



Automatic representation of knowledge structure: enhancing learning through knowledge structure reflection in an online course

Kyung Kim¹ · Roy B. Clarianay² · Yanghee Kim¹

Published online: 24 September 2018

© Association for Educational Communications and Technology 2018

Abstract

Summary writing is an important skill that students use throughout their academic careers, writing supports reading and vocabulary skills as well as the acquisition of content knowledge. This exploratory and development-oriented investigation appraises the recently released online writing system, *Graphical Interface of Knowledge Structure (GIKS)* that provides structural feedback of students' essays as network graphs for reflection and revision. Is the quality of students' summary essays better with *GIKS* relative to some other common approaches? Using the learning materials, treatments, and procedure of a dissertation by Sarwar (Doctoral Thesis, University of Ottawa, 2012) but adapted for this setting, over a three-week period Grade 10 students ($n = 180$) read one of three physics lesson texts each week, wrote a summary essay of it, and then immediately received one of three counterbalanced treatments including reflection with *GIKS*, solving physics problems as multiple-choice questions, and viewing video information, and finally students rewrote the summary essay. All three treatments showed pre-to-post essay improvement in the central concepts subgraph structure that almost exactly matched the results obtained in the previous dissertation. *GIKS* with reflection obtained the largest improvement due to the largest increase in relevant links and the largest decrease for irrelevant links. The different treatments led to different knowledge structures in a regular way. These findings confirm those of Sarwar (2012) and support the use of *GIKS* as immediate focused formative feedback that supports summary writing in online settings.

Keywords Knowledge structure · Reflection · Writing · *GIKS* · Feedback

✉ Kyung Kim
kkim2@niu.edu

Roy B. Clarianay
rbc4@psu.edu

Yanghee Kim
ykim9@niu.edu

¹ Educational Technology, Research and Assessment, Northern Illinois University, DeKalb, IL, USA

² Learning, Design, and Technology, The Pennsylvania State University, State College, PA, USA

Introduction

A primary goal of secondary education is to promote meaningful learning through the delivery of core academic content in effective and engaging ways that allow students to understand, apply, and transfer what they have learned. Writing-to-learn is one well-established approach that is widely used in secondary education that supports deeper understanding of conceptual knowledge of a given course content through various kinds of writing activities, such as summary writing, reflection essays, and discussion (Graham and Hebert 2010). The literature clearly shows that formative feedback during writing-to-learn activities is critical; however, it is very labor intensive for teachers to provide timely formative feedback on students' writings. To address this need, we present a novel browser-based automatic writing assessment tool, *Graphical Interface of Knowledge Structure (GIKS)* that generates formative visual feedback as network graphs to support the acquisition and development of conceptual knowledge in writing-to-learn activities. Is the quality of students' summary essays better with *GIKS* relative to other common approaches?

Reading, writing, and learning

Reading, writing, and learning are intimately related. Following Emig's (1977) view, writing, thinking, and learning are closely interrelated, and writing is a concrete artifact of cognition. In a meta-analysis of the effects of frequent "writing to learn" with or without feedback, Bangert-Drowns et al. (2004) said,

... learning entails active, personal, and self-regulated construction of organized conceptual associations, refined by feedback processes. The same features characterize writing. Writing requires the active organization of personal understandings. The externalization of those understandings in symbolic form makes them available for feedback in self-reflection and revision, in review of a record of the evolution of ideas and understanding, and in documentation for public discourse. (p. 29)

Bangert-Drowns et al. (2004) reported that for 47 studies that compared the effects of school-based writing with and without feedback, feedback was more effective for academic achievement (e.g., domain content knowledge) with an effect size of 0.32; and for 46 studies that used writing with metacognitive reflection or not, metacognitive reflection was more effective for content knowledge with an effect size of 0.44.

A meta-analysis by Graham and Hebert (2010) supports the findings of Bangert-Drowns et al. (2004) and extends those findings beyond lesson content knowledge to consider the effects of writing on general reading comprehension. Among other findings, Graham and Hebert reported an effect size of 0.52 on experimenter-designed reading comprehension tests in 19 studies for those students who wrote text summaries compared to other treatment interventions that included reading only, re-reading the passage, reading and studying, or receiving reading instruction. They reported an effect size of 0.77 on experimenter-designed reading comprehension tests across 9 studies in favor of metacognitive reflection (i.e., categorized as personalization, analysis, or interpretation) compared to reading the text, reading and rereading it, reading and studying it, reading and discussing it, or receiving reading instruction. So combining reading and summary writing with feedback and reflection has been demonstrated to substantially improve learning course content and even general reading comprehension skills.

