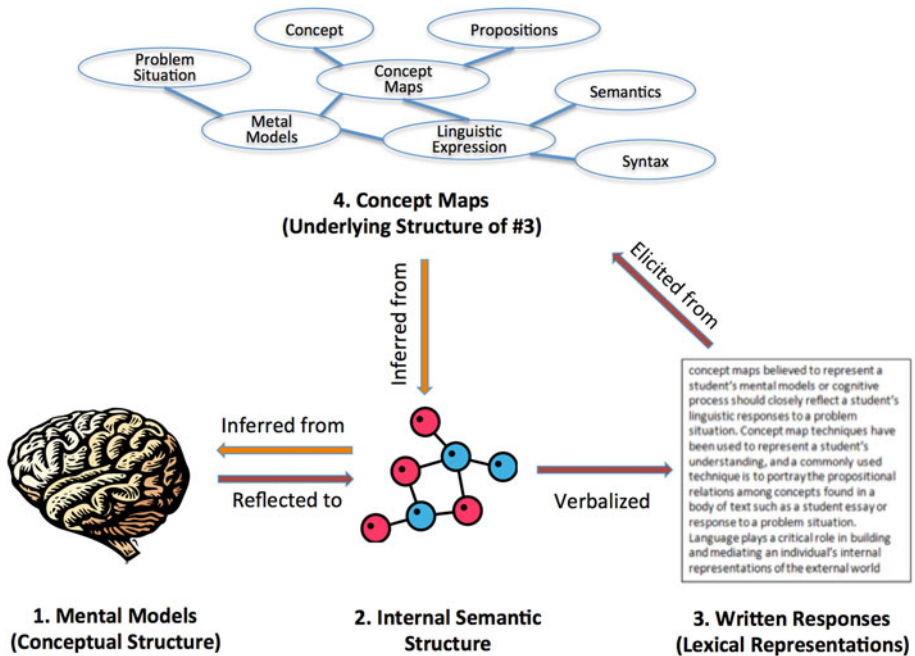
Fig. 1 The relations of internal and external representations

situation and feasible solutions (3. Lexical representations). Concept maps are external knowledge representations commonly elicited from natural language responses. The internal semantic structure is assumed to provide the underlying structure embedded in a visually rendered concept map.

Approaches to constructing concept maps can be characterized as ways of identifying concepts and determining the relations among those concepts. Defined as the associatedness of concepts in a text, *semantic relation* is a general term that does not indicate the specific nature of the relations that might be identified. However, this study operationalizes the term semantic relation (SR) as a way to extract underlying relations from deeper syntactic analysis. This study focuses on linguistic semantics, particularly Montague's (1974) semantics. His seminal work was to establish a systematic connection between syntax and semantics. From his point of view, natural language is a formal language based on predicate logic (Janssen 2012). The meanings in natural language expressions are determined by "a function of the meanings of its parts and of the way they are syntactically combined" (Partee 1984, p. 281).

Approaches to eliciting concept maps

The classification of concept map approaches in this study used the following procedure: First, promising concept map technologies that use natural language expressions as inputs were selected. The selection was based on three criteria: (a) technology invented within the last 5 years (i.e., since 2008); (b) technology with an accessible publication depicting its underlying mechanisms; and (c) technology eliciting a concept map on the basis of graph theory (Wasserman and Faust 1994). Second, selected technologies, related publications,

Table 1 Classification of concept-mapping approaches

	Adjacent relation (AR)	Proximity relation (PR)	Semantic relation (SR)
Approach	Implicit	Implicit	Explicit
Origin of relation	Spatial model	Spatial model	Syntactic model
Relation judgment	Adjacency	Distance	Syntactic/semantic
Direction of relation	Non-directional only	Non-directional only	Directional/non-directional
Exemplary tool	ALA-reader	T-MITOCAR CMM	OntoCmap[a]

ALA-Reader analysis of Lexical-Reader, *T-MITOCAR* text-model inspection trace of concepts and relations, *CMM* concept map mining, and OntoCmap

[a] OntoCmap is not open to the public

and technical documents were deeply analyzed. Last, upper-level characteristics that likely explain and distinguish the technologies were identified. The selected concept map technologies were ALA-Reader, T-MITOCAR, CMM, and OntoCmap.

The following questions guided the classification of concept map approaches:

- How do they distill concepts from a written response?
- How do they identify key concepts from a list of concepts?
- How do they identify relations between concepts?

This study grouped concept-mapping approaches into three classifications based on how propositional relations are treated: adjacent relation (AR), proximity relation (PR), and semantic relation (SR). In particular, the SR approach was determined to be an advanced way to create more authentic and complex concept maps when natural language is used for learner input.

Adjacent relation (AR)

AR is a spatial model because it assumes that closely connected concepts (words) tend to be physically closer to each other within a text (see Table 1). According to this model, verbs and other annotating words do not matter in a knowledge structure. When there is no annotated account between nouns, their relations are "implicit." Certainly, identifying concept–concept within a text is far easier than counting concept-annotation (e.g., verb-concept) structures. The AR method is simple; any two adjacent concepts in a text are considered strongly associated with each other, while concepts farther from each other are considered less directly associated. The former relationship has a strength value of 1, and the latter is coded as 0 (Clariana et al. 2009). For example, in the sentence "The success of active learning is determined by individuals' active engagement and contribution," there are maximum four identified relations: (a) success and leaning; (b) learning and individual; (c) individual and engagement; and (d) engagement and contribution.

ALA-Reader (Clariana and Koul 2008) is a tool designed to capture relations according to the adjacent relations of concepts (nouns) in a text. However, for parsimonious results, this technology only identifies a predefined set of concepts (up to a total of 30 nouns) in a student's text response. In an AR analysis, the two concepts need not be located in the same sentence. Consequently, the whole text is treated as one space, without any separation between sentences. AR cannot capture directional information about the relations because it is not concerned with annotating words (i.e., linking words such as verbs).

Proximity relation (PR)

Although proximity is commonly used as a general term to denote similarity, relatedness, and distance between concepts (Schvaneveldt 1990), PR is an approach to measuring the strength of relations according to the distance between concepts in a text (see Table 1). Because the relations are weighted according to the geometric distance between a pair of concepts, the PR approach is a spatial model (Schvanevldt et al. 1989) that assumes that the more two concepts are associated with each other, the closer they will be within or across sentences (Pirnay-Dummer and Ifenthaler 2010). The spatial model assumes that all concepts in a written artifact are basically associated with one another but have different degrees of distance. Due to the relations determined by spatial distance rather than semantic relation, PR does not capture logical information about relations. In the sentence used above ("The success of active learning is determined by individuals' active engagement and contribution"), PR could possibly extract ten pairs from five concepts (i.e., success, learning, individual, engagement, and contribution).

The PR approach must include a procedure for reducing the information in the network in order to identify meaningful relationships, for all possible pairs of all concepts can occur in a network. T-MITOCAR (Pirnay-Dummer and Ifenthaler 2010) is the technology used for the PR approach. The algorithms of T-MITOCAR generate distance data directly from a text using the number of words between two concepts. Detailed information about T-MITOCAR can be found in Pirnary-Dummer and Ifenthaler (2010). Like the ALA-Reader, the current version of T-MITOCAR limits the number of concepts to a total of 30 most frequent nouns.

Another tool is CMM (Villalon et al. 2010; Villalon and Calvo 2011). Unlike T-MITOCAR, this tool determines related words based on their locations in a hypothetical space modeled by a statistical method. Notably, the relations are not directly interpretable, so this approach is characterized as implicit. A distinguishable feature of the tool is its concept reduction algorithm, latent semantic analysis (LSA), which is a statistical dimensionality reduction technique (see Deerwester et al. 1990; Landauer et al. 1998). Briefly speaking, LSA locates extracted words in a multi-dimensional space using a method similar to factor analysis. LSA is usually based on a large corpus, but Villalon and Calvo (2009) used the technique to analyze a single text to generate a concept map.

Semantic relation (SR)

This study emphasizes the SR approach as a way to obtain richer, more authentic concept models, for the atomic units of meaning are clearly indicated by assigning a logical sense to the associated elements of a sentence (Zouaq et al. 2011b). Semantic relations are the underlying relations between two concepts expressed by words or phrases (Beamer et al. 2008; Girju et al. 2009). SR is similar to the general meaning of a proposition, in terms of the meaning of words, but SR involves diverse types of concept relations beyond the typical noun–verb–noun relation form (Adriana et al. 2004; Cañas 2009; Girju et al. 2009). Types of SR include complex noun compounds (e.g., "knowledge analysis"), genitives (e.g., "teachers' participation"), prepositional phrases attached to nouns (e.g., "community of practice"), or sentences (e.g., "Emerging new media have always led to instructional changes"). The example sentence ("The success of active learning is determined by individuals' active engagement and contribution") is deconstructed into the following units of meaning, each of which is assigned a semantic relation in parentheses (Adriana et al. 2004; Girju et al. 2009): (a) success of a active learning (Attribute-Holder); (b) individual's active engagement (Agent); (c) individual's active contribution (Agent); (d) engagement

and contributions (Associated With); (e) success is determined by active engagement (Cause-Effect); and (f) success is determined by active contributions (Cause-Effect).

