Education Tech Research Dev (2021) 69:971–1002
https://doi.org/10.1007/s11423-021-09942-1

DEVELOPMENT ARTICLE

Using graph centrality as a global index to assess students' mental model structure development during summary writing

Min Kyu Kim[1] · Kathryn S. McCarthy[2]

Accepted: 13 January 2021 / Published online: 28 January 2021

Abstract
During reading, students construct mental models of what they read. Summaries can be used to evaluate the latent knowledge structure of these mental models. We used indices from Student Mental Model Analyzer for Research and Teaching (SMART) to explore the potential of a global index, Graph Centrality (GC), as a measure to describe mental model structure and its relation to the quality of student summaries (e.g., the amount of content-coverage). Students ($n=73$) in an online graduate-level course wrote and revised summaries of their course readings. Data preview left the total count of 32 cases to evaluate how students' mental representations changed from initial to final version. These summaries were analyzed using indices derived from the 3S model (surface, structure, semantic) as well as a measure of GC. The results of this initial investigation are promising, demonstrating that Graph Centrality captures important differences in students' summaries, including revision behaviors to the wholistic structure of mental models, modification trajectories toward a cohesive and solid mental representation that is semantically similar to the expert model.

Keywords Summary writing · Mental models · Formative assessment · Knowledge structure · Reading comprehension

✉ Min Kyu Kim
mkim120@gsu.edu

Kathryn S. McCarthy
kmccarthy12@gsu.edu

[1] Department of Learning Sciences, Georgia State University, 238 College of Education and Human Development, 30 Pryor St SW, Atlanta, GA 30303, USA

[2] Department of Learning Sciences, College of Education & Human Development, Georgia State University, 776 College of Education and Human Development, 30 Pryor St SW, Atlanta, GA 30303, USA

Introduction

Summarization is a common and effective classroom practice (Dunlosky et al. 2013; Singer and Alexander 2017; Stevens et al. 2019). In order to write a summary, students must examine information throughout the text, separate important from less important ideas, synthesize these important or main ideas according to their propositional relationships, and create a new abstract knowledge of the whole text (Duke and Pearson 2009). Writing a quality summary requires students to utilize higher-order thinking skills (e.g., analysis, synthesis, and evaluation) and helps develop deep comprehension of the reading materials (Duke and Pearson 2009; Kintsch 1988; Westby et al. 2010). Thus, learning from text involves a collection of diverse cognitive processes (Singer and Alexander 2017). Researchers and educators have analyzed student summaries to examine students' comprehension processes (Kim et al. 2019; He et al. 2009; Kim et al. 2018; Li et al. 2018; Lin 2004; Santos Jr et al. 2004; Sung et al. 2016; Wade-Stein and Kintsch 2004). The purpose of the current study is to investigate how a computer-generated index, *Graph Centrality*, of a students' written summary can be used to evaluate the quality of their mental model and, in turn, their understanding.

Theories of mental models assume that readers construct mental models of a text as they read (Helbig 2006; Johnson et al. 2011; Jonassen and Cho 2008; Pirnay-Dummer and Ifenthaler 2011). Readers connect important ideas from different parts of the text as well as integrates information from prior knowledge to construct a coherent and elaborated mental model (Graesser et al. 1994; Johnson-Laird 2005). A written response, such as a summary, can be thought of as a re-representation of a students' internal mental model of the text (Jonassen et al. 1993; Kintsch 1988). One limitation in using summaries as a means of learning and formative assessment is that the process of evaluating open-ended responses is time-intensive, complex, and thus demanding to instructors. Due to this challenge, timely formative assessment of students' summaries hardly happens in classrooms (Graham et al. 2013; Li et al. 2018).

However, recent advancements in natural language processing (NLP) have made it feasible to automatically analyze students' written work (Crossley and McNamara 2016; Passonneau et al. 2018; Ifenthaler 2014; Kim et al. 2018; McNamara et al. 2017; Strobl et al. 2019). NLP tools use a variety of statistical techniques to analyze qualities of language at multiple dimensions. For example, the Tool for the Automatic Assessment of Lexical Sophistication (TAALES, Kyle et al. 2018) calculates more than 400 indices that describe the simplicity (or conversely, the complexity) of the words used in the writing. Tools like Coh-Metrix (McNamara et al. 2014) and SEMILAR (Rus et al. 2013) use latent semantic analysis (LSA, Landauer and Dumais 1997) to go beyond word-level metrics to calculate the relations or similarity between ideas in a text.

Of particular interest to the present work, this increase in NLP-driven tools has led to the development of automated summary evaluators (ASEs) that use NLP indices to determine the amount of content-coverage from the source text as well as the overall writing quality of student summaries (Strobl et al. 2019; Sung et al. 2016). These ASEs, such as Summary Street (Wade-Stein and Kintsch 2004), Online Summary Assessment and Feedback System (Sung et al. 2016), crowd-source summary evaluation (Li et al. 2018), ROUGE (Lin 2004), and PryEval (Gao et al. 2019), provide actionable feedback that can help students to improve their summary writing skills. Notably, however, these ASE tools use a series of descriptive linguistic indices, and the focus of the tools is to help develop their general summary writing skills, as opposed to an evaluation of the reader's mental model and their deep comprehension of the source text content.

One means of providing a more comprehensive evaluation of the student's summary is a model-based approach. A model-based approach elicits a concept map from a summary. A concept map is a network of interrelated concepts and is a re-representation of the students' knowledge structure embedded in the summary (Allen et al. 2015; Axelrod 1976; Ifenthaler 2014; Kim 2018; Koszalka and Epling 2010; Narayanan 2005; Schvaneveldt and Cohen 2010; Spector 2010). Model-based tools generate indices from elicited concept maps along multiple dimensions of mental models. Some scholars have proposed three dimensions (i.e., surface, structure, and semantic; Kim 2012; Ifenthaler 2014; Pirnay-Dummer and Ifenthaler 2011; Spector and Koszalka 2004), while some have suggested two dimensions, including surface and deep structure (Bransford and Johnson 1972; Gentner and Medina 1998; Katz and Postal 1964; Kintsch and van Dijk 1978). Model-based ASEs includes AKOVIA (Ifenthaler 2014), GISK (Kim 2018), HIMATT (Pirnay-Dummer and Ifenthaler 2011), and SMART (Kim et al. 2019). Using these tools, previous studies have shown that these dimensions capture meaningful change in student's summary revisions (Kim 2015; Clariana 2010; Ifenthaler et al. 2011; Kim 2018; Pirnay-Dummer and Ifenthaler 2011).

Despite the prominent findings, previous work tends to be limited to some indices and shows that there were inconsistent relationships and various non-linear patterns among indices in different dimensions (Ifenthaler et al. 2014). For example, diverse patterns of mean scores of indices indicated a greater variation within and across measures (Ifenthaler et al. 2014). Also, a previous study (Kim and McCarthy 2020) demonstrated that a greater inconsistency existed in structure-related indices. The effect of revision was less pronounced in the structural dimension. The literature calls for further investigation on mental model change in terms of structure in order to better evaluate students' learning.

The purpose of the current study is to advance a model-based approach to summary evaluation. Thus, we deploy an additional index, Graph Centrality (GC), that indicates the extent of relations among concepts in a concept map (Clariana et al. 2011; Newman 2010). Specifically, we explore the potential of GC as a global index that can characterize overall changes in a student's knowledge structure across individual indices in multiple dimensions. We examine how GC changes across student revisions and how GC compares across student summaries and expert benchmark summaries.

Theoretical background

Mental models in reading comprehension

Mental models include the ideas and concepts explicit in the text as well as the connections between those ideas and concepts. Thus, assessing comprehension includes evaluating not only the amount of knowledge, but also the organization of that knowledge within the mental model (Bransford et al. 2000; Kintsch 1998; Jonassen et al. 1993; Segers 1997).

Theories of mental models explain that people develop expertise through the manipulation of their cognitive artifacts that represent certain aspects of a problem situation (e.g., a complex reading material) (Anzai and Yokoyama 1984; Collins and Gentner 1987; Johnson-Laird 2005; Seel 2004; Smith et al. 1993). This line of work is grounded in the mental model theories with three assumptions: (a) readers construct a mental representation, or mental model, of the text as they read (Graesser et al. 1994; Johnson-Laird 2005; Kintsch 1998); (b) people's language, in terms of what people say or write as well as how they convey this information, is an external representation of the reader's internal mental model (Garnham 1987, 2001; Greeno 1989); and (c) one can evaluate learner comprehension by examining multi-layered, multi-dimensional mental representations demonstrated in people's language (Kim et al. 2019; Clariana and Taricani 2010; Gijbels et al. 2005; Zimmerman et al. 2018). More simply put, in the context of reading and summary writing, these assumptions mean that researchers can analyze student summaries to understand what a reader understand about a text that they have read. Importantly, mental models are not static, but rather can change gradually or abruptly (Kim 2015; Gentner et al. 2001; Pirnay-Dummer and Ifenthaler 2011; Spector 2010). Thus, a critical assumption of this work is that change in students' mental models elicited from their summary revisions can indicate their evolving understanding of the text. In turn, drawing the reader's attention to critical ideas and relations in an externally represented expert's mental model can help reader to write cohesive and complete summary of the text (Kim and Clariana 2017; Mayer 1989).

For example, students may write an initial version of a summary that centers on only a few, minor concepts. Given the feedback information about their summary, students may modify their thoughts and rewrite summaries, adding more important ideas and making more connections between these ideas. In contrast, some students who lack pre-existing knowledge of the text may write a summary with as many concepts as possible due to their lack of focus, which may make their summary complex and less structured. Feedback on the main ideas of the text may help those students focus on more important concepts and revise their summary more concise and cohesive, dropping unnecessary ideas (Kim 2017, 2018). We can take advantage of concept maps elicited from summaries to describe students' comprehension of the text in terms of the extent to which a student's holistic mental structure covers the substantial parts of the text and the degree of which the structure of the model is as cohesive and solid as an expert model (Kim et al. 2019; Spector and Koszalka 2004).

Model modification processes involve changes in the number of concepts and their propositional relations when students add or remove concepts from their summary during revisions (Norman 1983; Rumelhart and Norman 1978). Although, at its simplest, students are adding or deleting information, the interpretation of dynamic changes in a concept map is complex. For example, one student might revise their summary by haphazardly adding concepts, but not optimize them in terms of concept relationships. This results in a large concept map, but the overall network is not cohesive. A different student may add fewer, but more carefully selected concepts that have close connections with the existing concepts in the model. This concept map would be only slightly larger than the student's original but would be a far more cohesive network. Thus, considering the structure of mental models may be better suited to describe these qualitative differences and changes more than merely calculating the increase or decrease in the numbers of concepts and relations. Although this approach is promising, there is a dearth of empirical evidence related to how these structural changes of mental models reflect the improvements in the quality of the reader's comprehension (Kim et al. 2019; Westby et al. 2010).