Since essays are time consuming for teachers to review and grade, will teachers use writing? Kihuhara et al. (2009) in a survey of 350 high school science teachers reported that about half of these teachers used summary writing or responding to writing assignments at least weekly, and this proportion is likely larger for online courses where writing is a primary mode of interaction. Thus, asking students to complete written assignments is compatible with what many teachers already do. Online tools to make summary writing easier and more frequent then are an important potential addition to any learning setting (Mørch et al. 2017). This investigation seeks to explore and further develop one such online tool, *GIKS*.

Knowledge structure as a way to envision learning

Structure is a characteristic of domain content knowledge elements and of individual personal mental models of that content (Johnson-Laird 2004; Tripto et al. 2018). An important goal of learning from science texts is to arrive at a level of conception that reflects the underlying domain-specific knowledge structure intended by the author or writer (e.g., Van Dijk and Kintsch 1983; Jonassen et al. 1993). Knowledge structure (KS) here is defined as the organization of domain concepts stored in long-term memory (Clariana 2010; Ifenthaler 2010; Jonassen et al. 1993). Eliciting and then representing these domain structures and individuals' mental structures are a way to more fully understand the nature of knowledge in human cognition relative to that of others and of the domain. The KS measures used in this current investigation have been shown to correlate with STEM-content reading comprehension, for example, in monolingual settings using English (Clariana et al. 2014), Dutch (Fesel et al. 2015), and Mandarin (Su and Hung 2010); and also in bilingual settings with participants who are Dutch-English (Mun 2015), Chinese-English (Tang and Clariana 2017), and Korean-English (Kim and Clariana 2015, 2017, 2018).

So how do readers derive an internal mental representation based on their interpretation of a text? During reading, readers build an individualized situation model which is their structured mental model consisting of their background knowledge (general and specific) and the new information attained from the text (Van Dijk and Kintsch 1983). In this perspective, the reader's post-reading situation model may or may not match the author's situation model reflected in the text, but for expository science texts, the reader's ability to build the specifically structured situation model that the author intended is a key to successful STEM-content reading comprehension (Van Dijk and Kintsch 1983; Fesel et al. 2015; Zwaan and Radvansky 1998).

Measuring knowledge structure comprehension and misconceptions

There are many techniques to identify learners' understanding such as interviews, open-ended questions, multiple-choice tests, and other domain-specific tests (see for review, DiCerbo 2007). Each of these has different strengths and weaknesses in terms of the information produced, the time and effort required, and so on. Because it is moderately demanding to measure learners' missing conceptions, errors, and misconceptions (Coştu and Ayas 2005), in many cases, the reasons underlying learners' wrong answers have to be implied, assumed, or overlooked. Domain-specific KS measures provide a straightforward way to identify missing information, errors, misconceptions, and domain normative correct conceptions all at once.

Of the many ways available to measure KS, Ifenthaler et al. (2010) argued that using natural language responses, especially students' essays, are likely to most accurately tap

into student's internal mental structure and as such, essays are a manifestation of cognition and a concrete artifact (e.g., Emig 1977; Mørch et al. 2017). This current investigation applies the well-established *ALA-Reader* approach (Koul et al. 2005) embedded in *GIKS* for representing and comparing KS inherent in essays as node-link network graphs.

Knowledge structure and *ALA-Reader*

The Analysis of Lexical Aggregates-Reader (*ALA-Reader*) approach captures the sequence of important key terms in a text as links in an $n \times n$ array (an adjacency matrix proximity file; see Clariana et al. 2014; Kim 2012, for validity of *ALA-Reader*). *Pathfinder KNOT* is a computer-based data-reduction network scaling technique that seeks the most direct path between nodes, the *Pathfinder* algorithm eliminates much of the complexity, or visual noise, of the original network (see Sarwar 2012, for the validity of the *Pathfinder* algorithm for this purpose). *Pathfinder* is applied to (a) convert the raw proximity data from *ALA-Reader* into a *Pathfinder* network and then to (b) compare these networks with a benchmark referent network derived in the same way from the lesson texts or other referents.

The *ALA-Reader* approach has been used in diverse domains within and also across languages, for example to determine individual and group knowledge convergence in an online course (Draper 2013; Tawfik et al. 2018), to determine lesson text structures for comparison to student-readers' KS (Clariana et al. 2014; Fesel et al. 2015; Kim 2018), to describe KS transfer from first language to second language (Kim and Clariana 2015; Kim 2017a), and to understand neural mechanisms underlying reading comprehension by comparing KS network patterns with brain network patterns (Li and Clariana 2018).