As Table 1 summarizes, semantic relations are explicit in the sense that the meanings in a sentence are determined by the syntactic combination of its parts (syntactic models) (Montague 1974; Partee 1984; Zouaq et al. 2011b). In addition, the meaning of the relations is a core feature in the SR approach, helping define the direction of each relation. For example, in the phrase "learning progression toward expert level," the first concept (i.e., learning progression) is going in the direction of the second concept (i.e., expert level).

We assume that mental models are effectively depicted in integrated semantic networks and that core meanings (i.e., deep structures) can be filtered through the analysis of the semantic relations (i.e., surface structures). More recently, linguists have come to recognize that a sentence includes both a surface and an underlying deep structure, and that denotation and connation are relevant to semantic analysis. According to Katz and Postal (1964), the surface structure (i.e., syntax) characterizes the shape of the sentence, while the semantic information of the deep structure accounts for a substantial part of the meaning (Bransford and Franks 1972; Bransford et al. 1972; Bransford and Johnson 1972; Fodor et al. 1974).

To the best of our knowledge, the state-of-the-art technology that best employs the SR approach is the OntoCmap (Zouaq et al. 2010, 2011a, b). The purpose of the tool is not to elicit or assess student models but to obtain the domain ontology by filtering key elements from the concept maps elicited from plain text documents in a given domain. In spite of its distinct goal, the embedded ideas and mechanisms needed to build concept maps are very similar to those envisioned in the study. Thus, the current study developed the SR approach based on the OntoCmap research while the goals and mechanism details focused on formative assessment and instructional support. Unfortunately, because the OntoCmap is not accessible to the public, its analysis was not included. Note that the SR approach in this study is neither semi nor fully automated technology. This study first defined a sequence of phases and associated mechanisms for building concept maps; then the semantic relations were obtained manually (except for the concept map metrics). The following sections briefly introduce the process depicted in Fig. 2.

Phase 1: identify concepts

The first step is to analyze the syntax of sentences and distill concepts from a written response based on a set of rules:

- Rule 1: A concept can take one of several noun forms (Girju 2011; Girju et al. 2009, 2010; Moldovan and Girju 2001; Murphy 2003; Rijkhoff 2002): (a) one-word noun, (b) noun compounds, and (c) noun and adjective pre-modifier.
- Rule 2: Distilled concepts (nouns) are primarily stored as singular.
- Rule 3: Pronouns are not replaced with the nouns they represent (Kim 2012b).

The distilled concepts include three types of nouns (Girju 2008): one-word nouns (e.g., practice, technology, and classroom); noun compounds that consist of the head noun and noun modifier(s) (e.g., "bus station" and "technology implementation"); and noun and adjective pre-modifiers (e.g., "technological intervention").

Phase 2: build concept library

Once concepts are distilled from the text, the concepts are stored for the individual response and then all concepts from all responses are grouped based on synonym sets. This
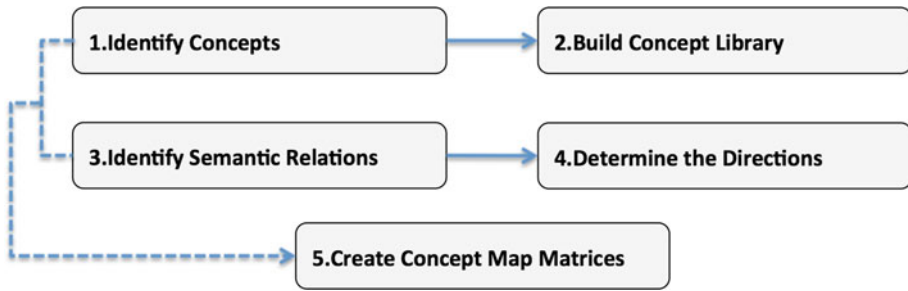
Fig. 2 The process of the Semantic Relation (SR) approach

study argues that an assessment for a certain problem situation requires a situation-specific set of synonyms. The sets can be built based on two sources: general synonym library and situation-specific terms to address polysemy. Although each concept is basically regarded as having a unique meaning, concepts sharing the same meaning or very similar meaning in the domain should belong to a single category (Moldovan and Girju 2001). For example, "chalkboard" could be contextually synonymous with "blackboard" and "whiteboard."

Phase 3: identify semantic relations

Identifying semantic relations is the most important step because it provides information used for creating concept maps. This step is to select pairs of concepts, C_i and C_j, linked by a particular semantic relation. The principles determining the pairs of concepts are established according to linguistics studies (Downing 1977; Girju 2008; Hearst 1992; Levi 1978; Moldovan and Girju 2001). The semantic relations are determined by syntactic patterns classified as phrase-level patterns and sentence-level patterns (Girju 2008; Hearst 1992). Zouaq et al. (2011b) divided the patterns into modifiers and core syntax.

First, phrase-level patterns (modifier patterns) include prepositional phrases attached to nouns (noun phrases) or s-genitives. For example, "the library of the school" is interpreted as having a semantic relation of "part-whole." A list of eight prepositions (of, for, in, at, on, from, with, and about), as compiled by Lauer (1995), plays a critical role in determining the semantic relations based on algorithmic patterns. Second, the semantic relation is determined by the sentence (core patterns). For example, in the sentence "the school has a new technology," the relation of "school" and "technology" is categorized as possessive. Moldovan and Girju (2001) defined twenty-two types of semantic relations. However, a natural language expression is not always a simple sentence, and multiple patterns can exist in a single sentence. Thus, this study added fourteen rules for determining semantic relations at the sentence level (see Appendix A).

Phase 4: determine the direction of relations

This phase is to determine the directional relations between two paired concepts, if necessary. The relation in which the path begins with the first concept and ends with the second concept is classified as follows: subject to object; source to target; from A to B; cause to effect; mutual relation (the first to the second); A belongs to B or B includes A; superior to inferior; A exists for B; A serves for B; tool to object; person to object; nouns linked with eight prepositional modifiers (the first to the second).

Phase 5: compute concept map metrics

Finally, in order to construct a concept map, all concepts distilled from a text response are listed and paired with one another in a matrix. The paired concepts that are semantically related are given a vector value of 1 in an n-by-n concept array, where n is the number of concepts; otherwise, the vector value is 0. In addition, to include directional information, the first concept (C_i) is considered the source, and the second concept (C_j) becomes the target. A semantic analysis technique, such as social network analysis based on graph theory, can generate a variety of concept map metrics that characterize an individual concept map—an individual's knowledge representation (Wasserman and Faust 1994). Some metrics could be indicators used to find the deep structure (i.e., key concepts and relations) emerging from the whole semantic structure.

Methods

This study is grounded on two assumptions similar to the ideas of Zouaq et al. (2011a, b): (a) the SR approach is one way to elicit an authentic and rich concept map from a text response, and (b) the core meaning (key variables and propositions) can be filtered from an authentic concept map based on semantic relations. In order to test the assumptions, this study compared concept maps constructed using the SR approach with those from CMM, T-MITOCAR, and ALA-Reader, exemplary technologies for the PR and AR approaches. For the second assumption, drawing on various concept map matrix measures, different sets of key concepts were filtered and compared with the set of concepts identified by human experts (i.e., standards) in order to develop a matrix measure that could reliably and accurately filter core meanings. Notably, the interpretation of comparison results should be limited to the selected technologies and carefully generalized to the associated approaches because each technology has particular pros and cons.

Participants

Participants included seven professors teaching at six major universities in the United States. As panel members, the professors were asked to complete Delphi surveys to establish a reference model for a complex problem related to technology implementation in K-12 schools. The professors were selected based on several pre-set criteria: (a) professor in Instructional Technology or related fields; (b) professor teaching a course titled Instructional Design or Technology Integration in Learning; (c) professor researching technology-integration in classroom learning; and (d) professor whose doctorate was received at least 3 years earlier. An invitation letter was sent out to prospective professors, and seven of the professors agreed to participate.

The problem-solving task

The panel was asked to respond to a complex problem situation using natural language. The task simulated participation in an evaluation project, the purpose of which was to investigate an unsuccessful project that had as its goal to adapt a technology (i.e., a tablet PC) for classroom teaching. In order to elicit the professors' knowledge in detail, the questions asked them to describe in detail the concepts, issues, factors, and variables likely to have contributed to the result, that the introduction of tablet PCs had very little effect on the instructional practices employed in the classes (see Appendix A).