Model-based 3S dimensions

Scholars generally agree that knowledge structures are multifaceted or multidimensional (Clariana 2010; Ifenthaler and Pirnay-Dummer 2014; Spector and Koszalka 2004). In this current study, we focus on 3S knowledge structure dimensions: (a) surface, (b) structural, and (c) semantic dimensions. Spector and Koszalka (2004) first introduced the 3S dimensions that have provided a theoretical framework for mental model assessment (Kim 2015; Ifenthaler 2014; Pirnay-Dummer et al. 2010).

The *surface* dimension reflects overall number of words, concepts, and relations in concept maps that build the basic information of mental models (Fodor et al. 1974; Holyoak and Koh 1987; Katz and Postal 1964). It provides the descriptive information of components of a knowledge structure. From a linguistics perspective, the surface dimension in terms of concepts and their relations in text characterize the shape of the sentences (Katz and Postal 1964).

The *structural* dimension characterizes the whole network of mental models in terms of the degree to which information in the text is organized and connected (Bransford and Franks 1972; Bransford and Johnson 1972; Gentner and Medina 1998; Kintsch and van Dijk 1978). The focus of the structural dimension is on the "extent to which the student's knowledge structure is organized around key concepts and principles that are linked to conditions and procedures for application" (Gijbels et al. 2005, p. 35).

The *semantic* dimension focuses on the underlying ideas in the text and, specifically, relates to whether key concepts and relations that students must learn from the text are embedded in their mental models (Kim 2013; Bransford and Franks 1972; Bransford and Johnson 1972; Katz and Postal 1964; Kintsch and van Dijk 1978). Studies of linguistic comprehension explain that a substantial part of the meaning emerges from information integrated from the whole corpus (Bransford and Franks 1972; Bransford and Johnson 1972; Kintsch and van Dijk 1978).

Previous studies demonstrated that the three dimensions explain different aspects of mental models (Kim 2012, 2015; Kim and McCarthy 2020; Clariana 2010; Ifenthaler 2009; Kim 2018; Pirnay-Dummer and Ifenthaler 2011). For example, an empirical study (Kim and McCarthy 2020) showed that the surface and semantic dimensions tended to change in a similar direction, while the indices of the structural dimension appeared to exhibit different patterns within and across dimensions. The findings from this study were generally consistent with the existing theory suggesting that deep comprehension emerges from constructing an interconnected mental model (Kintsch 1998). However, the findings also revealed inconsistent patterns in structure-related indices. Thus, further investigation required an advanced structural index to detect and describe changes in the overall quality of a students' knowledge structure consistent with changes in the surface and semantic dimensions.

A global index of a knowledge structure: graph centrality (GC)

Concept map morphology

Concept map morphology studies indicate that concept maps tend to follow three different types of structure: spoke, chain, and net structure (see Fig. 1). Simple addition of new concepts may create new links and become disruptive to the existing structure. Thus, changes in mental model structure are reflected by how learners switch, change, and link among the three types of structure (Kinchin 2008). For example, Kinchin (2008) notes that the spoke structure often serves as an interim structure from which either chains or nets can arise as additional information is added. These model modification processes may cause

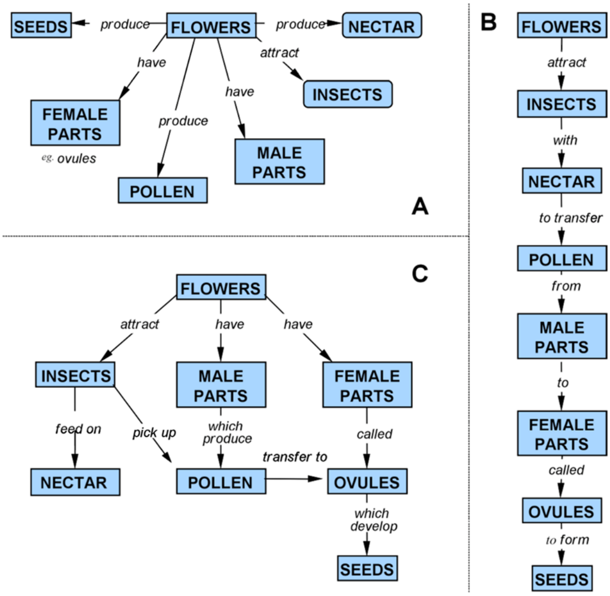


Fig. 1 Morphological variation in concept maps. A = Spoke, B = Chain, C = Network (Kinchin 2008, p. 2)

the gradual or abrupt emergence of different types of structure and increase complexity of the whole network (Hay and Kinchin 2006).

Chain structures on their own often reflect *microstructures*, indicative of a key idea, in one or two sentences (Hay and Kinchin 2006; van Dijk and Kintsch 1983). For example, this concise and well-written definition, “E-learning is a mode of learning that integrates pedagogies empowered by technologies,” shows a chain structure (i.e., [e-learning]-[mode]-[learning]-[pedagogy]-[technology]). Individual chains can be connected into more elaborate structures they share concepts. Connected chain structures form an integrated net structure together, which is called a cohesive macrostructure (Kintsch 1998). Key concepts likely link more subgroups and thus tend to play an important role of the cohesion of a concept map (Kim et al. 2016, 2019).

Graph centrality (GC)

Generally speaking, the concept maps that emerge from complex texts are likely to take on a net structure combining several substructures. However, even within this grouping, there are qualitatively different networks. For example, optimal concept maps include well-defined chain structures connected by carefully selected key concepts that stand toward the center of the network (Kim et al. 2019). On the other hand, a net structure composed of ill-defined chains (connecting many concepts that are not grouped in chunks of proper propositional relations) reflects a naïve epistemology (Hay and Kinchin 2006). A holistic evaluation of the student knowledge structure can be a means to detect, describe, and interpret a dynamic transformation of knowledge structure beyond individual indices and dimensions.

Thus, in the current work, we examine a more holistic measure of a concept map analysis: graph centrality. Graph centrality (GC) as a global index that considers the 3S dimensions in tandem. GC is calculated based on the degree centrality that describes the number of relations connected to a concept in a concept map (Newman 2010). For example, in a sentence of the summary (“Some says motivation is a cognitive domain in terms of a person's belief about the current achievement, while some describe motivation as part of affective domains."), the concept ("motivation") can be connected to "cognitive domain" and "affective domain," which makes the degree centrality of the concept ("motivation") two. GC is the degree centrality of a whole concept map that is computed by incorporating individual concepts' degree centrality values into the entire network level (Clariana et al. 2011; Kim 2017).

Our assumption is that GC can serve as a means to characterize the chain and net structures of a student's concept map. Also, we predict that GC values of ideal knowledge structures of the texts would be dependent on the way to write a reference summary and to elicit concept maps by the technology.

The current study

The current study explores Graph Centrality (GC) as a measure of student mental model construction and revision. Specifically, we examined how changes in the GC (i.e., the centrality of information in a concept map based on a student summary) captured changes in students' summaries of texts from initial draft to final version. The following research questions guided the current study:

RQ1. To what extent do GC values of student summaries change from initial to final version?
RQ2. To what extent do GC measures relate to changes of 3S indices?

To generate GC along with 3S (surface, structure, semantic) indices, we used the Student Mental Model Analyzer for Research and Teaching (SMART) tool (further described in the following section). A descriptive case study approach was used to examine how GC captured structural change of students' mental models. We combined two case studies, both of which used a different group of students and different reading material. Two cases served for the cross-validation of the findings. For each case, we inspected overall shifts in the global index as indicative of knowledge structure change, examined the relationships of the global index with the 3S indices and similarity measures from the SMART tool, and also conducted visual inspections of selected students' mental representations.

Method

Participants

Participants were 73 students enrolled in multiple sections of the same graduate-level online course. Group 1 included 38 students enrolled in two sections during a fall semester. Group 2 consisted of 35 students enrolled in two sections in the following spring semester. Demographic information appears in Table 1. The sample was predominantly female and was composed of mostly students coming from the corporate sector, but participants' professional backgrounds varied, including PK12, higher education, and non-governmental organization (NGO).

The students wrote and revised summaries of their course readings. Data preview showed that 14 students (36%) in Group 1 and 22 students (63%) in Group 2 submitted multiple summary revisions. Four students in Group 1 made submissions with no edits in their summaries. These four students were omitted, leaving 10 cases for this exploratory

Table 1 Participants

		Participants	
		Group 1 ($n=38$)	Group 2 ($n=35$)
Gender	Male	9	8
	Female	29	27
Professional background	PK12	9	7
	Corporate	20	13
	NGO	1	3
	Higher education	7	6
	Unknown	1	6

analysis. All 22 students in Group 2 submitted edited versions and thus were retained for the study. We examined the total count of 32 cases to evaluate how students' mental representations changed from initial to final version.

Context

SMART technology

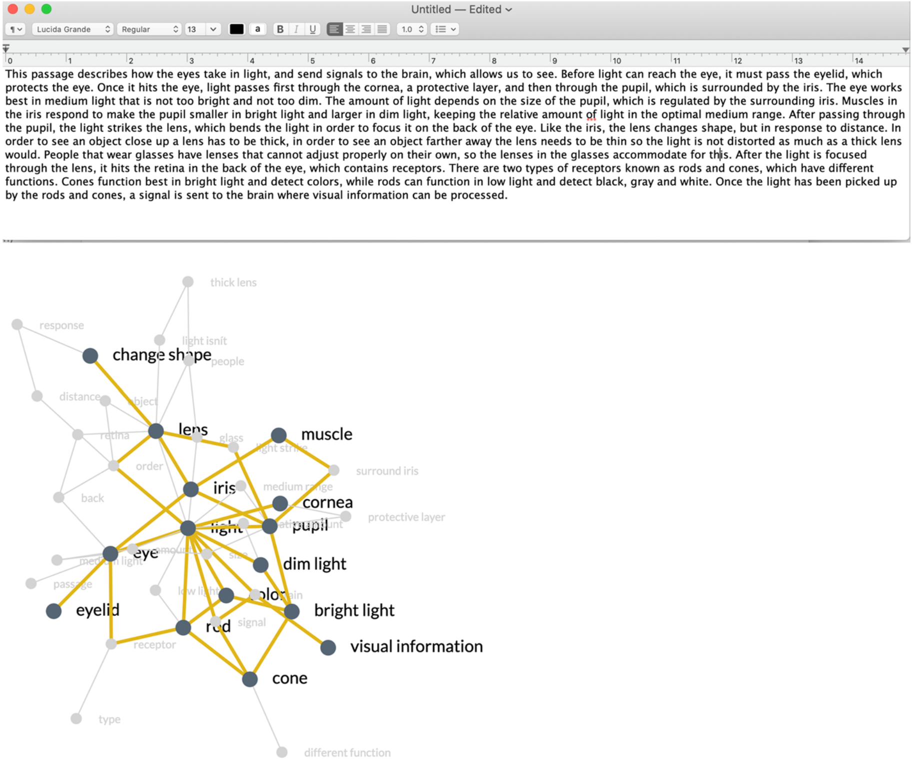
In this study, we used SMART technology due to its three advantages: First, SMART generates the highest number of indices and similarity measures in the 3S dimensions of mental models. Second, SMART provides students with personalized feedback that is delivered in various modalities, including an expert's reference summary, a visualized comparison between the expert and student concept maps, and feedback messages, which, taken together, prompts students to develop a more cohesive summary focused on the key ideas and relations in the text. Lastly, SMART supports file download that includes raw data for each student response so that one may compute new indices.