The *ALA-Reader* approach has been favorably compared to other text assessment technologies, including global-scale (implicit) semantic space models such as Hyperspace Analogue of Language and also Latent Semantic Analysis (HAL and *LSA*; e.g., Su and Hung 2010) and also to local-scale (explicit) semantic models such as Text-Model Inspection Trace of Concepts and Relations (T-MITOCAR) and Concept Map Modeling (CMM) (e.g., Ifenthaler et al. 2010; Kim 2012). Kim (2012) reported that "*ALA-Reader* had a higher correlation with the manually constructed benchmark model than did T-MITOCAR" (p. 703). Koul et al. (2005) compared *ALA-Reader* to *LSA*, *ALA-Reader* essay scores had a larger correlation with the 11 human-rater scores than did the *LSA* essay scores, e.g., *ALA-Reader* Pearson $r = .72$ vs. *LSA* $r = 0.61$. Similarly, Su and Hung (2010) reported that native Mandarin students' essay scores from *ALA-Reader* outperformed *LSA* compared to keyword pattern matching, e.g., *ALA-Reader* $r = 0.67$ vs. *LSA* $r = 0.34$; while the two scoring approaches were equivalent compared to two human rater scores, e.g., *ALA-Reader* $r = 0.63$ vs. *LSA* $r = 0.62$.

GIKS iterative development

In previous research, *ALA-Reader* is described as a Windows OS-based standalone essay analysis tool designed for researchers (Clariana et al. 2009; Clariana et al. 2014), but actually *ALA-Reader* is a text analysis algorithm that translates text into an adjacency matrix (i.e., a symmetrical $n \times n$ array, with n as the number of key terms included) that can be embedded in other tools (e.g., Su and Hung, 2010). *GIKS* (<http://giks.herokuapp.com/>) is a recently developed browser-based writing support tool designed especially for teachers based on the *ALA-Reader* algorithm (Kim 2017b; Zimmerman et al. 2018). Currently, the *GIKS* student interface has two screens, the first screen provides content as

any HTML 5 source (text, video, etc.), a writing prompt, and a submit text box (with spell checking); the second screen that immediately appears after pressing Submit provides a network graph referent at the top left called the Master, and a network graph of the student's essay at the top right. An essay percent score based on the relationship between the student and referent networks, and a text table of these results is shown below the two networks. The individualized KS network graph uses different colored link lines to show similarity (green) and differences (red for incorrect, yellow for missing) of a student's essay network compared to a referent KS network (see Fig. 1). Clicking the buttons at the bottom of the student's network (Your network, Missing Link/Node, Incorrect Links) hides all other links and just reveals the incorrect links (red) or the missing links (yellow).

As with *GIKS*, Gogus (2013) notes that T-MITCAR is designed to immediately provide a cutaway visual network to allow comparisons between participants' and a referent representation (e.g., the expert's solution). Gogus (2013) used this kind of visual network representation to provide feedback to learners after the experiment was done (p. 180), but not immediately after writing, as in this current investigation.

For *GIKS* to effectively provide immediate formative feedback during the writing process, there are a number of design issues yet to work out. Most critically for this current investigation, as noted above, *GIKS* allows for immediate visual side-by-side examination and comparison between the student and expert network graphs. During the initial design and ongoing iterative development of *GIKS*, a base assumption has been that side-by-side comparison of an entire portion of domain-specific content is best, but this immediately became an issue in pilots due to screen size limitations when displaying large network graphs, the networks can be extensive and complex. The simplest solution was to adjust the chunk size of domain content so that the referent network graph would have no more than