Reference model

The reference model, in the form of a written response, was created according to the Delphi survey procedures (Goodman 1987; Hsu et al. 2007; Okoli and Pawlowski 2004), and the panel selected 23 key concepts—gold standards—from all of the concepts in the reference response. The Delphi survey included three rounds to refine the reference response (see Table 2). In the first round, the participating experts created their own responses to the problem. All responses from the panel were consolidated. In the second round, a document including all statements and a list of identified concepts was sent back to the panel. The experts were asked to add their comments to the listed statements and concepts and to rank them. After the second survey, the researcher created a final list of ranked statements and concepts. Based on the summary, a draft of the reference model was created. In the final round, the results of the second survey were sent to the panel and revised according to their comments.

Analysis procedure

Constructing concept maps

A total of eight responses, including (a) the initial responses of seven professors and (b) the reference model, were used for the study. Concepts and relations in the SR approach were manually distilled according to the procedure described earlier. Competitive data were created using the CMM, T-MITOCAR, and ALA-Reader. The ALA-Reader tool required a predefined list of concepts in order to distill the relations of the concepts from the text. The 23 terms (concepts) defined by the expert panel were used for the ALA-Reader. The data took the form of a matrix and were transformed into concept maps processed by NodeXL version 1.0 software (http://nodexl.codeplex.com/).

Identifying key concepts

This study suggests that the core elements of a concept map can be filtered from an authentic concept map elicited using the SR approach. Some metrics calculated from graph theory can be used to weigh individual concepts and then select more important elements. For example, Zouaq et al. (2011b) used metrics such as Degree, Betweenness Centrality, and PageRank (see Wasserman and Faust 1994), which yield values between 0

Table 2 Delphi procedure

Round	Activity
R1. Brainstorming	• Collect and consolidate all responses from experts
R2. Narrowing down/ ranking	• Send refined final version of consolidated lists, including statements and used concepts • Ask experts to add comments if they disagree with or have different opinions about a statement • Ask experts to rank key statements and concepts
R3. Refinement	• Send each panelist ranked statements and concepts summarized by the investigators • Ask for revision of judgments or specification of reasons for remaining outside the consensus

and 1. This study used NodeXL to calculate those metrics. According to the current study, Betweenness Centrality is the most meaningful measure, given its theoretical definition. The centrality measure is based on the assertion that a concept can exert control over the interaction between other pairs of concepts in a network (Anthonisse 1971; Freeman 1977).

Zouaq et al. (2011b) proposed a combination of the metrics Degree, Betweenness, and Page Rank to rank concepts, according to the following scheme:

$$\text{TVoted} = \text{TDegree} \cap \text{TBetweenness} \cap \text{TPage Rank}$$

In this scheme, all metrics are considered equally important, and a term is only a candidate when its value is greater than or equal to the mean value of the metrics. They suggested that any candidate term is important when the term belongs to all three metrics. However, as they admitted, the mean value as the threshold might be too restrictive and lack empirical evidence.

The current study validated the competitive measures that might indicate which terms are more important using a sample text response, the reference model. For the TVoted terms, we set two thresholds: 25 and 50 % quartile. The 25 % quartile means that a term is considered a candidate if its value is greater than or equal to the value of 25 % quartile of the metrics, a more generous criteria for term selection. Terms that had a Betweenness Centrality value of no less than zero were included as well. In addition, the terms from CMM and T-MITOCAR were used for comparison. However, the ALA-Reader tool was not included because it used a predefined set of 23 terms that human experts determined to be key concepts. The 23 terms that human experts selected worked as gold standards for the comparison.

Data comparison

Comparisons between the different technologies (SR, CMM, T-MITOCAR, and ALA-Reader) followed two paths: (a) general features of concept maps elicited from different technologies were reviewed using a quantitative and qualitative method, and (b) the sets of key concepts identified by the technologies or filtering methods were compared against the gold standards set by experts.

To compare these sets of key concepts, two types of similarity measures were applied: (a) numerical similarity and (b) conceptual similarity. The comparisons of the number of key concepts were derived from

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

where f_1 and f_2 denote the numerical frequency of each method compared. Conceptual similarity, indicating the extent to which the paired models shares the same concepts and relations, was calculated using Tversky's (1977) formula:

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

where α and β are weights for differentiating the quantities of A and B. This study assumes that there is no difference in the weights. Thus, α and β were both set to 0.5 ($\alpha = \beta = 0.5$). As for the qualitative review, we visually inspected the elicited concept maps.

Comparisons of concept maps

Concepts and relations

The numbers of concepts and relations obtained from the different technologies were investigated (see Table 3). The values were higher for the SR approach than the others. For example, the number of concepts obtained using T-MITOCAR ranged from 26 to 40 % the number obtained using SR; the ALA-Reader identified only 13 to 43 %. The number of relations identified using T-MITOCAR ranged from 34 to 75 % the number identified by SR; the ARA-Reader managed only 11 to 55 %. CMM reached 48 to 90 % the number of concepts obtained by SR and 38 to 70 % the number of relations. However, the numbers for expert 7 exceeded those achieved by SR: 116 % for concepts and 111 % for relations.

Overall, the SR approach generated the richest concepts, regardless of the relations, followed by CMM, T-MITOCAR, and ALA-Reader. This finding was in part affected by the fact that, unlike SR, the other technologies contain data reduction mechanisms that restrict the amount of concept map information. The number of concepts in CMM sharply increased at a certain corpus size (between 500 and 750 words). This finding implies that the reduction of dimensionality in LSA is too restrictive given a smaller set of words; indeed, LSA was originally used to analyze large corpora with multiple texts.

As Fig. 3 illustrates, measuring the association of response word count and the number of identified concepts or relations revealed high variation among the approaches. Admittedly, a sample size of 8 is not enough to produce generalizable trends. Nonetheless, this exploratory review provides some useful findings: SR showed a moderately positive line with word count. In contrast, the other three technologies showed little relation to word count. These results imply that the latter tend to be abstractive in the way they describe knowledge structures and are likely missing the semantic information of the entire structure. Some sharp drops were observed in AR, even in responses with higher word counts; in CMM, there was a steep increase.

Key concepts

This study assumes that a substantial part of the meaning of a text can be derived from the complex concept map composed of semantic relations (Fodor et al. 1974; Katz and Postal 1964). This assumption was tested by comparing various filtering methods using the reference response. As Table 4 demonstrates, human experts selected a total of 23 key concepts (Standards). Three filtering methods used in SR rendered three different sets of key concepts: Betweenness (24), 50 % quartile threshold (10), and 25 % quartile threshold (24). T-MITOCAR filtered 14 key concepts based on their frequency, while CMM rendered 26 key concepts. The metrics used for SR seem to be helpful in identifying key concepts (matched % ≥ 70).

Numerical and conceptual similarities were calculated according to the formulas introduced in the Methods section. The metrics obtained by graph theory—SR (BT) and SR (25)—outperformed the metrics from the competitive technologies (see Tables 5 and 6). The similarities achieved using SR were reasonable. The concepts identified using the Betweenness and 25 % quartile filters for TVoted in SR were identical (i.e., similarity of 1). The numerical similarity with Standards was 0.96, and the conceptual similarity was 0.72, whereas the numerical and conceptual similarity between the 50 % quartile for TVoted and Standards were 0.43 and 0.42, respectively. Overall, these results prove that SR can create a rich and authentic concept map embedding key concepts that are filtered by

Table 3 Descriptive statistics: the number of concepts and relations

Model	Word count	Concepts (nouns)				Relations (links)			
		SR	LSA	TM	AR	SR	LSA	TM	AR
Reference	397	54	26 (48 %)	14 (26 %)	23 (43 %)	64	24 (38 %)	22 (34 %)	35 (55 %)
Expert 1	319	41	26 (63 %)	13 (32 %)	8 (20 %)	43	20 (47 %)	18 (42 %)	7 (16 %)
Expert 2	411	42	33 (79 %)	11 (26 %)	10 (24 %)	54	34 (63 %)	24 (44 %)	15 (28 %)
Expert 3	331	37	27 (73 %)	13 (35 %)	9 (24 %)	50	28 (56 %)	22 (44 %)	15 (30 %)
Expert 4	297	29	26 (90 %)	10 (34 %)	5 (17 %)	30	21 (70 %)	12 (40 %)	5 (17 %)
Expert 5	436	48	27 (56 %)	19 (40 %)	6 (13 %)	55	25 (45 %)	41 (75 %)	6 (11 %)
Expert 6	481	52	30 (58 %)	17 (33 %)	7 (13 %)	62	34 (55 %)	33 (53 %)	8 (13 %)
Expert 7	751	50	58 (116 %)	13 (26 %)	10 (20 %)	63	70 (111 %)	23 (37 %)	22 (35 %)