SMART relies on the study of semantics to obtain text variables (e.g., concepts) and elicit structural information (e.g., an array of the relations between concepts) from a text (Kim et al. 2019). Students read a text and then enter a summary into SMART. SMART then analyzes the summary by comparing it to an expert summary. Students are presented with a concept map of the expert summary as well as a concept map of their summary with colored nodes and line indicating missing concepts and relations. Students also see written feedback about what information they can add or remove from their summary to improve it.

In order to evaluate student learning and provide feedback, SMART models the student's current knowledge (learner model) and compares it to an ideal or expert model (Clariana et al. 2009; Ifenthaler 2014; Kim 2018). This approach starts with the analysis of a text input (i.e., student summary) to identify concepts and concept-to-concept relations in the syntactic structure of the text afforded by natural language processing (NLP) techniques. An array of concept relations builds up a concept map (Axelrod 1976). Analytic algorithms based on graph theory (Rupp et al. 2010; Schvaneveldt et al. 1985; Wasserman and Faust 1994) generate various indices of the student model along the 3S dimensions: surface, structure, semantic. The comparison between a student model and an expert model reveals quantified similarity values and also provides qualitative information about which concepts and relations are overlapping or missing. This comparison drives formative feedback for individual students.

Figure 2 shows a summary of the text about human vision collected in previous studies (Hinze et al. 2013) and the concept map that SMART produces based on this summary. From the first two lines ("This passage describes how the eyes take in light, and send signals to the brain, which allows us to see. Before light can reach the eye, it must pass the eyelid, which protects the eye."), SMART can detect six unique concepts underlined in the example (i.e., passage, eye, light, signal, brain, and eyelid). These concepts build up 4 unique pairwise relations ([passage, eye], [eye, light], [light, signal], [signal, brain], [eye, eyelid]). SMART leverages the NLP dependency analysis to identify the semantic relations (i.e., concept-to-concept) according to the way parts of a sentence are syntactically combined. Theses relations are based on the premise that connected concepts are placed closer to each other with contextual information of the relations (Kim et al. 2019; Baroni et al. 2014; Clariana et al. 2009; Turney and Pantel 2010). The adjacent relation approach is regarded as beneficial because its easiest way to determine concept relations in a text (Clariana 2010).

The number of concepts and relations are basic text variables that constitute the information of the surface dimension. Then, an array of six concepts are transformed to a concept map from which SMART extracts structure-related indices. The concepts and propositional relations used by a student provide the semantic information of the summary.



This passage describes how the eyes take in light, and send signals to the brain, which allows us to see. Before light can reach the eye, it must pass the eyelid, which protects the eye. Once it hits the eye, light passes first through the cornea, a protective layer, and then through the pupil, which is surrounded by the iris. The eye works best in medium light that is not too bright and not too dim. The amount of light depends on the size of the pupil, which is regulated by the surrounding iris. Muscles in the iris respond to make the pupil smaller in bright light and larger in dim light, keeping the relative amount of light in the optimal medium range. After passing through the pupil, the light strikes the lens, which bends the light in order to focus it on the back of the eye. Like the iris, the lens changes shape, but in response to distance. In order to see an object close up a lens has to be thick, in order to see an object farther away the lens needs to be thin so the light is not distorted as much as a thick lens would. People that wear glasses have lenses that cannot adjust properly on their own, so the lenses in the glasses accommodate for this. After the light is focused through the lens, it hits the retina in the back of the eye, which contains receptors. There are two types of receptors known as rods and cones, which have different functions. Cones function best in bright light and detect colors, while rods can function in low light and detect black, gray and white. Once the light has been picked up by the rods and cones, a signal is sent to the brain where visual information can be processed.

Fig. 2 Sample summary of human vision and its network structure

SMART compares individual concepts and relations from the student model to those from the expert model to generate feedback on what key concepts and propositions should be considered for revision.

Summary writing assignments

Students in both groups read an assigned chapter of the text and wrote summaries in the ASE, SMART. They were instructed to write 250–300 words summaries that thoroughly covered the key concepts of the reading. The students were allowed to make multiple revisions to their summaries based on the SMART feedback.

For the cross-validation of the findings, we examined the two groups of students (from different semesters) and had them complete different SMART assignments. Each was a chapter of the textbook, which was 10 pages long (7000–8000 words). Group 1 read a text about constructivism for active, authentic learning. Group 2 was assigned a text about evaluation models in instructional design.

To create the expert model summary, two doctoral students independently wrote an ideal summary of each reading assignment, discussed their drafts together, and then proposed an ideal version of the summaries. SMART analytics automatically generated a list of key concepts from each summary. Two doctoral students also independently evaluated the SMART-generated key concepts and suggested a corrected list of key concepts for each summary. The two lists demonstrated good reliability (Kappa scores; constructivism text: 0.82; evaluation models text: 0.91). These summaries and key concepts were reviewed and approved by the course instructor.

Measures

Standard SMART measures: 3S indices and similarity measures

Leveraging network analysis methods (Kim 2015; Coronges et al. 2007; Wasserman and Faust 1994), SMART computes six concept map indices and similarity indices. These indices maps to the 3S dimensions (Kim 2015; see Table 2). The first three indices, (1) number of concepts, (2) number of relations, and (3) density, relate to the surface dimension. For example, “density” denotes the proportion of possible relations which exist among the concepts of a concept map, indicating the extent of cohesion of a concept map. The density of a concept map is computed by dividing the number of relations actualized in a network by the number of all possible relations, ranging from 0 to 1.

The following three indices (4) average degree, (5) mean distance, and (6) diameter, correspond to the structure dimension. For example, “average degree of a concept map” indicates the average number of edges (relations) that are incident to the concepts in a concept map. In a concept map with n concepts (n is the number of concepts in the concept map), the maximum degree of a concept is ($n - 1$), and the minimum degree is 0. As the number of incoming and outgoing relations grows, the complexity of the cognitive structure is considered higher. The semantic dimension is assessed using similarity values based on particular concepts and their pair-wise relations used in a concept map.

The structure of an optimal concept map is guided by the content of the source text. Thus, it was important not only to understand the changes in a students’ concept map on its own, but also the degree to which the structure of the student concept map compared to an expert concept map elicited from an exemplary summary.

Table 2 Descriptions of similarity measures

Similarity measure	Definition	Indices compared	
		Operationalization	3S structure
Number of concepts[a]	Compare the number of concepts (nodes) in two models	The overall number of concepts	Surface
Number of relations[a]	Compare the number of links (edges) in two models	The overall number of relations of paired concepts	Surface
Density of graphs[a]	Compare the density of the two models	The density of a concept map indicates how cohesive it is.	Surface
Average degree[a]	Compare the average number of degrees in two models	As the number of incoming and outgoing relations grows, the complexity of the cognitive structure is considered higher.	Structure
Mean distance[a]	Compare the mean distances in two models	Indicates how close the concepts are to one another.	Structure
Diameter[a]	Compare the largest geodesics in two models	Represents how broad the understanding of a domain is	Structure
Concept matching	Compare semantically identical concepts, including contextual and principle variables	Qualitative comparison	Semantic
Propositional matching	Compare fully identical propositions (edges) between two concept maps	Qualitative comparison	Semantic
Recall-C	The proportion of key concepts that appear in a student summary	The number of key concepts	Semantic
Recall-P	The proportion of key relations that appear in a student summary	The number of key relations	Semantic

[a] 3S indices available in SMART. The table was modified from Kim (2015, p. 8)

We do not argue that the expert summary in SMART is the *only* correct way that information from the source text can be represented, but rather reflects a more optimal structure than what is typically produced by a novice. It is true that experts' concept maps can vary when experts build on their epistemological preference to make inferences connecting preferred information from the text to their prior knowledge. However, when writing summary focuses on identifying the information the author structures in a text, expert readers can build a similar knowledge structure as intended by the writer, connecting key ideas from different parts of the text (Graesser et al. 1994; Kintsch 1988). Experts represent information in qualitatively different ways than their novice counterparts (e.g., Alexander 2003, 2004; Chi 2006). Thus, there may be some variability from expert to expert, but they are likely to have "clearly recognizable patterns in the problem conceptualization" (Spector 2008 p. 31).

In this regard, SMART evaluate the quality of a student summary by comparing the student concept map to a reference model (expert model). This generates similarity measures, ranging 0 (completely dissimilar) to 1 (completely similar). SMART uses two types of similarity formulas: numerical and conceptual similarity (Kim et al. 2016, 2019). The numerical similarity formula which compares two numerical measures from a student and an expert model is used for all the 3S indices.

$$s = 1 - \frac{\left|v_1 - v_2\right|}{max\left(v_1, v_2\right)}$$

where $v1$ is the index value of a student model, and $v2$ is the value of an expert model. In contrast, the conceptual similarity that indicates the extent to which a student model embeds the same elements found in the expert model relates to the semantic dimension, including concept matching and propositional matching. The conceptual similarity draws on Tversky's (1977) similarity formula:

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

where '*A*' is a student model, and '*B*' is a reference model. The weighting functions, α and β, were set as 0.7 and 0.3, according to the suggestion that α should be weighted higher than β in an asymmetric relation wherein a student model resembles a reference model (Kim 2015).