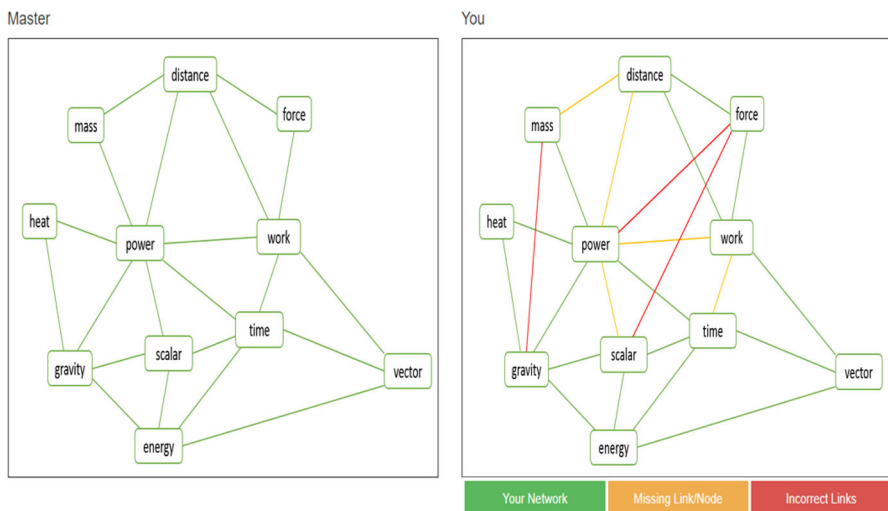


Fig. 1 *GIKS* screen capture of KS network graphs derived from lesson text “Work, Energy, and Power” (left), and a student summary essay (right). Students' KS network graphs include highlights showing the similarity and difference compared to the referent lesson text KS; yellow indicates 'missing' links (perhaps due to a lack of understanding of the specific relationships between the key concepts) and red indicates 'incorrect' links (perhaps due to a misunderstanding of the relationship between the key concepts) (Color figure online)

about 15 concepts, which means that the *GIKS* writing prompt must be set so that the students' essays would be about 500 words long (i.e., 3–5 paragraphs). This content chunk size allows for reasonably sized writing activities from the teachers' and students' perspectives, provides networks that are easier to inspect, and also this way aligns with both concept map research and also the *ALA-Reader* research so far.

Focused structural feedback

A meta-analysis of 55 studies by Nesbit and Adesope (2006) considered the influence of visual representations of central ideas compared to peripheral details, the effect sizes for central ideas were larger (0.59 versus 0.20, p. 433) suggesting that side-by-side network comparison of subgraphs of just the important central ideas are one way to reduce the visual complexity of the networks but retain their effectiveness.

In concert with our view of domain-specific KS, Trumpower and Sarwar (2010) have coined the term “structural feedback” for this kind of side-by-side comparison of an individual's KS network with a referent network. They used a network subgraph structural feedback approach is aimed at the central critical concepts so that the structural feedback is designed to alter the core structure of an individual's KS. They note that so far “very little research has been devoted to determining the effect of feedback on structural knowledge development” (Trumpower and Sarwar 2010, p. 9).

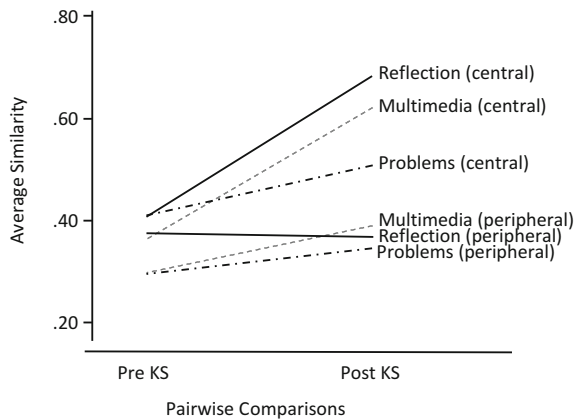
In a dissertation by Sarwar (2012), high school students learned physics concepts in the traditional way. Then students were provided with one of three quite diverse remediation treatments that were all created to influence the central concept in the referent concept maps of the content. The three treatments included (a) structural feedback that asked students to *reflect* on their individualized *Pathfinder* network derived from individuals' previous collected pairwise term associations compared to a benchmark referent central concept network subgraph, (b) *solving problems* presented as multiple-choice questions that separately address the same central concept relationships, or (c) *viewing a multimedia demonstration* of the same central concept relationships. In order to measure whether focusing on central concepts negatively affects other concepts, peripheral concepts were also identified for each referent network; peripheral concepts are least related to the central concept and hence should be least influenced by the structural feedback treatments.

In his dissertation study, pre-to-post Pathfinder network similarity as Cohen d effect sizes for the *central concepts* were: reflection, $d = 1.42 >$ view multimedia, $d = 1.39 >$ solve problems, $d = 0.53$ and this order is reversed for *peripheral concepts*: view multimedia, $d = 0.41 >$ solve problems, $d = 0.22 >$ reflection, $d = -0.05$ (p. 85, Table 3.3, Sarwar 2012). A clear interaction was observed, the central concept relations improved with all three treatments (significant) while the peripheral concepts were in the reverse order (though not significant). The greatest improvements were observed when students reflected on their own knowledge structure alongside the referent subgraph knowledge structure (see Fig. 2).

Purpose

Although text analysis systems such as *T-MITCAR* and *GIKS* can immediately provide network graphs of essays, there has been little research examining the optimal design related to using these network graphs to support summary writing, and whether such graphs actually help. The purpose of this exploratory and development-oriented

Fig. 2 Pre and post knowledge structure (i.e., essay KS) mean network similarity for central and peripheral concepts for the three treatments (from Sarwar 2012, p. 85)



investigation is to consider one feature of *GIKS*, structural feedback as network *subgraphs*. Is the quality of students' summary essays better with *GIKS* relative to these other common approaches? The investigation will also confirm and extend Sarwar's (2012) dissertation findings by examining the relative effectiveness of the new *GIKS* tool through using the same materials and treatments used in that previous study (adapted), but with *GIKS* providing the network graphs for immediate structural feedback for reflection. Findings should contribute to an improved understand of structural feedback relative to these other approaches and will support the iterative design of the *GIKS* tool for summary writing.