The values in parentheses are percentages of corresponding SR amount

SR semantic relation, *CMM* concept map mining, *TM* T-MITOCAR, *AR* ALA-Reader
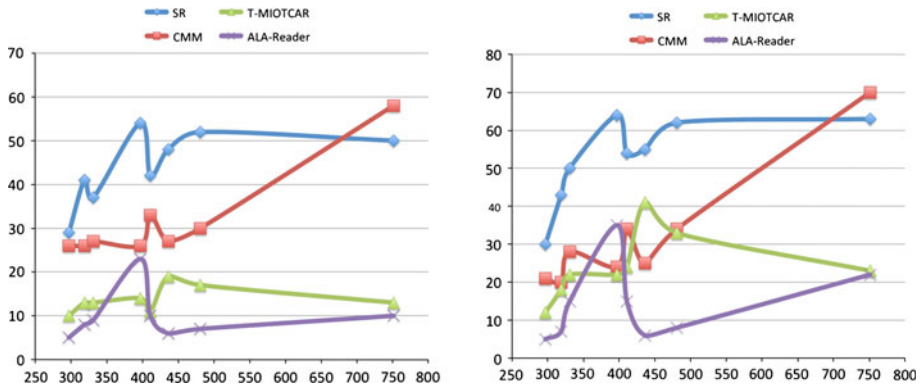
Fig. 3 Association of response word counts with the number of concepts or relations. The *left* associations between the number of concepts and word count, and the *right* represents associations between the number of relations and word count

Table 4 The number of key concepts in the reference model

	N	Matched N[a]	Matched (%)[b]
Standards	23	–	–
T-MITOCAR	14	9	64
CMM	26	11	42
SR (BT)	24	17	71
SR (50)	10	7	70
SR (25)	24	17	71

SR (BT) denotes the key concepts filtered by the Betweenness measures. SR (50) indicates the key concepts filtered by the 50 % quartile threshold for TVoted. SR (25) indicates the key concepts filtered by the 25 % quartile threshold for TVoted

[a] The number of concepts out of N that matched concepts in Standards

[b] The percentage of N that matched Standards

graph theory-related metrics. For analyzing a single student's written response to a problem in a learning situation, simple Betweenness or 25 % quartile threshold could be the most effective and accurate measure.

Visual inspection

A visual inspection of the concept maps drawn from the different approaches revealed a clear distinction among the selected technologies. First of all, as Fig. 4 depicts, the concept maps of SR were much richer and more complex for all samples, in both the number of concepts and the number of relations, than the other technologies. The concept maps of the reference response were highly coherent and connected, regardless of the approach, because the reference response was written very carefully to connect key concepts.

Substantial differences were found in the concept maps for Expert 6. The concept map drawn from SR contains two subsets that share no connection with the rest of the network. This study argues that the distinction made by SR between independent subsets reflects an actual knowledge structure and helps accurately describe the current under-standing of a student. However, the possibility that all concepts, without exception, will

Table 5 The number of key concepts similar or dissimilar among the approaches in the reference model

Pair (A with B)	$f(A - B)$	$f(B - A)$	$f(A \cap B)$
ST with TM	14	5	9
ST with CMM	12	15	11
ST with SR (BT)	6	7	17
ST with SR (50)	16	3	7
ST with SR (25)	6	7	17
TM with CMM	8	20	6
TM with SR (BT)	5	15	9
TM with SR (50)	9	5	5
TM with SR (25)	5	15	9
CMM with SR (BT)	14	12	12
CMM with SR (50)	19	3	7
CMM with SR (25)	14	12	12
SR (BT) with SR (50)	14	0	10
SR (BT) with SR (25)	0	0	24
SR (50) with SR (25)	0	14	10

ST is the 'Standards' human experts judged

Table 6 Similarities of key concepts in the reference model

	1	2	3	4	5	6
1. Standards	–	0.61	0.88	0.96	0.43	0.96
2. T-MITOCAR	0.49	–	0.54	0.58	0.71	0.58
3. CMM	0.45	0.30	–	0.92	0.38	0.92
4. SR (BT)	0.72	0.47	0.48	–	0.42	1.00
5. SR (50)	0.42	0.42	0.39	0.59	–	0.42
6. SR (25)	0.72	0.47	0.48	1.00	0.59	

Numerical similarities are in the upper diagonal, and conceptual similarities are located in the lower diagonal

always be represented as connected is remote. For example, in the technology implementation problem case, there might be a list of influential subsets that are all believed to be the major cause of the failure but have no direct connection to one another. A novice student with insufficient abstract knowledge might identify low teacher motivation as a key factor but fail to elaborate or connect it to other factors, such as organizational support and professional development. In that case, the subset (i.e., a group of concepts centering on low teacher motivation) will necessarily be isolated from the main body of concepts. Any concept map technology needs to detect that subset so that the concept map can be used for identifying problematic understanding and supporting personalized instructional support.

In contrast, the reference model drawn from CMM was less integrated than the concept map for Expert 6 (see Fig. 5). This result implies that concept selection from LSA in CMM might fail to detect some meaningful information about concepts and relations. As seen in Figs. 6 and 7, T-MITOCAR and ALA-Reader allowed all elements of the concept maps to connect, suggesting that all the concepts are actually linked in the mind of the writer. Thus, the concept maps for Expert 6 drawn from T-MITOCAR and ALA-Reader still look completely connected, but they were deemed to be, in some cases, insufficient for obtaining appropriate information about the writer's cognitive status. Drawn from T-MITOCAR, the concept map for Expert 6 was more complex than the reference model (see Fig. 6), while the same from ALA-Reader was too abstract compared to the other approaches (see Fig. 7).
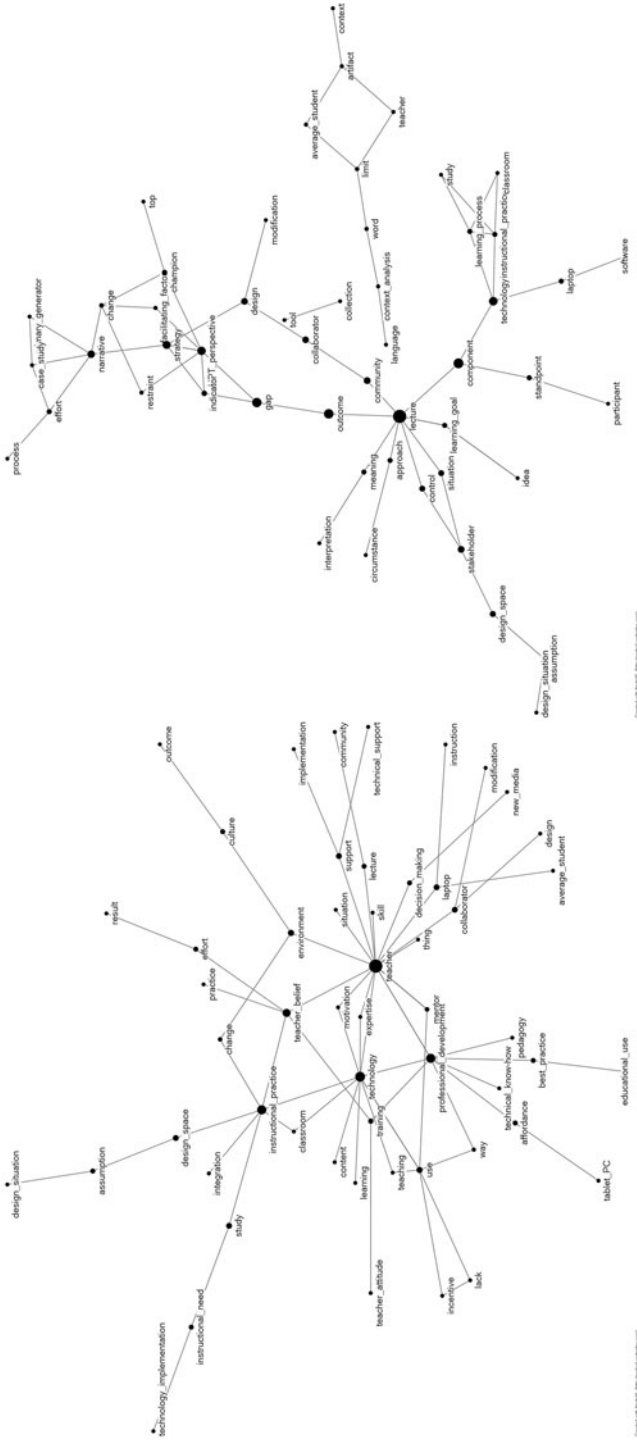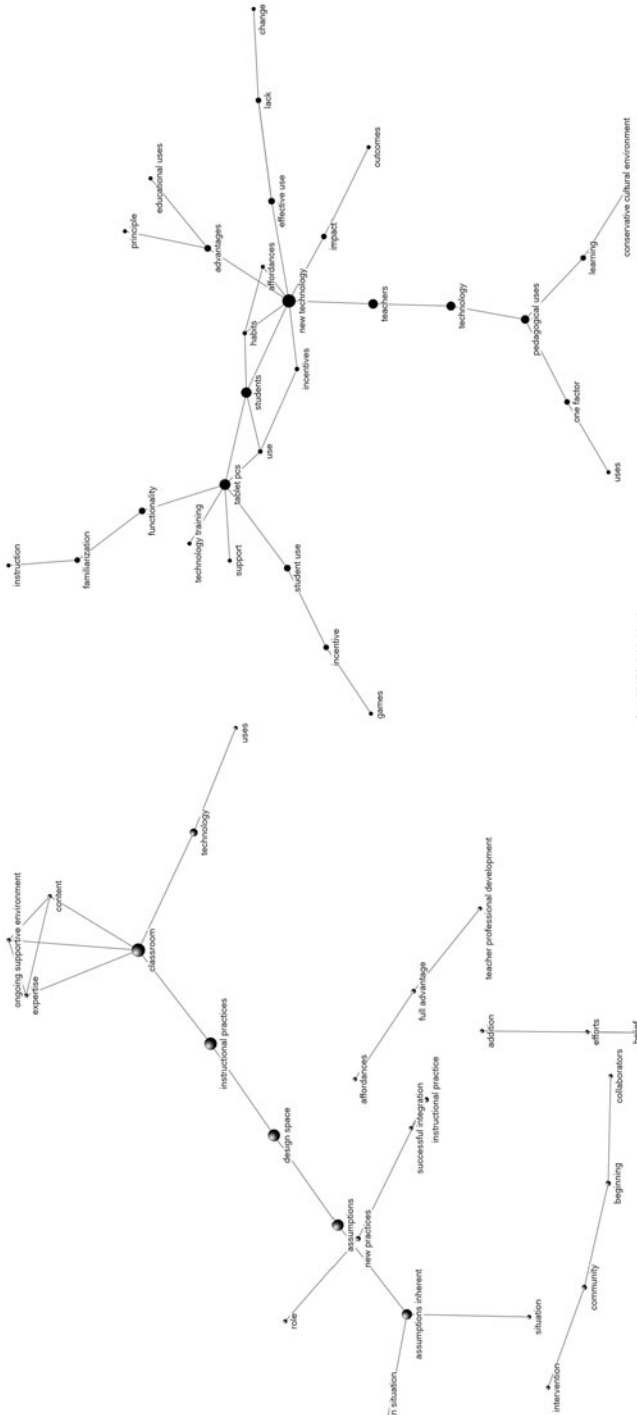
Fig. 4 Concept maps of the reference model (the *left*) and Expert 6 (the *right*) drawn from SR