In addition, SMART provides two semantic-related similarity measures, Recall-C and Recall-P, which indicate the proportion of fully identical key concepts (i.e., the central ideas of the text) and the relations of key concepts. To calculate these, SMART uses a simple formula: Recall measure = the number of key elements in a student model/the total number of key elements in an expert model.

Additional SMART measure: graph centrality

Central to this study is GC—the degree centrality of a whole concept map computed by extending individual concepts' degree centrality values into the entire network level (Clariana et al. 2011; Kim 2017). A value of GC is calculated as follows:

$$\text{Degree centrality of a concept}: \ \text{DC}(v) = \text{degree}(v)/(n - 1) \tag{1}$$

$$\text{Degree centrality of a graph}:\ \mathrm{DC}(G) = \sum (i = 1\ \mathrm{to} v)\,[\mathrm{DC}(v*) - \mathrm{DC}(vi)]/(n-2) \tag{2}$$

where v is a concept in a network graph, n indicates the total number of concepts in a network graph, DC (v*) indicates the highest degree centrality of a concept, and DC (vi) is the degree centrality of the ith concept. Thus, GC yields a value between 0 and 1. Lower scores (i.e., DC(G) below 0.1) are assumed to reflect a goal-orientation structure that optimally integrates concepts and propositions essential for the meaning of the text, whereas higher scores (i.e., DC(G) greater than 0.6) reflect naïve epistemology that indicates the lack of focused, organized understanding of the text (Hay and Kinchin 2006; Yin et al. 2005). We hypothesized that GC could serve as a global index that indicates students' learning progress in building a solid understanding of the text. Specifically, in the context of the SMART analytics, we presumed that a cohesive and integrated mental model could include many chain structures (thoughtful sentences in a summary) connected to a net shape of a concept map, which resulted in a GC index value closer to the goal-orientation threshold value of 0.1.

Analysis procedure

SMART produces concept map information that includes basic network data (i.e., an array of the concepts in a concept map) and the 3S indices and similarity measures of a student's concept map. We used the network data to compute GC values of the concept maps by means of a social network analysis application (NodeXL; Hansen et al. 2010). We identified those students who demonstrated the greatest GC changes from initial summary to final submission and then examined these students' summaries for changes in the 3S indices and similarity measures. Finally, we performed paired samples t-tests to examine the significant change between the initial and final summaries in GC, 3S indices, and similarity measures.

Results

GC change from initial to final version

GC values

We inspected students' revision trajectories by examining the GC values of the student models from the initial to the final models. We categorized each student as either positive or negative GC change and then further identified those students who demonstrated the largest change in each direction. Applying the 75th percentile value (i.e., highest change > 75th percentile value) resulted in four classifications: high-positive, low-medium positive, low-medium negative, and high-negative change. For each group, we examined GC values of the initial and the final summaries of each case (Table 3). Group 1 had only two positive GC changes and thus we did not divide them, resulting in three categories. In contrast, Group 2 showed the high-positive cases against the low-medium positive, forming four categories: high positive (over the 75th percentile value of the positive cases, GCI = 0.148 and GCF = 0.265), low-medium positive (below the 75th percentile value of the positive cases, GCI = 0.131

Table 3 Trends in graph centrality change

Group	Change direction		GCI	GCF	GCF-GCI
Group 1	Positive	Positive total ($n=2$)	0.11	0.13	0.02
		Subtotal ($n=2$)	0.11	0.13	0.02
	Negative	Low–medium ($n=6$)	0.14	0.12	– 0.02
		High ($n=2$)	0.20	0.09	– 0.11
		Subtotal ($n=8$)	0.17	0.10	– 0.07
	Total ($n=10$)		0.15	0.12	– 0.03
Group 2	Positive	High ($n=2$)	0.148	0.265	0.117
		Low–medium ($n=7$)	0.131	0.149	0.017
		Subtotal ($n=9$)	0.135	0.175	0.039
	Negative	Low–medium ($n=10$)	0.144	0.125	– 0.019
		High ($n=3$)	0.193	0.110	– 0.083
		Subtotal ($n=13$)	0.155	0.121	– 0.034
	Total ($n=22$)		0.147	0.143	– 0.004

GCI graph centrality in initial models, *GCF* graph centrality in final models

and GCF = 0.149), low-medium negative (below the 75th percentile value of the negative cases, GCI = 0.144 and GCF = 0.125), and high negative (over the 75th percentile value of the negative cases, GCI = 0.193 and GCF = 0.110).

Figure 3 shows each students' change in GC. The two figures are separated by semester and text group (i.e., Groups 1 and 2, respectively). The dotted line indicates the GC score of the expert model. Negative trends are indicated with downward arrows, and positive trends are indicated in upward arrows. There are two important things to note in this analysis. The first is that most students demonstrate a negative trend toward an expert line, while positive trends tend to depart from an expert's GC value (e.g., S10 and S16 in Group 1; and S62 and S71 in Group 2). The second is that the majority of GC trends tend to move toward the expert line. That is, most students revised their summaries in a way that made their new concept map look more like the expert concept map through *reduced* graph centrality score.

GC similarity

We used an expert model for each text to compute similarity values of GC measures for both initial and final models (Table 4). Intriguingly, students who demonstrated a negative GC trend showed a stronger increase in GC similarity to the expert model than those who had demonstrated positive GC trend.

These data, in consideration with Fig. 3, indicate that students tend to generate more diffuse concept maps (i.e., higher GC) than the experts, whose concept maps tend to be at or below the .10 GC threshold indicative of a mental model connecting well-reasoned chain-type substructures. Thus, students who were better able to reduce and centralize their ideas yielded more expert-like concept maps, while those who increased their graph centrality (i.e., positive GC trend) tended to move away from or "overshoot" the idea graph structure.

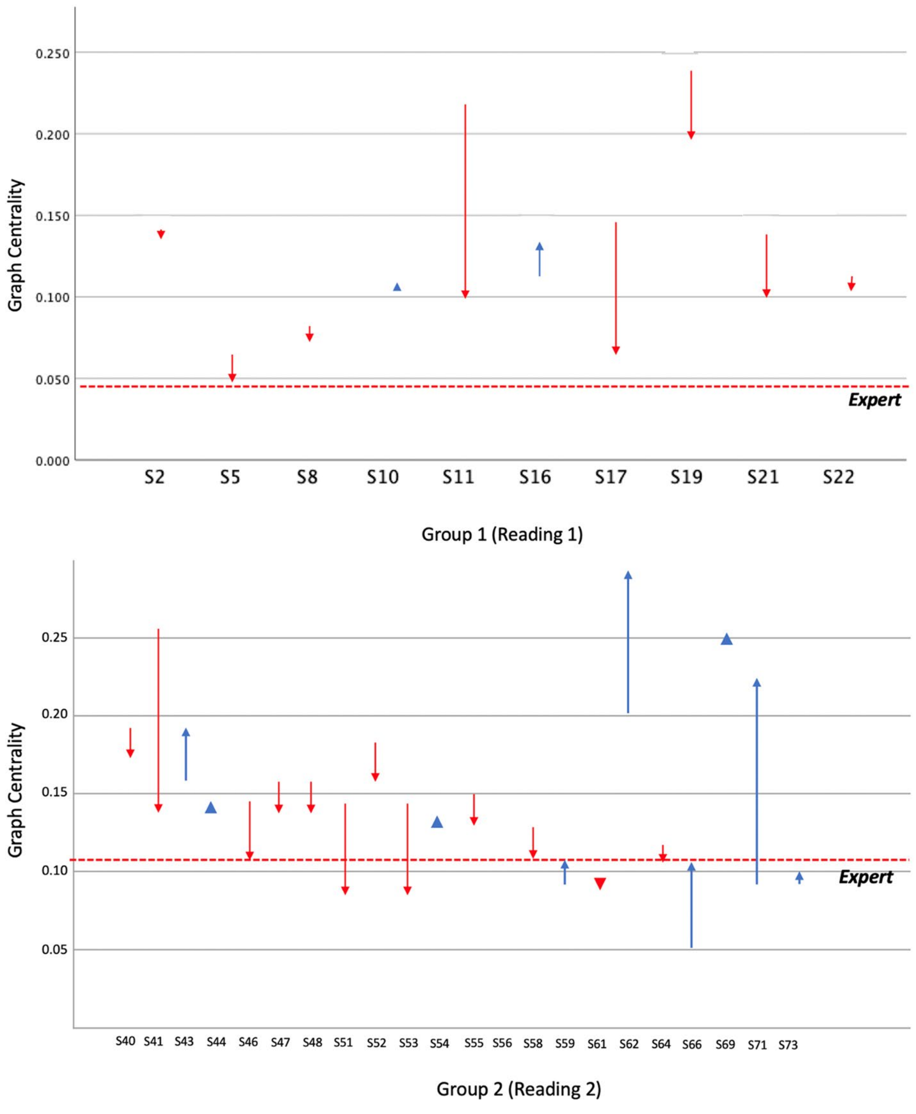


Fig. 3 Trends in graph centrality values. In Group 1 (Positive [S10, S16], Low-Medium Negative [S2, S5, S8, S19, S21, S22], and High Negative [S11, S17]). In Group 2 (High Positive [S62, S71], Low-Medium Positive [S44, S54, S73, S69, S59, S43, S66], Low-Medium Negative [S40, S46, S47, S48, S51, S52, S55, S56, S58, S61], and High Negative [S41, S64, S53])

Relationships between GC and 3S indices

3S indices by GC trends

We then examined how the changes in GC, as a reflection of overall changes of the mental models, related to changes in the 3S indices. To foreshadow our findings, analyses

Table 4 Trends in graph centrality similarity

Group	Change direction		SGCI	SGCF	SGCF-SGCI
Group 1	Positive	Positive total (*n* = 2)	0.494	0.425	– 0.069
		Subtotal (*n* = 2)	0.494	0.425	– 0.069
	Negative	Low–medium (*n* = 6)	0.459	0.560	0.101
		High (*n* = 2)	0.297	0.685	0.388
		Subtotal (*n* = 8)	0.419	0.591	0.172
	Total (*n* = 10)		0.434	0.558	0.124
Group 2	Positive	High (*n* = 2)	0.687	0.433	– 0.255
		Low–medium (*n* = 7)	0.722	0.787	0.065
		Subtotal (*n* = 9)	0.714	0.708	– 0.006
	Negative	Low–medium (*n* = 10)	0.739	0.813	0.074
		High (*n* = 3)	0.622	0.819	0.197
		Subtotal (*n* = 13)	0.712	0.814	0.103
	Total (*n* = 22)		0.713	0.771	0.058

SGCI similarity of the graph centrality in initial models, *SGCF* similarity of the graph centrality in final models

revealed that changes in GC were driven by complex combinations of the 3S dimensions (see Table 5).