Method

Participants

The participants were 180 students from low-level Grade 10 physics online courses, pre-classified by the After-School Online Home Learning System based on their physics scores. This system is for foreign K-12 students living in Korea, funded by the Korean Ministry of Education. All participants were native English speakers, typically children of foreigners working in embassy and foreign companies in Korea, and also of native Koreans who had lived in a foreign country for more than 10 years (aged from 17 to 18; male, 53%). After a briefing of the requirements of the investigation, volunteer participants signed the consent form to further individually assent to the consent form previously signed by their parents. They received course credits for their participation.

Procedure and materials

This investigation was conducted over a three-week period in May that was the last month of the school term, participants had already learned this course content to some degree. For the investigation, each week the participants accessed the online system with their assigned individual confidential ID number and read one of the three lesson texts in an individually pre-assigned order, either "Work, Energy, and Power", "Motion", or "Nature of Wave" (from Sarwar 2012), and then wrote and submitted an initial summary of that text (*before-feedback KS*). Next, one of three different types of structural feedback were immediately provided, either "KS reflection", "multiple-choice questions", or "video instruction"

(discussed in detail below). These three structural feedback treatments parallel those used by Sarwar (2012) who stated, “The techniques selected for this study, i.e., reflection, written exercises, and multimedia instruction act like an umbrella for other types of instructions” (p. 33). These three treatments are familiar to students and teachers, are fairly common in schools, and can be easily delivered online. Over the course of the investigation, all students read all three texts and received all three types of structural feedback treatments, the order of text and feedback types was fully counterbalanced (see Table 1). After the structural feedback treatment, students were asked again to write and submit a final summary of that lesson text as a measure of their *after-feedback* KS. These *before-feedback* and *after-feedback* KSs were compared to the lesson text KS network graph they read to identify improvements, if any, in the students’ domain-specific KS.

Comparison measures

For the analysis of students’ KS before and after feedback, following Sarwar (2012), this investigation compared students’ KS to a highly specific referent KS that purposefully focused on just the central critical concept in each text and its relations to the other terms. These are referred to here as *central concepts* versus *peripheral concepts* (explained in detail below). Any increase in similarity of students’ KS to the referent KS after structural feedback was considered as an indication of the effectiveness of the feedback, because if the student-readers comprehend the science texts as the author/expert intended, then the author’s KS would be reflected in the readers’ KS (e.g., Clariana et al. 2014; Fesel et al. 2015).

To identify all key terms for each text, five instructors reviewed all of the materials and then negotiated together to identify 11 terms for each lesson topic. Following the approach by Sarwar (2012), one *central concept* (targeted concept) and one *peripheral concept* (a selected non-targeted concept) were selected from each referent KS network graph of the three lesson texts (see for example, Fig. 3). To do this, the selected 11 key terms were used to generate a KS network of each text. Then the central and the peripheral concepts were chosen by the five instructors mainly based on node degree (links) as a measure of concept importance (Clariana et al. 2013) and the position of the concepts in the referent KS network graphs. All the structural feedback treatments (*GIKS*, MC questions, and video instruction) were designed to improve the KS related to these central concepts (see Figs. 3 and 4). Following Sarwar’s (2012) approach, to observe potential broader effects of the treatments, a peripheral concept was chosen for analysis along with the central concept. A peripheral concept is one that is least related to the central concept in the referent KS network graph and so it was assumed that the peripheral concept would receive the least or no effect of the structural feedback, since the feedback treatments are specifically designed and intended to influence the central concepts and their closely related concepts.