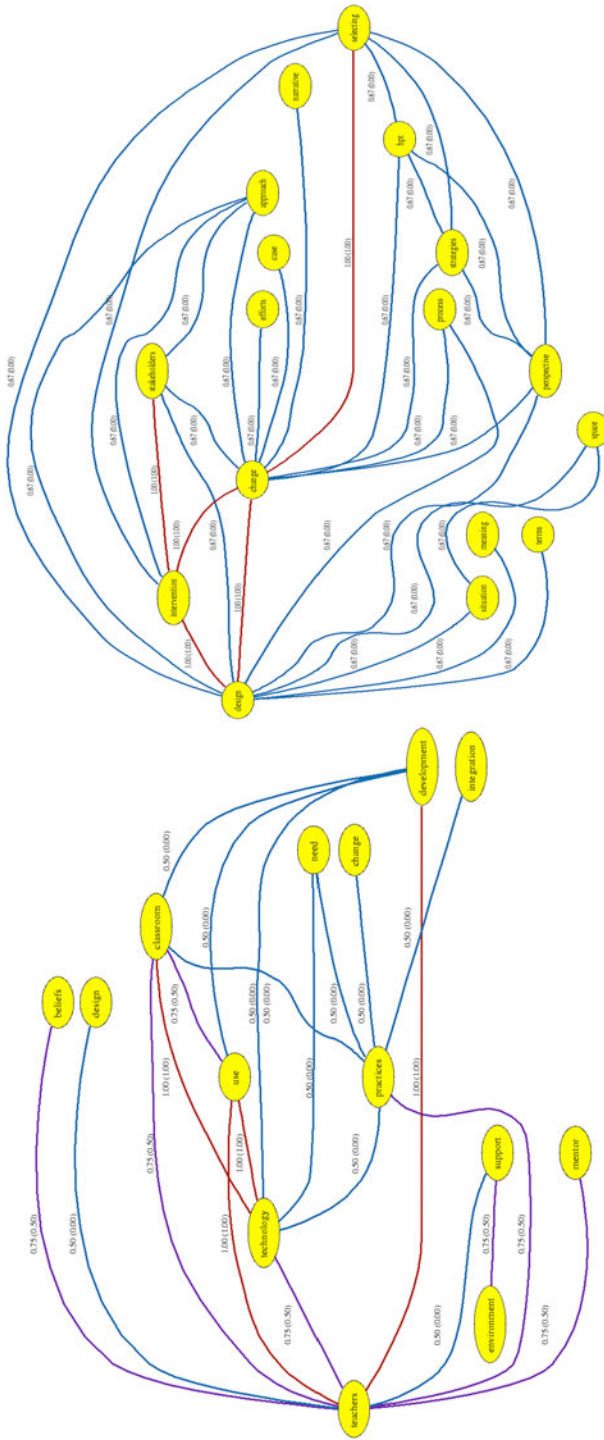**Fig. 5** Concept maps of the reference model and Expert 6 drawn from CMM (PR)

Fig. 6 Concept maps of the reference model and Expert 6 drawn from T-MITOCAR (PR)
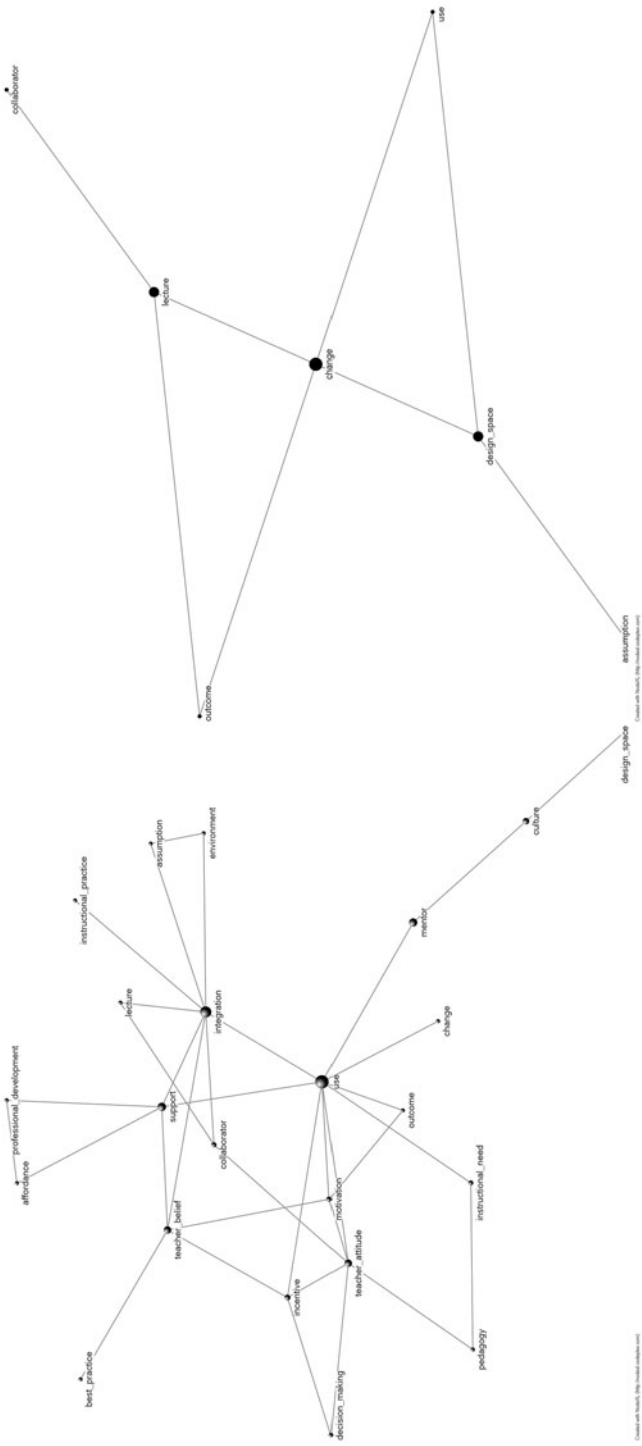
Fig. 7 Concept maps of the reference model and Expert 6 drawn from ALA-Reader (AR)

Discussion

Findings

This study proposed the SR approach for drawing rich and authentic concept maps of written responses that reflect students' internal representations of a problem situation. On the basis of underlying theoretical connections between cognitive and linguistic representation, this study argues that internal semantic structure can be inferred from external linguistic semantic structure. Semantic structure was assumed to be a better basis for representing an individual's mind in the visual form of a concept map. The results demonstrated that semantic relations distilled from a corpus constitute a concept map that is closer to linguistic representation than technologies that use PR or AR approaches.