Analysis indicated no uniform pattern of change in indices that reflected the surface dimension (i.e., number of concepts, number of relations, density). In general, students increased the number of words, concepts, and concept relations in their final versions, which resulted in a slight decrease in the density values. However, there was variability in this across the four GC trend groups. A uniform change (addition behavior) occurred in the low-medium directions, but not in the extreme GC trend (i.e., the high positive and negative). For example, the high negative trend in Group 1 included more words in their revisions, whereas that trend in Group 2 reduced the words. These findings imply that GC may not be directly associated with surface-level changes.

The structure-related indices of Average Degree and Mean Distance tended to slightly decrease or remain at a similar level across two groups. In contrast, Diameter remained at a similar network size or demonstrated modest increase. This suggests that students' summary revisions led to concept maps that had a slightly larger (Diameter), but more closely connected (Average Degree, Mean Distance) network. The results of these indices indicated that their potential relationships with GC values. However, individual indices might not be enough to explain goal-oriented changes toward an expert-like model.

The semantic-related indices are not directly measured from a concept map but computed by the comparison to the expert concept map. Since the semantic features of a student concept map are inherently similarity measures, the semantic changes are discussed in the following section.

3S similarity measures by GC trends

Similar to the way we compared student GC to expert GC, we wanted to explore how if changes in these 3S indices moved students toward more expert-like values. Thus, we calculated 3S similarity measures for each GC group (Table 6). In general, students' concept

Table 5 3S indices by GC trends

Group	3S index		Attempt	GC trend				
				N-H	N-LM	P-LM	P-H	Total
Group 1	Surface	NW	I	243.500	250.333	–	393.500	277.600
			F	259.000	318.000	–	330.500	308.700
			F–I	15.500	67.667	–	– 63.000	31.100
		NC	I	56.500	57.500	–	81.000	62.000
			F	63.500	74.167	–	74.500	72.100
			F–I	7.000	16.667	–	– 6.500	10.100
		NR	I	75.000	69.667	–	111.500	79.100
			F	79.000	90.333	–	97.000	89.400
			F–I	4.000	20.667	–	– 14.500	10.300
		DE	I	0.046	0.045	–	0.033	0.043
			F	0.039	0.037	–	0.036	0.037
			F–I	– 0.007	– 0.008	–	0.003	– 0.005
	Structure	AD	I	2.522	2.397	–	2.614	2.465
			F	2.403	2.425	–	2.551	2.446
			F–I	– 0.119	0.029	–	– 0.063	– 0.019
		MD	I	4.058	4.881	–	4.410	4.622
			F	4.037	4.272	–	4.451	4.261
			F–I	– 0.020	– 0.609	–	0.04	– 0.361
		DIA	I	11.500	14.500	–	13.000	13.600
			F	13.500	17.167	–	13.000	15.600
			F–I	2.000	2.667	–	0.000	2.000
Group 2	Surface	NW	I	242.333	291.500	274.857	270.500	277.591
			F	239.000	304.800	297.286	339.500	296.591
			F–I	– 3.333	13.300	22.429	69.000	19.000
		NC	I	60.000	72.900	69.429	61.500	69.000
			F	60.000	76.900	75.143	85.500	74.818
			F–I	0.000	4.000	5.714	24.000	5.818
		NR	I	79.667	97.000	88.429	75.000	89.909
			F	75.667	101.000	94.714	117.500	97.045
			F–I	– 4.000	4.000	6.286	42.500	7.136
		DE	I	0.043	0.037	0.038	0.040	0.039
			F	0.042	0.035	0.035	0.032	0.036
			F–I	– 0.002	– 0.003	– 0.003	– 0.008	– 0.003
	Structure	AD	I	2.449	2.549	2.464	2.422	2.497
			F	2.411	2.528	2.463	2.603	2.498
			F–I	– 0.038	– 0.021	– 0.001	0.181	0.001
		MD	I	4.110	3.941	3.818	3.811	3.913
			F	4.027	3.935	3.702	3.775	3.859
			F–I	– 0.083	– 0.006	– 0.116	– 0.035	– 0.054
		DIA	I	12.333	12.700	13.714	13.500	13.045
			F	12.333	13.300	13.714	12.000	13.182
			F–I	0.000	0.600	0.000	– 1.500	0.136

Number of Words (NW), Number of Concepts (NC), Number of Relations (NR), Density (DE), Average Degree (AD), Mean Distance (MD), Diameter (DIA), High Negative Trend (N-H), Low-Medium Negative

Table 5 (continued)

Trend (N-LM), Low-Medium Positive Trend (P-LM), High Positive Trend (P-H)

maps became more similar to the experts' in each of the three dimensions. Although some structure-related similarity measures showed some slight decreases.

In surface-related similarity measures, students with a high negative GC trend showed relatively strong similarity values and an increase from initial to final models in Similarity of the Number of Concepts (SNC) and Similarity of Density (SDE), while the positive GC trends yielded a high increase in Similarity of the Number of Relations (SNR). Overall, students who demonstrated a negative GC trend built a cohesive network (high SDE) due to the inclusion of appropriate number of concepts.

Structure-related similarity measures produced complex trends. For Group 1, Similarity of Average Degree (SAD), Similarity of Mean Distance (SMD), and Similarity of Diameter (SDI) tended to have a high value in the negative GC trends, but some showed a decrease in values from initial to final models (i.e., SAD in the high negative trend and SDI in the low-medium negative trend). Group 2 showed more diverse patterns. SAD, SMD, and SDI levels were similar across the GC trends. The highest value of SAD was found in the high positive trend, SMD value was dropped in the high-negative trend, and similar to Group 1, SDI in the high-negative trend remained at an appropriate level from initial to final models, which meant that the high-negative trend tended to form a cohesive concept map with a size similar to an expert model.

Notably, GC is calculated based on a structural measure (degree centrality). Intriguingly, GC appeared to be strongly associated with the semantic-related similarity measures. In both groups, the high-negative GC trend students demonstrated a higher value and a greater increase in Conceptual Matching (CM) and Propositional Matching (PM), while the high-positive trend showed a highest similarity value and a greater increase in the Recall-C and Recall-P. The results indicated that the negative GC trends, especially the high-negative, related to students' effort in revisions to include concepts and relations used in an expert model. It made sense that a higher Recall-C and Recall-P value was found in the high-positive GC trend since during revisions students in that trend tended to include more words (Table 5), likely including key concepts and key relations as directly suggested by the SMART feedback. However, overall concepts and propositions used in their summaries tended to be less similar than students who demonstrated negative trends, which indicating that they might simply adopt the suggested key words without much deliberation.

Validation of the descriptive findings

Statistical analysis

In order to examine these changes quantitatively, we conducted a series of paired samples t-tests. Given the small sample size of Group 1, we conducted these statistical analyses on only the Group 2 data. We computed two data sets: One used all cases ($n = 22$) and the other had 20 cases, dropping two high-positive cases as outliers. Post-hoc power analyses with effect size = 0.5, $\alpha = .05$, and power = 0.7 and 0.8 suggested samples sizes of 21 and 27, respectively. Thus, our sample of 22 is acceptable, but may be too small to detect more nuanced effects. Thus, caution is advised when interpreting the results.

Table 6 Similarity measures by GC trends

Group	3S index similarity		Attempt	GC Trend				
				N-H	N-LM	P-LM	P-H	Total
Group 1	Surface	SNC	I	0.843	0.770	–	0.835	0.798
			F	0.918	0.761	–	0.884	0.817
			F–I	0.075	− 0.009	–	0.049	0.020
		SNR	I	0.915	0.783	–	0.779	0.809
			F	0.837	0.794	–	0.845	0.813
			F–I	− 0.077	0.011	–	0.066	0.004
		SDE	I	0.794	0.759	–	0.909	0.796
			F	0.937	0.720	–	0.802	0.780
			F–I	0.143	− 0.039	–	− 0.107	− 0.016
	Structure	SAD	I	0.941	0.935	–	0.900	0.929
			F	0.935	0.943	–	0.920	0.937
			F–I	− 0.006	0.008	–	0.020	0.008
		SMD	I	0.889	0.767	–	0.921	0.822
			F	0.936	0.874	–	0.930	0.898
			F–I	0.046	0.107	–	0.009	0.075
		SDI	I	0.885	0.761	–	0.791	0.792
			F	0.964	0.627	–	0.791	0.727
			F–I	0.080	− 0.135	–	0.000	− 0.065
	Semantic	CM	I	0.227	0.180	–	0.145	0.182
			F	0.523	0.229	–	0.227	0.287
			F–I	0.296	0.049	–	0.082	0.105
		PM	I	0.059	0.016	–	0.012	0.024
			F	0.368	0.065	–	0.044	0.122
			F–I	0.310	0.049	–	0.032	0.098
		RC	I	0.450	0.267	–	0.350	0.320
			F	0.700	0.617	–	0.750	0.660
			F–I	0.250	0.350	–	0.400	0.340
		RP	I	0.117	0.056	–	0.100	0.077
			F	0.500	0.461	–	0.617	0.500
			F–I	0.383	0.406	–	0.517	0.423

Table 6 (continued)

Group	3S index similarity		Attempt	GC Trend				
				N-H	N-LM	P-LM	P-H	Total
Group 2	Surface	SNC	I	0.832	0.844	0.843	0.891	0.846
			F	0.870	0.850	0.859	0.831	0.854
			F–I	0.038	0.005	0.016	– 0.060	0.007
		SNR	I	0.842	0.807	0.840	0.806	0.822
			F	0.814	0.831	0.852	0.812	0.834
			F–I	– 0.029	0.025	0.012	0.005	0.012
		SDE	I	0.824	0.859	0.829	0.937	0.851
			F	0.909	0.858	0.830	0.852	0.856
			F–I	0.086	0.000	0.001	– 0.085	0.004
	Structure	SAD	I	0.944	0.952	0.948	0.955	0.950
			F	0.951	0.964	0.962	0.974	0.963
			F–I	0.006	0.012	0.014	0.019	0.013
		SMD	I	0.967	0.933	0.919	0.875	0.928
			F	0.918	0.936	0.913	0.931	0.926
			F–I	– 0.049	0.002	– 0.006	0.056	– 0.002
		SDI	I	0.853	0.816	0.817	0.964	0.835
			F	0.853	0.821	0.838	0.849	0.833
			F–I	0.000	0.004	0.021	– 0.115	– 0.002
	Semantic	CM	I	0.533	0.452	0.384	0.402	0.334
			F	0.306	0.377	0.330	0.174	0.437
			F–I	– 0.227	– 0.075	– 0.054	– 0.228	0.103
		PM	I	0.084	0.075	0.091	0.063	0.080
			F	0.248	0.137	0.121	0.181	0.151
			F–I	0.165	0.062	0.030	0.119	0.071
		RC	I	0.256	0.523	0.473	0.269	0.448
			F	0.667	0.754	0.659	0.808	0.717
			F–I	0.410	0.231	0.187	0.538	0.269
		RP	I	0.111	0.367	0.321	0.167	0.299
			F	0.597	0.683	0.565	0.729	0.638
			F–I	0.486	0.317	0.244	0.563	0.339