Table 1 The design of this investigation

	Work, energy, and power	Motion	Nature of wave
Week 1	Reflection (<i>GIKS</i>)	Video instruction	MC questions
Week 2	MC questions	Reflection (<i>GIKS</i>)	Video instruction
Week 3	Video instruction	MC questions	Reflection (<i>GIKS</i>)

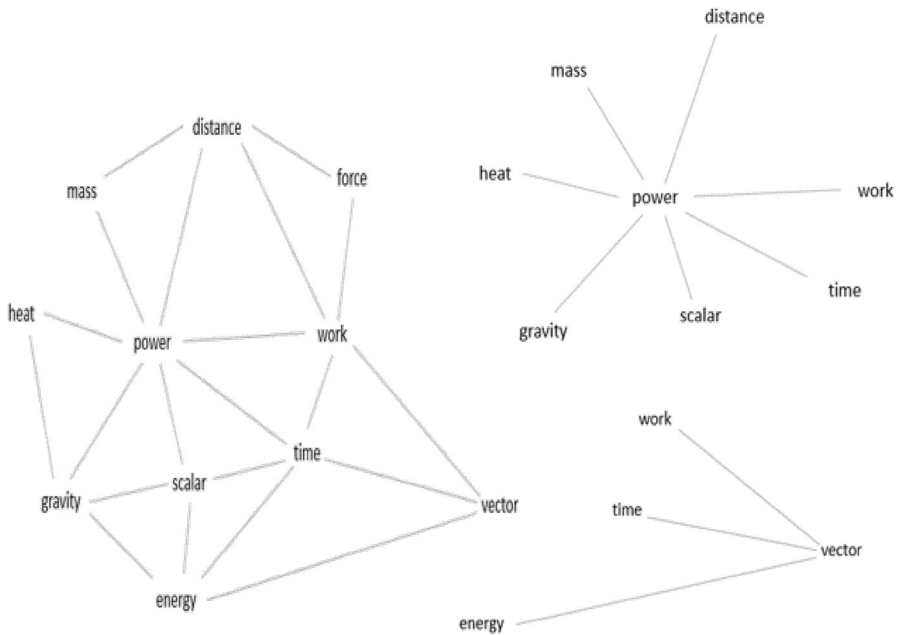


Fig. 3 The referent KS of the lesson text (left), the central concept of “power” from the referent KS (top right), and peripheral concept of “vector” from the referent KS (bottom right)

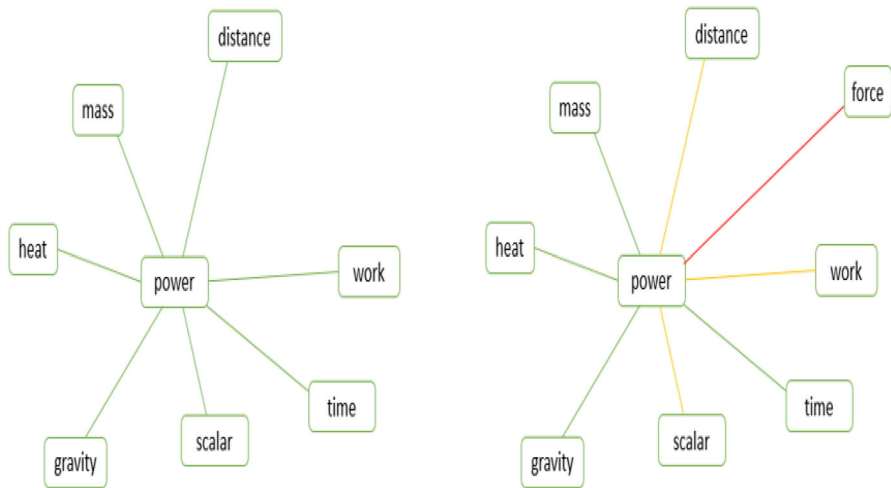


Fig. 4 Example KS reflection regarding the central concept of “power” from the referent KS (left) and a student’s essay KS of the central concept (right) (Color figure online)

Experiment conditions

KS reflection treatment Under the KS reflection treatment (i.e., *GIKS*), after submitting their *before-feedback* summary essays of each lesson text to the *GIKS* system embedded in

the online course, students were immediately shown the expert referent KS network side-by-side with their own KS essay network. Students were asked to compare their own KS with the referent KS, focusing on any discrepancies such as missing links that the referent KS had but they did not (in yellow) and incorrect links that they had but the referent KS did not (in red) and then reflect on their writing based on these comparisons (see Fig. 4).

The referent KS network subgraph in Fig. 4 shows that the central concept, *power*, is related to seven other concepts, *distance*, *mass*, *heat*, *gravity*, *scalar*, *time*, and *work* while the student network subgraph links the concept of *power* with only *mass*, *heat*, *gravity*, and *time* (missing the links of *power* with *distance*, *scalar*, and *work*, and incorrectly links the concept of *power* with *force*) as compared to *power* concept in the referent KS. After reflection, the students were asked to rewrite and resubmit their summary essay as their *after-feedback* essays.

Solving problems treatment Under the solving problems treatment, after submitting their *before-feedback* summary essays of each lesson text to the *GIKS* system embedded in the online course, students were immediately given typical physics problems to solve, in the form of multiple-choice items. The problems were specifically designed to relinquish missing and incorrect conceptions about the central concepts. These MC questions taken from Sarwar (2012) were designed around the same central target concepts to provide the opportunity for students to further establish their knowledge of the central concepts and of the correct relationships with other concepts.