In accordance with the belief that deep structure is constituted by surface structure (Bransford et al. 1972; Katz and Postal 1964; Spector and Koszalka 2004), using graph-related metrics to filter key concepts as elements of deep structure proved that SR elicited key concepts from the composition of semantic relations (i.e., microstructure). The combination of metrics with the mean threshold (50 % quartile TVoted), as Zouaq et al. (2011b) suggested, was too restrictive. Rather, the 25 % quartile threshold and Betweenness measures produced a set of key concepts closely similar to the set human experts selected (i.e., Standards). Assuming that a single written document with 350–400 words is a common student response in an educational assessment, the thresholds set for the "25 % quartile TVoted" and "Betweenness" could be an effective and efficient way mathematically to identify key meanings in a written response.

Visual inspection of the concept maps revealed the sensitivity of each approach to the assessment context and the writing style. SR was more robust and capable of distinguishing a better response from less qualified responses than the PR and AR tools, but some situations might favor the other approaches. For example, when key concepts are explicitly defined in conjunction with learning goals and the goal of instruction is to help students correctly internalize them along with their prior knowledge, ALA-Reader (AR) might yield more consistent and accurate concept maps on the condition that the concepts are introduced and sufficiently explained.

The SR approach is likely to be an effective approach when the goal of the concept map is to obtain more meaningful (and thus formative) information about students' cognitive changes. However, this study does not intend to argue that the proposed SR approach is always superior to other methods and technologies. PR and AR approaches, represented by CMM (PR), T-MITOCAR (PR), and ALA-Reader (AR), could be useful in providing information about cognitive status, succinctly and economically focusing on key concepts and relations.

Enhancing the design of an adaptive learning environment

Learning Analytics, an emergent field of research, is defined as the mining of student-related data to improve pedagogy (Horizon Report 2013). A potential benefit of this approach is to inform the design of an adaptive learning environment in terms of automatic and simultaneous understanding of student progress in problem solving and adapting instruction to individual learning needs. In this respect, this study provides a set of mechanisms to assess student cognition in problem solving. The following examples depict potential applications.

The proposed mechanisms for eliciting a concept map and filtering out key concepts can work for automatic formative assessment. Given a problem related to global warming in an

eighth-grade science class, a student could compose his/her initial written response (350–400 words) in a computer-based learning system. The system could then analyze the student's text and transform it into a concept map, which an SR-based assessment technology could use to identify the student's key ideas. The student could then review his/her own cognition in the form of a concept map and could ask the system whether there are key concepts that are missing or wrong. The student could keep rebuilding his/her own understanding supported by the system, as the teaching agent system stores changes in the concept maps to monitor the student's progress.

The same techniques can be employed to model expert responses and domain knowledge structure by which learning materials are indexed. Building domain knowledge is critical for intelligent tutoring systems (ITSs) but demands time-consuming and high costs (Zouaq and Nkambou 2008, 2009, 2010). The approaches in this study could address these issues. For example, in a computer-based problem-solving learning environment, a group of experts could first create a set of problems and then submit their own written responses that describe the causes of the problems and associated factors. The ITSs could automatically process those inputs as expert models from which considerable variables are extracted and established as a domain knowledge structure. Associated learning resources such as documents, cases, media files, web sites, and teaching materials could be organized using a distilled knowledge structure so that a student is guided to appropriate learning resources based on his/her diagnostic results, obtained by comparing the student response to expert models. An agent in a system would work as a virtual facilitator to guide and help learners evolve toward the expert level, linking learner models, expert models, and the tutor model.

Improving the scientific accuracy of an educational research

Although the SR approach needs further elaboration and development as an automatic assessment technology, it opens new opportunities in educational research. First, applying SR to the measurement of problem-solving performance can boost the reliability of an experimental study. For example, as performance measures, concept map technologies have often been used to investigate the impact of a treatment in complex problem solving (Kim 2008; McKeown 2009; Schlomske and Pirnay-Dummer 2008). McKeown (2009) asked participants to draw their concept map directly, while T-MITOCAR was used in two other studies to elicit concept maps. Drawing concept maps has limitations because the activity leads to cognitive overload in the working memory, requiring high levels of knowledge abstraction to identify certain concepts and relations (Brown 1992). The effective use of the PR and AR tools is typically dependent on the research context. In contrast, a tool based on SR can provide an alternative way to enhance the scientific accuracy of a comparison study. Another implication is that the SR approach embedded in automated technologies has the potential to cut across disciplinary boundaries (e.g., traditional language comprehension studies). In learning and instruction, SR is applicable to a wide range of areas: automated essay evaluation, expertise modeling, competency diagnosis in adult learning, technology-enhanced adaptive learning systems (e.g., intelligent tutoring system), longitudinal studies of learning progress, and formative assessment and feedback. For instance, SR technology enables researchers to keep track of structural changes in individuals' concept maps over time so that longitudinal changes are measured and described. Another example is automated expertise modeling of a complex problem task. We can postulate a situation and process expert responses to a certain problem through SR, and then share expert understanding established by drawing on the concepts and relations across the responses.

Suggestions

Complex sentences are not easily interpreted using a concept map because correctly distilling paired concepts is more difficult. Future studies might address the following topics:

(1) Elaborate the algorithm to identify semantic relations from a text. This study proposed a set of algorithms and methods to deconstruct a student's written response based on basic principles of computational linguistics. Admittedly, additional ways of identifying logical relations in a text require development.

(2) Build diverse measures that capture the attributes of the knowledge structure. When an individual's understanding is visually represented in the form of a concept map, the characteristics need to be described. Studies have defined some parameters of a concept map that quantify its features (Ifenthaler 2006; Kim 2012a; Spector and Koszalka 2004). Nonetheless, additional parameters need to be identified and validated due to limited evidence and lack of consensus regarding current measures (Kim 2012b).

(3) Elaborate the methodology to compare concept maps to the reference model(s) and to monitor structural changes as learning trajectories. A simple way to conduct concept map comparison is to compare two matrices. For example, Schwartz et al. (2009) compared student-generated and expert-generated matrices to find missing concepts and relations. Furthermore, drawing on similarity formulas, the alignment of a student's concept model with an expert model was used as a learning achievement indicator (Clariana and Taricani 2010; Kim 2008; McKeown 2009; Schlomske and Pirnay-Dummer 2008). However, as pointed out previously, some concept map parameters describe different features of a knowledge structure and change patterns over time (Kim 2012b). Thus, a comprehensive methodology for explaining and monitoring changes in concept maps as they approach an expert model is highly desired.

Appendix A

Case study

Directions: read the case study described below and then prepare a response to the questions below

Assume that you have been involved in evaluating a media implementation project in an urban inner middle school. At the beginning of the school year all of the students assigned to four subject area teachers (math, language arts, social studies and science) in the seventh grade at the middle school were given tablet PCs (laptop computers also equipped with a stylus/pen and a touchscreen that can be written upon) and were also given wireless internet access at home and in school for a entire year.

The students took the tablet PCs home every evening and brought them into classes every day. The teachers were also provided with tablet PCs 24/7 (24 h a day, every day of the week) for the entire year. The teachers and students were trained on how to use the tablet PCs. Moreover, all of the curriculum materials (textbooks, workbooks, student study guides, teacher curriculum guides, some activities, tests, etc.) were installed on the tablet PCs or were accessible through the tablet PCs.

Your job as one of the evaluators for the project was to examine how this innovation (providing teachers and students with tablet PCs 24/7) changed the way instruction was presented in the classrooms of the four teachers. Results indicated that the innovation had very little effect on the manner in which instruction took place in the teachers' classrooms.

Written response with at least 350 words is required for each question

1. Based on what you have learned about the use of technology in education, describe what concepts, issues, factors, and variables are likely to have contributed to the fact that the introduction of the tablet PCs had very little effect on the instructional practices that were employed in the classes.
2. Describe the strategies that could have been employed to help mitigate the factors that you think contributed to the minimal effect the tablet PCs had on instructional practices. When you answering this question, use the concepts, factors, and variables you described in the question 1 or add other assumptions and information that would be required to solve this problem.