Similarity of the Number of Concepts (SNC), Similarity of the Number of Relations (SNR), Similarity of the Density (SDE), Similarity of the Average Degree (SAD), Similarity of the Mean Distance (SMD), Similarity of the Diameter (SDI), Concept Matching (CM), Propositional Matching (PM), Recall-C (RC), and Recall-P (RP), High Negative Trend (N-H), Low-Medium Negative Trend (N-LM), Low-Medium Positive Trend (P-LM), High Positive Trend (P-H)

Table 7 summarizes the paired sample *t*-tests. We used one-tailed *t* test since we hypothesized one directional change in those measures. Graph Centrality (GC) showed significant change when two extreme cases were removed with $t(19) = 1.881$, $p = .038$. As to model indices, the Number of Concepts (NW) was only significant with a larger sample ($n = 22$). However, Density (DE) demonstrated a significance in both $n = 22$ and $n = 20$. DE may be an indicator of change that combines changes in NW and NR together. None of the structure indices yielded a significant change, implying various, multi-direction shifts happened in knowledge structure while students revised.

From the perspective of assessment, similarity measures are more important than model indices. GC similarity was improved from initial to final models, $t(19) = 2.412$, $p = .013$, indicating students built a knowledge structure similar to the expert structure. No surface similarity measure was significant in change. Out of the structural similarity measures, Average Degree (AD) similarity was only significant in change with the sample size of 22. As observed in the descriptive statistics (Table 6), similarity measures in the surface and structure dimensions did not show a significant pattern due to nonuniform changes in their indices. Conversely, a strong positive change in the similarities of the semantic-related measures was found. All semantic similarity changes were significant, including Concept Matching (CM), Propositional Matching (PM), Recall-C, and Recall-P in both $n = 22$ and $n = 20$, with p = .000. These results suggest that when the students revise their summaries, they wrote a summary with more semantically similar information to the expert's summary.

Table 7 Paired samples tests for group 2

	Measure	*t-value*	*df*	Sig.	Excluding S62 and S71 (*df* = 19)	
					t-value	*Sig.*
Model index	Graph centrality (GC)	0.340	21	0.369	1.881	0.038
	Number of words (NW)	1.397	21	0.088	0.996	0.166
	Number of concepts (NC)	2.084	21	0.025	1.549	0.069
	Number of relations (NR)	1.518	21	0.072	0.840	0.206
	Density (DE)	2.972	21	0.004	2.534	0.010
	Average degree (AD)	0.041	21	0.484	0.512	0.307
	Mean distance (MD)	0.865	21	0.198	0.853	0.202
	Diameter (DIA)	0.191	21	0.425	0.391	0.350
Similarity measure	Similarity-GC	1.444	21	0.082	2.412	0.013
	Similarity-NC	0.357	21	0.362	0.765	0.227
	Similarity-NR	0.534	21	0.300	0.585	0.283
	Similarity-DE	0.189	21	0.426	0.646	0.263
	Similarity-AD	1.975	21	0.031	1.707	0.052
	Similarity-MD	0.209	21	0.418	0.678	0.253
	Similarity-DIA	0.069	21	0.473	0.347	0.366
	Concept matching	5.385	21	0.000	4.936	0.000
	Propositional matching	4.710	21	0.000	4.097	0.000
	Recall C	6.882	21	0.000	6.453	0.000
	Recall P	6.794	21	0.000	6.054	0.000

$N = 22$. $p < .05$. One-tailed *t*-test

Lastly, we tested the correlations between GC similarity change and semantic similarity change, using the Group 2 data ($n = 20$). As depicted in Table 8, no correlation was found in the whole data ($n = 20$) or in the positive trend ($n = 7$). However, GC similarity change in the negative trend ($n = 13$) demonstrated a positive correlation with the change in Concept Matching (CM) and Propositional Matching (PM), $r = 0.622$ and 0.609, with $p < .05$, respectively. Given the effect size, a minimum of expected samples ranged from 13 (with power = 0.7) to 15 (with power = 0.8). Accordingly, the sample size of 13 was acceptable. A negative trend in GC values indicated an increase in GC similarity (meaning that student models become similar to an expert model in terms of GC) and in the overall quality of the models in regard to the used concepts and propositions.

Qualitative review: visual inspections

To further examine, we conducted qualitative analyses for three cases: the reference model (expert model), student 71 (who demonstrated the highest positive change in GC values), and student 41 (who uniquely showed great negative changes in GC values).

As Fig. 4 shows, the expert concept map forms a cohesive macrostructure in which many key concepts (the squares) weaved sub-chain structures together (Kintsch 1998). The local structures of the summary (i.e., a sentence or a paragraph) appear linear and goal-directed (Hay and Kinchin 2006; van Dijk and Kintsch 1983). In contrast, Student 71 starts with a model that embeds no single key concept and unconnected line shapes. The student's final summary create a complex net structure with few sub-chains and a unconnected tail, indicative of naïve epistemology (Hay and Kinchin 2006). Although this model embeds several key concepts, the key concepts do not serve as cohesion cues (van Dijk and Kintsch 1983) mapping local and distal constituents in the text (i.e., relationships between linear subgroups–sentences). While the student reduces the size of the model, the key concepts are located off to the side of the model.

Conversely, Student 41 draws a cohesive model that includes some key concepts standing toward the center of the network. The final model of student 41 demonstrates a goal-directed structure with most key concepts mapping sub-chain components similar to the expert model. This case shows that the higher degree of negative GC change in a model structure perhaps relates to the mental model change toward a proper knowledge structure of the text.

Table 8 Correlations between GC similarity change and semantic similarity changes

	CMD	PMD	RCD	RPD
Total ($n = 20$)	0.392	0.361	− 0.172	− 0.168
Negative Direction ($n = 13$)	0.622*	0.609*	− 0.043	− 0.003
Positive direction ($n = 7$)	− 0.128	− 0.196	− 0.430	− 0.481

Concept matching difference (CMD) = concept matching value of the final model—concept matching value of the initial model. Similar formula was applied to propositional matching, Recall-C and Recall-P as well. *$p < .05$, **$p < .01$

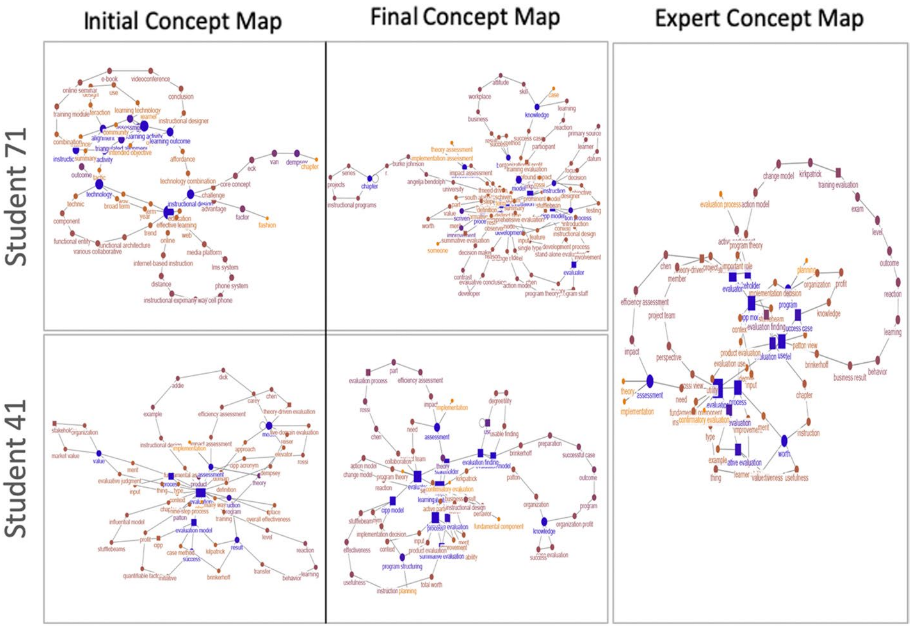


Fig. 4 Graphical representations of knowledge structure

Discussion

In this current study, we investigated the potential of Graph Centrality (GC) as a global index that describes the overall quality of students' knowledge structure elicited from their summary of a complex reading material. This initial exploratory review of the global index yielded that GC could serve as an indicator of students' progression in writing a quality summary of the text.

Referring to changes in GC and the 3S indices, GC measures showed an interesting trend that many students started with a larger GC value and shifted to a smaller GC value similar to the expert's GC value. The negative GC pattern indicated that students revised their summaries to embed a mental structure similar to the expert's model. Although the 3S indices (i.e., the surface and structure indices) showed that the students tended to include more concepts and relations and thus created a larger and more cohesive concept map than their initial version, there was no uniform change in those indices.

The following similarity measures confirmed that students in the negative trend constructed a mental representation toward the expert model, while their counterparts in the positive trend (an increase in the GC value) created a naïve model structure dissimilar to the reference when they revised their summaries. The extent of the changes in GC values in both directions indicated the strength of medication in the students' mental representations. Overall, the negative trend in GC values meant an improved similarity of Density, meaning that the students' model models became more cohesive.

A significant similarity increase was found in the semantic-related similarity along with the GC similarity. Especially, GC similarity change in the negative trend showed a positive correlation with changes in Concept Matching and Propositional Matching similarity, indicating that structural similarity change described by GC measures also explained the quality of the students' mental models in terms of the inclusion of concepts and propositions used by the expert model.

Implications

We discuss the implications of this study in theory, methodology, and pedagogy. Table 9 recapitulates the model-based approach and graph centrality in comparison to the typical text-based approach that use lexical components to compute hundreds of indices (Crossley et al. 2016; Kyle et al. 2018).