For example, here are three of the problems regarding the *power-work* relationship: (1) a mass of 4 kg is raised vertically a *distance* of 2 m in 5 s. Calculate (a) the *work* done in raising the mass, (b) the average *power* required. (2) The rate at which *work* is done is called (a) *power*, (b) *scalar*, (c) *energy*; and (3) If 100 Joules of *work* was done in 10 s, what *power* was used? (a) 1 kW, (b) 10 W. The key terms in the problems were shown in italics to direct the students' attention to these terms. The central concept *power* was related to seven other concepts, *distance*, *mass*, *heat*, *gravity*, *scalar*, *time*, and *work* (compare these to Fig. 4). Students were asked to solve these problems around all seven of the related links (i.e., between power and distance, mass, heat, gravity, scalar, time, and work), consisting of 4–6 questions per central term, and then received correct answer feedback on all questions at the end.

Video instruction treatment Under the video instruction treatment, after submitting their *before-feedback* summary essays in the *GIKS* system embedded in the online course, students immediately viewed the video instruction that was similarly designed to relinquish missing and incorrect conceptions about the central concepts. For example, the central concept *power* was related to seven other concepts so students were provided with video instructions around all seven of the related links (i.e., between power and distance, mass, heat, gravity, scalar, time, and work). Figure 5 below is an example video instruction screen capture of the *power-work* relationship. The video clips, around 5–14 min per relationship principle, were available in the online course and individual students watched the video using their own computer. The five instructors designed these videos for these learners in this setting and so they were especially interested in the impact of the videos.

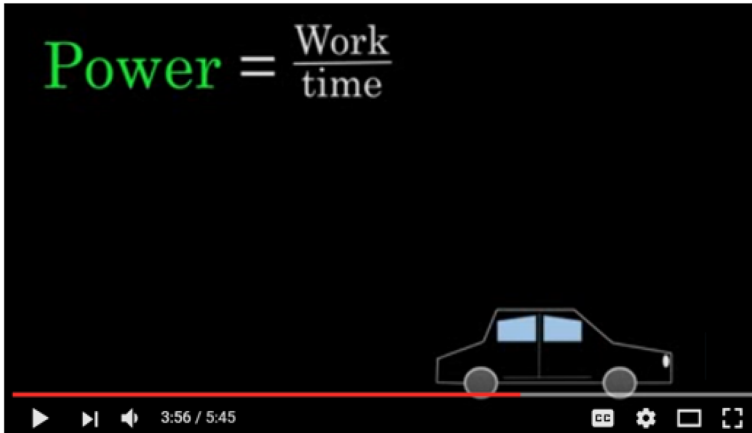


Fig. 5 Example video instruction regarding power–work relationship

Results

Similarity of central and peripheral concepts to the text referent subnetworks

In this investigation, the similarity between students' KS and the lesson text referent KS was calculated as *percent overlap* measured as “links in common” divided by the average number of links in the two KS network graphs (see for detail, Clariana et al. 2014; Kim and Clariana 2015). The descriptive statistics of the analysis are provided in Table 2.

A 3-within-subjects analysis of variance (ANOVA) with type of feedback (reflection, MC questions, and video instruction) \times type of concept (central and peripheral) \times time (pre-intervention and post-intervention) on average similarity data measures (as percent overlap with the text benchmark referent subnetworks) was computed using SPSS (version 20). There were no outliers (assessed by inspection of a boxplot), residuals were normally distributed ($p > 0.05$; assessed by Shapiro–Wilk), and there was homogeneity of variances ($p = 0.061$; assessed by Levene's test). A standard alpha level of 0.05 was used to test for statistical significance in all analyses in this investigation.

There was a significant three-way interaction between type of feedback, type of concept, and time, $F(1, 54) = 6.101$, $p = 0.004$, partial $\eta^2 = 0.881$. There was a significant simple

Table 2 Average similarity of student essay network to lesson text referent subnetwork expressed as percent overlap with the text subnetworks for each treatment by a type of feedback (reflections, MC questions, or video instruction), type of concept (central or peripheral), and time (pre-intervention or post-intervention)

Treatment	Central concepts		Peripheral concepts	
	Pre-KS	Post-KS	Pre-KS	Post-KS
Reflections (<i>GIKS</i>)	0.43 (0.21)	0.75 (0.25)	0.31 (0.21)	0.33 (0.22)
MC questions	0.39 (0.19)	0.50 (0.25)	0.29 (0.26)	0.31 (0.28)
Video instruction	0.44 (0.23)	0.62 (0.23)	0.26 (0.19)	0.37 (0.21)