Appendix B

Rules for determining pairs of concepts from complex lexico-syntactic patterns

Complex lexico-syntactic patterns	Pairs of concepts–$R(C_i, C_j)$
N0 is N1 and N2 (e.g., technology is hardware and software)	(N0, N1); (N0, N2); (N1, N2)
N0 of N1 and N2 verb N3 (e.g., the use of technology and access to internet allow students to…)	(N0, N1); (N0, N2); (N0, N3); (N2, N3)
N0 such as N1, N2,…, Nn (e.g., classroom technologies such as laptops, internet, and electronic whiteboard)	(N0, N1); (N0, N2);…; (N0, Nn)
Such N0 as N1, N2,…, Nn (e.g., such new technologies as Web 2.0, cloud computing, and mobile internet)	(N0, N1); (N0, N2);…; (N0, Nn)
Np are N1, N2,…, Nn or other N0 (e.g., magnetism is positive or negative)	(Np, N1); (Np, N2);…; (Np, Nn); or (Np, N0)
N0 include N2 and N3 (e.g., internal representation includes conceptual structure and linguistic semantic structure)	(N0, N1); (N0, N2)
N0, especially N1, verb… (e.g., supportive environments, especially leadership support, are the most important)	(N0, N1)
N1 of N2 in N3 of N4 (e.g., the use of technology in the classrooms of participating schools)	(N1, N2); (N3, N4); (N1, N3)
By –ing N1 and N2, Np verb N3[a] (e.g., by using the Internet and Smartphone, students can access learning materials anytime, anywhere)	(Np, N1); (Np, N2); (Np, N3)
N1 provide N2 with N3 (e.g., the Internet provides us with)	(N1, N2); (N1, N3)
N1 and N2 verb N3 (e.g., teachers and students are not used to using a computer)	(N1, N2); (N1, N3); (N2, N3)
N1 verb that-clause (e.g., the witness hated that the boy attacked the victim)	(N1, the first N in that-clause)[c]
N1 between N2 and N3 (e.g., discrepancy between boys and girls)	(N1, N2); (N1, N3)
N1 that N2 verb N3[b] (e.g., teachers maintain the belief that these efforts will have positive results)	(N1, N2); (N2, N3)

[a] Subordinate clause in which the subject is omitted

[b] Conjunction clause

[c] It connects N1 to the that-clause

References

Adriana, D. M., Moldovan, D., Badulescu, A., Tatu, M., Antohe, D., & Girju, R. (2004). Models for the semantic classification of noun phrases. *In HLT-NAACL 2004: Workshop on Computational Lexical Semantics*, 60-67.

Anthonisse, J. M. (1971). *The rush in a graph*. Amsterdam: Mathematische Centrum.

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

Axelrod, R. (1976). *Structure of decision: The cognitive maps of political elites*. Princeton: Princeton University Press.

Azevedo, R. (2011). Can we design artificial pedagogical agents to be intelligent enough to detect, model, and foster regulatory learning processes? In R. Pirrone & F. Sorbello (Eds.), *AI*IA 2011: Artificial intelligence around man and beyond* (Vol. 6934, p. 9). Berlin: Springer.

Beamer, B., Rozovskaya, A.,& Girju, R. (2008) Automatic semantic relation extraction with multiple boundary generation. *Proceedings of AAAI* (824–829), Chicago: AAAI Press.

Bierwisch, M., & Schreuder, R. (1992). From concepts to lexical items. *Cognition, 42*(1–3), 23–60. doi:16/0010-0277(92)90039-K.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3*, 193–209.

Bransford, J. D., & Franks, J. J. (1972). The abstraction of linguistic ideas. *Cognitive Psychology, 2*, 331–350.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 717–726. doi:16/S0022-5371(72)80006-9.

Brown, D. (1992). *An introduction to object-oriented analysis: Objects in plain English*. New York: Wiley.

Cañas, A. J. (2009). *What are Propositions?… from a Concept Mapping Perspective*. http://cmap.ihmc.us/docs/Proposition.html. Accessed 31 Jul 2011.

Champagne, A. B., Kouba, V. L., & Hurley, M. (2000). Assessing inquiry. In J. Minstrell & E. H. van Zee (Eds.), *Inquiry into inquiry learning and teaching in science* (pp. 456–461). Washington DC: American Association for the Advanced of Science.

Clariana, R. B. (2010). Multi-decision approaches for eliciting knowledge structure. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 41–59). New York: Springer.

Clariana, R. B., & Koul, P. (2008). The effects of learner prior knowledge when creating concept maps from a text passage. *International Journal of Instructional Media, 35*(2), 229–236.

Clariana, R. B., & Taricani, E. M. (2010). The consequences of increasing the number of terms used to score open-ended concept maps. *International Journal of Instructional Media, 37*(2), 218–226.

Clariana, R., Wallace, P., & Godshalk, V. (2009). Deriving and measuring group knowledge structure from essays: The effects of anaphoric reference. *Educational Technology Research and Development, 57*(6), 725–737. doi:10.1007/s11423-009-9115-z.

Collins, A., & Gentner, D. (1987). How people construct mental models. In D. Holland & N. Quinn (Eds.), *Cultural models in language and thought* (pp. 243–265). Cambridge: Cambridge University Press.

Curti, M. B., & Viator, R. E. (2000). An investigation of multidimensional knowledge structure and computer auditor performance. *Auditing: A Journal of Practice and Theory, 19*(Fall), 83–103.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science (1986-1998), 41*(6), 391-407.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: a meta-analysis. *Learning and Instruction, 13*(5), 533–568. doi:10.1016/S0959-4752(02)00025-7.

Downing, P. (1977). On the creation and use of English compound nouns. *Language, 53*(4), 810–842.

Duschl, R. (2003). Assessment of inquiry. In J. M. Atkins & J. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 41–60). Arlington, VA: NSTA Press.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.

Frederick, D. M. (1991). Auditors' representation and retrieval of internal control knowledge. *The Accounting Review, 66*, 240–258.

Frederick, D. M., Heiman-Hoffmann, V. B., & Libby, R. (1994). The structure of auditors' knowledge of financial statement errors. *Auditing: A Journal of Practice and Theory, 13*(Spring), 1–21.

Freeman, L. C. (1977). A set of measures of centrality based on betweeness. *Sociometry, 40*, 35–41.

Funke, J. (1985). Steuerung dynamischer Systeme durch Aufbau und Anwendung subjektiver Kausalmodelle [Control of dynamic systems by building up and using subjective causal models]. *Zeitschrift für Psychologie, 193*, 443–466.

Garnham, A. (1987). *Mental models as representations of discourse and text*. Chicheser: Ellis Horwood Ltd.

Garnham, A. (2001). *Mental models and the interpretation of anaphora*. Hove: Psychology Press.

Geeslin, W. E., & Shavelson, R. J. (1975). Comparison of content and cognitive structure in high school students' learning of probability. *Journal of Research in Mathematics Education, 6*, 109–120.

Girju, R. (2008, August). *Semantic relation extraction and its applications*. Paper presented at 20th European Summer School in Logic, Language and Information, Freie und Hansestadt Hamburg, Germany.

Girju, R. (2011). The syntax and semantics of prepositions in the task of automatic interpretation of nominal phrases and compounds: A cross-linguistic study. *Computational Linguistics, 35*(2), 185–228. doi:10.1162/coli.06-77.

Girju, R., Beamer, B., Rozovskaya, A., Fister, A., & Bhat, S. (2010). A knowledge-rich approach to identifying semantic relations between nominals. *Information Processing and Management, 46*(5), 589–610. doi:10.1016/j.ipm.2009.09.002.

Girju, R., Nakov, P., Nastase, V., Szpakowicz, S., Turney, P., & Yuret, D. (2009). Classification of semantic relations between nominals. *Language Resources and Evaluation, 43*(2), 105–121. doi:10.1007/s10579-009-9083-2.

Goodman, C. M. (1987). The Delphi technique: A critique. *Journal of Advanced Nursing, 12*(6), 729–734.

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

Hearst, M. (1992). Automatic acquisition of hyponyms from large text corpora. *Proceedings on 14th international conference on computational linguistics* (COLING-92) (539–545). Stroudsburg: Association for Computational Linguistics.

Hsu, Chia-Chien & Sandford, Brian A. (2007). The Delphi Technique: Making Sense of Consensus. *Practical Assessment Research & Evaluation, 12* (10): 1–8. Available online: http://pareonline.net/getvn.asp?v=12&n=10.

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

Ifenthaler, D., Masduki, I., & Seel, N. M. (2009). The mystery of cognitive structure and how we can detect it: Tracking the development of cognitive structures over time. *Instructional Science, 39*(1), 41–61.