Theoretically, the findings of the study espouse the assumptions of knowledge representations. Compared to the text-based approach, the model-based approach takes advantage of a concept map–externally re-represented learner mental model. Graph theory suggests that human mind reflected in a written summary of a text can be modeled in the form of a graphical representation (i.e., concept map) that in turn explains the quality of comprehension (Rupp et al. 2010; Schvaneveldt et al. 1985; Wasserman and Faust 1994). GC as an index of mental models is only available for model-based ASE tools. GC measures are able to indicate the overall quality of individuals' mental representations, and the changes in GC measures denotes whether students develop appropriate understanding of the text when reconfiguring their mental structure.

Methodologically, the model-based approach provides more elaborated mental model dimensionality than the text-based approach (i.e., three vs. two dimensions). The 3S dimensions have been significantly explained by corresponding indices in previous validation studies (Kim 2015; Ifenthaler 2009). Upon such an empirically tested framework, we can compute Graph Centrality (GC), using the degree of individual concepts that are derived from the network analysis. In spite of its computational simplicity, GC can serve as a global index that describes a holistic structure of a mental model. Specifically, GC measures explains the extent to which students develop a mental structure that takes an appropriately complex net structure (macrostructure) connecting well-defined chain-type structures (microstructure). Also, GC as a global index shows high correlations with trends in the surface and semantic indices. Li et al. (2017) warn that many studies have shown no effect of ASEs due to the use of a wide variety of measures. The validated 3S structure of the model-based approach can provide a reliable assessment framework, and GC can serve as a cross-reference to diagnose students' growth in their understanding of the texts.

Pedagogically, GC computing formula can be easily added to model-based automated summary evaluation (ASE) tools such as SMART. GC can serve a feedback index that inform students of their learning progression in writing an expert-like summary on top of the existing feedback information–learning progression feedback (Fonger et al. 2018; Nadolski and Hummel 2017). Given the trends explained by GC values, one can use the 3S indices and similarity measures to further detail the characteristics of individual students' mental representations while they revise summaries. In other words, we can create two-step formative feedback. At the first step, GC enables us to provide students with the information of their general progression toward an expert-like summary. This feedback information is an overall judgement similar to latent semantic analysis (LSA) and the percentage of n-gram co-occurrences. LSA determines related words based on their locations in a hypothetical space and thus is not able to specify exact concepts and relations in the text (Deerwester et al. 1990). N-gram stands for a contiguous sequence of n items from a given

Table 9 Comparison of model-based and text-based approaches to summary analysis

	Model-based	Text-based	
		LSA	Machine-translated
Tool	AKOVIA (Ifenthaler 2014); GISK (Kim 2018); HIMATT (Pirnay-Dummer and Ifenthaler 2011); SMART (Kim et al. 2019)	Coh-Metrix (McNamara et al. 2014); SEMILAR (Rus et al. 2013); Summary Street (Wade-Stein and Kintsch 2004); Online Summary Assessment and Feedback System (Sung et al. 2016); crowd-source summary evaluation (Li et al. 2018)	ROUGE (Lin 2004); PryEval (Gao et al. 2019); TAALES (Kyle et al. 2018)
Assessment target	Concept map	Text-base (word, sentence, and discourse levels)	
Dimension	Surface	Surface	
	Structure	Deep (situational model)	
	Semantic		
Graph centrality	Applicable	N/A	
Key measure	3S indices	LSA cosine	N-gram co-occurrence
Key advantage	Concept maps as feedback information; deep comprehension focused; multi-level feedback (e.g., two-step feedback)	Feedback on writing styles (e.g., content word overlap, lexical diversity, and cohesive text structure; over a hundred linguistic and discourse measures.	

text (He et al. 2009; Lin 2004). Together with a generic feedback on GC, a more elaborated feedback information can be created, referring to the 3S indices in terms of what concepts and relations are missing and how to organize the summary cohesively. Thus, model-based metrics and GC may be a powerful tool for automatic writing evaluation (AWE) and learning analytics more generally. Automated writing evaluation and feedback has largely focused on improving the writing itself. While this is an important endeavor, it makes these tools less suitable for content courses. Writing and revising summaries with GC-driven feedback supports students in identifying key concepts and relations prior to class so that they can be prepared to engage in more meaningful activities. In addition, instructors can refer to GC values for student summaries to evaluate the class-level and the student-level progressions in real-time, which will be much simpler than looking a combination of the 3S indices together. That is, concept maps with a GC value could serve as a useful at-a-glance dashboard feature so that instructors could more easily spot outliers or patterns that need to be addressed. For example, initial GC values can serve for the early detection of students who lack pre-existing understanding of the text, and the changing direction of GC values (i.e., positive trend over the expert's value) can signify a naïve epistemology of students who may need further assistance of instructors.

Limitations and suggestions

Despite promising findings of this initial work, more studies are required to fully validate the value of GC index for ASE. One limitation is the small number of cases. The current study used 32 cases to examine the potential of GC as a global index. Admittedly, a single case study with small sample size is insufficient to justify the index's validity. The utility value of the GC index needs to be calibrated by means of repeated tests with different types of texts, various expert models, and larger samples of student summaries. Future studies using a large sample size can allow us to break down the cases according to the GC trends and test a significance of the changes in the 3S indices and the similarity measures per each GC trend with sufficient power.

Another limitation is the research context. This study used summaries of the texts in the areas of learning science from a graduate-level course. A knowledge structure could be positioned in summary in different ways according to the contexts, including disciplines, the genre of the texts, sizes of the corpus, and student levels. For example, K-12 students in science classrooms could write a summary in different styles, depending on the genre of reading materials. Students might summarize sequencing events in cause-effect relationships or scientific definitions and informational reports. The former could take more linear and chain structures, while the latter became more net structures. Both what information should be foregrounded in the mental model and how that information is conveyed in writing are dependent upon the conventions and demands of different genres and disciplines (e.g., Goldman et al. 2016; León et al. 2006). Writing styles in various genres and topics of texts could engender different forms of knowledge structures embedded in summaries.

In addition to theoretical differences across disciplines, there are also methodological considerations. In the present experiment, students were asked to write summaries of approximately 250–300 words. Longer summaries are likely to contain a greater number of concepts and relations and thus create a bigger and more complex concept map. Accordingly, we need to test and validate GC across a greater variety of summary lengths and explore how properties of the concept map might change across length constraints.

Lastly, GC values may depend on writing styles and ASE tools' modeling algorithms. The experts' models used in this study generated GC values less than 0.1, which met the goal-orientation threshold suggested by the literature (Hay and Kinchin 2006; Yin et al. 2005). However, depending on the functions to extract concepts and relations from texts and techniques to form a concept map, ASE tools may create different mental representations, using the same written summary. Accordingly, future studies need to explore GC values calculated from data derived from other model-based ASE tools so as to see if GC trends are similar across ASE platforms as suggested by the theory.

Conclusion

The present case study explored the use of Graph Centrality (GC) as a global index of students' mental model quality and, in particular, students' mental model change as they wrote and revised summaries in SMART. Analyses revealed that GC change (i.e., negative vs. positive change) related to how the students modified the wholistic structure of their mental models. The graph centrality index may serve as a way to better characterize these important qualitative shifts in mental model structure that may not be immediately apparent in the 3S indices. Methodologically, this study suggests that the GC index could be beneficial to describe the overall change of a student mental model along with the 3S indices (e.g., surface, structure, semantic). Pedagogically, the GC index available in a formative assessment and feedback technology like SMART could track the overall mental model trajectories in real-time and inform students of the quality of their understanding of the text and help instructors to deliver targeted feedback and support.

Compliance with ethical standards

Conflict of interest Min Kyu Kim declares that he has no conflict of interest. Kathryn S. McCarthy declares that she has no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or the national research committee.

References

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher, 32*(8), 10–14.

Alexander, P. A. (2004). A model of domain learning: Reinterpreting expertise as a multidimensional, multistage process. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 273–298). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). Are you reading my mind? In *Proceedings of the fifth international conference on learning analytics and knowledge—LAK'15* (pp. 246–254). New York, USA: ACM Press.

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, NJ: Princeton University Press.

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd annual meeting of the association for computational linguistics* (Volume 1: Long Papers) (Vol. 1, pp. 238–247).

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). Learning and transfer. In *How people learn: Brain, mind, experience, and school* (pp. 31–78). Washington, DC: National Academy Press.

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.

Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 121–130). Cambridge: Cambridge University Press.

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., Draper, D., & Land, S. M. (2011). An automated measure of group knowledge structure convergence. *Annual meeting of the AECT international convention*.

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.

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, MA: Cambridge University Press.

Coronges, K. A., Stacy, A. W., & Valente, T. W. (2007). Structural comparison of cognitive associative networks in two populations. *Journal of Applied Social Psychology, 37*(9), 2097–2129.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods, 48*(4), 1227–1237. https://doi.org/10.3758/s13428-015-0651-7.

Crossley, S. A., & McNamara, D. S. (Eds.). (2016). *Adaptive educational technologies for literacy instruction*. Routledge: Taylor & Francis.

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.

Duke, N. K., & Pearson, P. D. (2009). Effective practices for developing reading comprehension. *Journal of Education, 189*(1–2), 107–122.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58.

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.

Fonger, N. L., Stephens, A., Blanton, M., Isler, I., Knuth, E., & Gardiner, A. M. (2018). Developing a learning progression for curriculum, instruction, and student learning: An example from mathematics education. *Cognition and Instruction, 36*(1), 30–55.

Gao, Y., Warner, A., & Passonneau, R. J. (2019). Pyreval: An automated method for summary content analysis. In H. Isahara, B. Maegaard, S. Piperidis, C. Cieri, T. Declerck, K. Hasida, H. Mazo, K. Choukri, S. Goggi, J. Mariani, A. Moreno, N. Calzolari, J. Odijk, et al. (Eds.), *LREC 2018—11th international conference on language resources and evaluation* (pp. 3234–3239). (LREC 2018—11th international conference on language resources and evaluation). European Language Resources Association (ELRA).

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, UK: Psychology Press.

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (2001). *The analogical mind*. Cambridge, MA: MIT Press.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65,* 263–297.

Gijbels, D., Dochy, F., den Bossche, P. V., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research, 75*(1), 27–61.

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., et al. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist, 51*(2), 219–246.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371–395.

Graham, S., MacArthur, C. A., & Fitzgerald, J. (Eds.). (2013). *Best practices in writing instruction*. NY: Guilford Press.

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

Hansen, D., Schneiderman, B., & Smith, M. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. Burlington, MA: Morgan Kaufmann.