Janssen, T. M. V. (2012). Montague semantics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2012.). Retrieved from http://plato.stanford.edu/archives/win2012/entries/montague-semantics/.

Johnson-Laird, P. N. (2005a). Mental models and thoughts. In K. J. Holyoak (Ed.), *The Cambridge handbook of thinking and reasoning* (pp. 185–208). Cambridge: Cambridge University Press.

Johnson-Laird, P. N. (2005b). The history of mental models. In K. I. Manktelow & M. C. Chung (Eds.), *Psychology of reasoning: theoretical and historical perspectives* (pp. 179–212). New York: Psychology Press.

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

Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language* (pp. 277–322). Amsterdam: Mathematisch Centrum.

Katz, J. J., & Postal, P. M. (1964). *An integrated theory of linguistic descriptions*. Cambridge: M.I.T. Press.

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

Kim, M. (2012a). Theoretically grounded guidelines for assessing learning progress: Cognitive changes in ill-structured complex problem-solving contexts. *Educational Technology Research and Development, 60*(4), 601–622. doi:10.1007/s11423-012-9247-4.

Kim, M. (2012b). Cross-validation study on methods and technologies to assess mental models in a complex problem solving situation. *Computers in Human Behavior, 28*(2), 703–717.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*(4), 294–303. doi:10.1037/0003-066X.49.4.294.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363–394. doi:10.1037/0033-295X.85.5.363.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes, 25*, 259–284.

Lauer, M. (1995).Corpus statistics meet the noun compound: Some empirical results. *Proceedings of the Association for Computational Linguistics Conference* (ACL) (47–54), Cambridge, MA.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.

Levi, J. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.

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

Moldovan, D. I., & Girju, R. C. (2001). An interactive tool for the rapid development of knowledge bases. *International Journal on Artificial Intelligence Tools, 10*(1/2), 65–86.

Montague, R. (1974). In R. Thomason (Ed.), *Formal Philosophy: The Selected Papers of Richard Montague*. New Haven: Yale University Press.

Murphy, M. L. (2003). *Semantic relations and the lexicon: antonymy, synonymy and other paradigms*. Cambridge: University Press.

Narayanan, V. K. (2005). Causal mapping: An historical overview. In V. K. Narayanan & D. J. Armstrong (Eds.), *Causal mapping for research in information technology* (pp. 1–19). Hershey: Idea Group Publishing.

Naveh-Benjamin, M., McKeachie, W. J., Lin, Y., & Tucker, D. (1986). Inferring students' cognitive structures and their development using the ordered tree technique. *Journal of Educational Psychology, 78*, 130–140.

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

Norman, D. A. (1986). Reflections on cognition and parallel distributed processing. In J. L. McClelland, D. E. Rumelhart, & The POP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 531–546). Cambridge: MIT Press.

Novak, J. D., & Cañas, A. J. (2006). *The origins of the concept mapping tool and the continuing evolution of the tool*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.106.3382.

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & Management, 42*(1), 15–29. doi:10.1016/j.im.2003.11.002.

Partee, B. H. (1984). Compositionality. In F. Landman & F. Veltman (eds.), *Varieties of formal semantics: Proceedings of the 4th Amsterdam colloquium* (Groningen-Amsterdam Studies in Semantics, No. 3) (pp. 281–311). Dordrecht: Foris.

Phelan, J., Kang, T., Niemi, D. N., Vendlinski, T., & Choi, K. (2009). *Some aspects of the technical quality of formative assessments in middle school mathematics*. (CRESST Report 750). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

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*. New York: Springer.

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

Quellmalz, E. S., & Haertel, G. D. (2004). *Use of technology-supported tools for large-scale science assessment: implications for assessment practice and policy at the state level*. Commissioned paper prepared for the National Research Council's Committee on Test Design for K-12 Science Achievement, Washington, DC.

Rijkhoff, J. (2002). *The noun phrase*. Oxford: Oxford University Press.

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

Schlomske, N., & Pirnay-Dummer, P. (2008). Model based assessment of learning dependent change during a two semester class. In Kinshuk, Sampson, D., & Spector, M. (Eds.), *Proceedings of IADIS International Conference Cognition and Exploratory Learning in Digital Age 2008* (pp. 45053). Freiburg, Germany: IADIS.

Schvaneveldt, R. W. (Ed.). (1990). *Pathfinder associative networks: Studies in knowledge organization*. Ablex Publishing Corp. Retrieved from http://portal.acm.org/citation.cfm?id=SERIES9012.119801.

Schvanevldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. In G. H. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory* (pp. 249–284). San Diego: Academic Press.

Schwartz, D. L., Chase, C., Chin, D. B., Oppezzo, M., Kwong, H., Okita, S., et al. (2009). Interactive metacognition: Monitoring and regulating a teachable agent. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 340–358). New York: Routledge.

Seel, N. M. (1999). Semiotics and structural learning theory. *Journal of Structural Learning and Intelligent Systems, 14*(1), 11–28.

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

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

Seel, N. M. (2004). Model-centered learning environments: Theory, Instructional design, and ethics. In N. M. Seel & S. Dijkstra (Eds.), *Curriculum, plans, and processes in instruction design: International perspectives* (pp. 49–74). Mahwah: Lawrence Erlbaum Associates Inc.

Seel, N. M., & Dinter, F. R. (1995). Instruction and mental model progression: Learner-dependent effects of teaching strategies on knowledge acquisition and analogical transfer. *Educational Research and Evaluation, 1*(1), 4–35.

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*. New York: The Greenwood Publishing Group.

Shute, V. J., & Zapata-Rivera, D. (2007). Adaptive technologies. In J. M. Spector, M. D. Merill, J. van Merrienboer, & M. P. Driscoll (Eds.), *Handbook of research for educational communications and technology* (pp. 227–294). New York: Routledge: Taylor & Francis Group.

Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115–163.

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., & 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.

Taricani, E. M., & Clariana, R. B. (2006). A technique for automatically scoring open-ended concept maps. *Educational Technology Research and Development, 54*(1), 65–82.

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

Villalon, J., & Calvo, R. A. (2009, December). *Single document semantic spaces*. Paper presented at the Australian Joint Conference on Artificial Intelligence, Melbourne, Australia.

Villalon, J., & Calvo, R. A. (2011). Concept Maps as Cognitive Visualizations of Writing Assignments. *Educational Technology & Society, 14*(3), 16–27.

Villalon, J., Calvo, R. A., & Montenegro, R. (2010). Analysis of a gold standard for Concept Map Mining – How humans summarize text using concept maps. *Proceedings of the Fourth International Conference on Concept Mapping* (pp. 14–22). Viña del Mar, Chile.

Wasserman, S., & Faust, K. (1994). *Social networks analysis: Methods and applications*. Cambridge: Cambridge University.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45*(4), 477–501.

Zouaq, A., Gasevic, D., & Hatala, M. (2011a). Towards open ontology learning and filtering. *Information Systems, 36*(7), 1064–1081. doi:10.1016/j.is.2011.03.005.

Zouaq, A., Gasevic, D., & Hatala, M. (2011b). Ontologizing concept maps using graph theory. *Proceedings of the 2011 ACM Symposium on applied computing* (pp. 1687–1692). New York, NY, USA: ACM. doi:10.1145/1982185.1982537.

Zouaq, A., Michel, G., & Ozell, B. (2010). Semantic analysis using dependency-based grammars and upper-level ontologies. *International Journal of Computational Linguistics and Applications, 1*(1–2), 85–101.

Zouaq, A., & Nkambou, R. (2008). Building domain ontologies from text for educational purposes. *IEEE Transactions on Learning Technologies, 1*(1), 49–62. doi:10.1109/TLT.2008.12.

Zouaq, A., & Nkambou, R. (2009). Evaluating the generation of domain ontologies in the knowledge puzzle project. *IEEE Transactions on Knowledge and Data Engineering, 21*(11), 1559–1572. doi:10.1109/TKDE.2009.25.

Zouaq, A., & Nkambou, R. (2010). A survey of domain ontology engineering: methods and tools. In R. Nkambou, J. Bourdeau, & R. Mizoguchi (Eds.), *Advances in intelligent tutoring systems* (pp. 103–119). Springer Berlin Heidelberg. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-14363-2_6.

Minkyu Kim received his PhD from the Learning, Design, and Technology program at the University of Georgia in May 2012, and he has been working as a postdoctoral research associate in the Center for Scholarly Technology at the University of Southern California since August 1, 2012. He has over 8-year work experience as a e-learning consultant and instructional designer. His work experience employed

technology in a variety of learning situations ranging from K-12 to adult learners in both online and face-to-face settings. Combined with his work experience, his active research agenda centers on personalized instructional supports and feedback in online and hybrid learning environments, focusing on the studies in terms of learner modeling, assessment of cognitive changes in complex problem solving situations, and the development of assessment technologies.