Hay, D. B., & Kinchin, I. M. (2006). Using concept maps to reveal conceptual typologies. *Education + Training, 48*(2/3), 127–142. https://doi.org/10.1108/00400910610651764.

He, Y., Hui, S. C., & Quan, T. T. (2009). Automatic summary assessment for intelligent tutoring systems. *Computers & Education, 53*(3), 890–899.

Helbig, H. (2006). *Knowledge representation and the semantics of natural language*. Berlin: Springer.

Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language, 69*(2), 151–164. https://doi.org/10.1016/j.jml.2013.03.002.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15,* 332–340.

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

Ifenthaler, D. (2014). AKOVIA: Automated Knowledge Visualization and Assessment. *Technology, Knowledge and Learning, 19*(1–2), 241–248. https://doi.org/10.1007/s10758-014-9224-6.

Ifenthaler, D., Masduki, I., & Seel, N. M. (2011). The mystery of cognitive structure and how we can detect it: tracking the development of cognitive structures over time. *Instructional Science*, *39*(1), 41–61.

Ifenthaler, D., & Pirnay-Dummer, P. (2014). Model-based tools for knowledge assessment. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of **r**esearch on **e**ducational **c**ommunications and **t**echnology* (pp. 289–301). Retrieved from http://link.springer.com/chapter/10.1007/978-1-4614-3185-5_23.

Johnson, T. E., Pirnay-Dummer, P., Ifenthaler, D., Mendenhall, A., Karaman, S., & Tennenbaum, G. (2011). Text summaries or concept maps: Which better represents reading text conceptualization? *Technology, Instruction, Cognition and Learning, 8*(3–4), 297–312.

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

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

Jonassen, D. H., & Cho, Y. H. (2008). Externalizing mental models with mindtools. In D. Ifenthaler, P. Pirnay-Dummer, & J. M. Spector (Eds.), *Understanding models for learning and instruction. Essays in Honor of Norbert M. Seel* (pp. 145–160). New York: Springer.

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

Kim, M. (2012). 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. https://doi.org/10.1007/s11423-012-9247-4.

Kim, M. (2013). Concept map engineering: Methods and tools based on the semantic relation approach. *Educational Technology Research and Development, 61*(6), 951–978. https://doi.org/10.1007/s11423-013-9316-3.

Kim, M. (2015). Models of learning progress in solving complex problems: Expertise development in teaching and learning. *Contemporary Educational Psychology, 42,* 1–16. https://doi.org/10.1016/j.cedpsych.2015.03.005.

Kim, K. (2017). Graphical interface of knowledge structure: A web-based research tool for representing knowledge structure in text. *Technology Knowledge and Learning*. https://doi.org/10.1007/s10758-017-9321-4.

Kim, K. (2018). An automatic measure of cross-language text structures. *Technology, Knowledge and Learning, 23*(2), 301–314.

Kim, K., & Clariana, R. B. (2017). Text signals influence second language expository text comprehension: Knowledge structure analysis. *Educational Technology Research and Development, 65*(4), 909–930.

Kim, K., Clarianay, R. B., & Kim, Y. (2018). Automatic representation of knowledge structure: enhancing learning through knowledge structure reflection in an online course. *Educational Technology Research and Development, 67*(1), 105–122.

Kim, M., Gaul, C., Kim, S., & Madathany, R. (2019). Advance in detecting key concepts as an expert model: Using Student Mental Model Analyzer for Research and Teaching (SMART). *Technology, Knowledge and Learning, 25*(4), 953–976.

Kim, M., & McCarthy, K. S. (2020). Improving summary writing through formative feedback in a technology-enhanced learning environment. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.12516.

Kim, M., Zouaq, A., & Kim, S. (2016). Automatic detection of expert models: The exploration of expert modeling methods applicable to technology-based assessment and instruction. *Computers & Education, 101,* 55–69. https://doi.org/10.1016/j.compedu.2016.05.007.

Kinchin, I. M. (2008). The qualitative analysis of concept maps: Some unforeseen consequence and emerging opportunities. In A. J. Cañas, P. Reiska, M. Ahlberg, & J. D. Novak (Eds.), *Proceedings of the 3rd international conference on concept mapping*. Tallinn, Estonia & Helsinki, Finland.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163–182.

Kintsch, W. (1998). The representation of knowledge in minds and machines. *International Journal of Psychology, 33*(6), 411–420.

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

Koszalka, T. A., & Epling, J. (2010). A methodology for assessing elicitation of knowledge in complex domains: identifying conceptual representations of Ill-structured problems in medical diagnosis. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 311–334). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4419-5662-0_16.

Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior Research Methods, 50*(3), 1030–1046.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211–240.

León, J. A., Olmos, R., Escudero, I., Cañas, J. J., & Salmerón, L. (2006). Assessing short summaries with human judgments procedure and latent semantic analysis in narrative and expository texts. *Behavior Research Methods, 38*(4), 616–627.

Li, H., Cai, Z., & Graesser, A. C. (2018). Computerized summary scoring: crowdsourcing-based latent semantic analysis. *Behavior Research Methods, 50*(5), 2144–2161.

Li, H., Gobert, J., & Dickler, R. (2017). Automated Assessment for Scientific Explanations in On-Line Science Inquiry. *Proceedings of the 10th international conference on educational data mining, EDM 2017* (pp. 214–219). Wuhan, China

Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. *Text Summarization Branches Out*, 74–81. Retrieved from https://www.aclweb.org/anthology/W04-1013.

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

McNamara, D. S., Allen, L., Crossley, S., Dascalu, M., & Perret, C. A. (2017). Natural language processing and learning analytics. *Handbook of Learning Analytics*, 93–104.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.

Nadolski, R. J., & Hummel, H. G. (2017). Retrospective cognitive feedback for progress monitoring in serious games. *British Journal of Educational Technology, 48*(6), 1368–1379.

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.

Newman, M. (2010). *Network: An introduction*. New York: Oxford University Press.

Norman, D. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 7–14). Hillsdale, NJ: Erlbaum.

Passonneau, R. J., Poddar, A., Gite, G., Krivokapic, A., Yang, Q., & Perin, D. (2018). Wise crowd content assessment and educational rubrics. *International Journal of Artificial Intelligence in Education, 28*(1), 29–55.

Pirnay-Dummer, P., & Ifenthaler, D. (2011). Reading guided by automated graphical representations: How model-based text visualizations facilitate learning in reading comprehension tasks. *Instructional Science, 39*(6), 901–919. https://doi.org/10.1007/s11251-010-9153-2.

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

Rumelhart, D. E., & Norman, D. A. (1978). Accretion, tuning and restructuring: Three model of learning. In R. L. Klatzky & J. W. Cotton (Eds.), *Semantic factors in cognition* (pp. 37–53). Hillsdale, NJ: Lawrence Erlbaum.

Rupp, A. A., Sweet, S., & Choi, Y. (2010). Modeling learning trajectories with epistemic network analysis: A simulation-based investigation of a novel analytic method for epistemic games. *Presented at the annual meeting of the International Society for Educational Data Mining (EDM)*, Pittsburgh, PA.

Rus, V., Lintean, M., Banjade, R., Niraula, N. B., & Stefanescu, D. (2013, August). Semilar: The semantic similarity toolkit. In Butt, M., & Hussain, S. (Eds.), *Proceedings of the 51st annual meeting of the association for computational linguistics: system demonstrations* (pp. 163–168). Sofia, Bulgaria.

Santos Jr, E., Mohamed, A. A., & Zhao, Q. (2004). Automatic evaluation of summaries using document graphs. *Proceedings of ACL 2004 workshop on text summarization branches out* (pp 66–73). Barcelona, Spain.

Schvaneveldt, R. W., & Cohen, T. A. (2010). Abductive Reasoning and Similarity: Some Computational Tools. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-Based Diagnostics and Systematic Analysis of Knowledge* (pp. 189–211). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4419-5662-0_11.

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., & Cooke, N. M. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies, 23,* 699–728.

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.

Segers, M. (1997). An alternative for assessing problem-solving skills: The overall test. *Studies in Educational Evaluation, 23*(4), 373–398.

Singer, L. M., & Alexander, P. A. (2017). Reading across mediums: Effects of reading digital and print texts on comprehension and calibration. *The Journal of Experimental Education, 85*(1), 155–172.

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.

Spector, J. M. (2008). Cognition and learning in the digital age: Promising research and practice. *Computers in Human Behavior, 24*(2), 249–262.

Spector, J. M. (2010). Mental representations and their analysis: An epistemological perspective. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 27–40). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4419-5662-0_3.

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.

Stevens, E. A., Park, S., & Vaughn, S. (2019). A review of summarizing and main idea interventions for struggling readers in Grades 3 through 12: 1978–2016. *Remedial and Special Education, 40*(3), 131–149.

Strobl, C., Ailhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A., et al. (2019). Digital support for academic writing: A review of technologies and pedagogies. *Computers & Education, 131,* 33–48.

Sung, Y.-T., Liao, C.-N., Chang, T.-H., Chen, C.-L., & Chang, K.-E. (2016). The effect of online summary assessment and feedback system on the summary writing on 6th graders: The LSA-based technique. *Computers & Education, 95,* 1–18.

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research, 37,* 141–188.

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

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Wade-Stein, D., & Kintsch, E. (2004). Summary street: Interactive computer support for writing. *Cognition and Instruction, 22*(3), 333–362.

Wasserman, S., & Faust, K. (1994). *Social network analysis: methods and applications*. Cambridge: Cambridge University Press.

Westby, C., Culatta, B., Lawrence, B., & Hall-Kenyon, K. (2010). Summarizing expository texts. *Topics in Language Disorders, 30*(4), 275–287.

Yin, Y., Vanides, J., Ruiz-Primo, M. A., Ayala, C. C., & Shavelson, R. J. (2005). Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use. *Journal of Research in Science Teaching, 42*(2), 166–184.

Zimmerman, W. A., Kang, H. B., Kim, K., Gao, M., Johnson, G., Clariana, R., et al. (2018). Computer-automated approach for scoring short essays in an introductory statistics course. *Journal of Statistics Education, 26*(1), 40–47.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Min Kyu Kim is an assistant professor of Learning Technologies in the Department of Learning Sciences at Georgia State University. He is committed to exploring new diagnostic models of learning progression and design solutions to create innovative technology-rich learning experiences.

Kathryn S. McCarthy is an Assistant Professor of Educational Psychology in the Department of Learning Sciences at Georgia State University. Her work examines the cognitive processes involved in reading comprehension and how these processes vary across disciplines. She also explores how educational technologies can support learning from text.