Standard deviations shown in parentheses

two-way interactions of type of feedback \times time for the central concepts, $F(1, 54) = 8.78$, $p = 0.000$, partial $\eta^2 = 0.557$, and for peripheral concepts, $F(1, 54) = 5.90$, $p = 0.041$, partial $\eta^2 = 0.444$. There was a significant simple main effect of type of feedback at post-intervention phase for central concepts, $F(1, 54) = 13.408$, $p = 0.004$, partial $\eta^2 = 0.546$, and for peripheral concepts, $F(1, 54) = 5.04$, $p = 0.045$, partial $\eta^2 = 0.411$. The significant interaction is shown in Fig. 6, compare this to the previous data from Sarwar (2012) shown in Fig. 2, the two are nearly identical.

Essay network similarity with the lesson texts' network subgraphs reported as Cohen effect sizes for the *Central concept subnetworks* from pre-to-post are (larger is better): solve problems (0.58) < view video (0.78) < KS reflection (1.52) and for the *Peripheral concept subnetworks* (smaller is better) are solve problems (0.08) < KS reflection (0.10) < view video (0.58).

Analysis of relevant and irrelevant links to central concepts

The analysis above considered central and peripheral (control) concepts, but note that central concepts may be subdivided into relevant and irrelevant links. Irrelevant links are not quite errors, but students' essay KS can be improved not only by increasing relevant links but also by decreasing irrelevant links. To consider this, the number of relevant and irrelevant links from pre-to-post intervention by type of central concept were calculated and then compared to the corresponding referent KS as a follow-up analysis. The descriptive statistics of the analysis are summarized in Table 3.

Analysis of the average number of links consisted of a 1—between, 2—within-subjects analysis of variance (ANOVA) with 3 types of essay feedback (reflection, MC questions, or video instruction) \times 2 types of links (relevant, irrelevant) \times 2 time periods (pre-intervention, post-intervention). There were no outliers (assessed by inspection of a boxplot), residuals were normally distributed ($p > 0.05$; assessed by Shapiro–Wilk), and there was homogeneity of variances ($p = 0.071$; assessed by Levene's test).

There was a significant simple main effect of type of feedback at post-intervention phase for relevant links, $F(1, 54) = 9.90$, $p = 0.002$, partial $\eta^2 = 0.630$, and for irrelevant links, $F(1, 54) = 7.315$, $p = 0.002$, partial $\eta^2 = 0.220$. There was a significant simple two-way interactions of type of feedback \times time for the relevant links, $F(1, 54) = 9.78$, $p = 0.000$, partial $\eta^2 = 0.957$, and for irrelevant links, $F(1, 54) = 62.50$, $p = 0.002$, partial $\eta^2 = 0.608$ that is subsumed in the significant three-way interaction observed between type

Fig. 6 The two-way interaction of central concepts (top) and peripheral concepts (bottom)

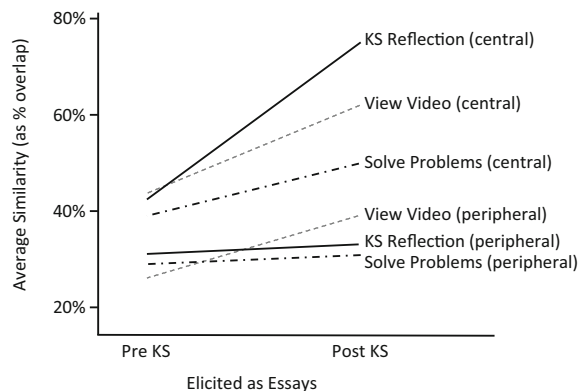
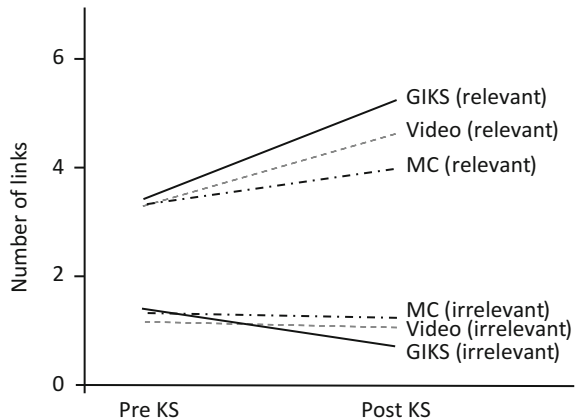


Table 3 Average number of central concept links by treatment (*GIKS* reflections, MC questions, or video instruction), type of links (relevant, irrelevant), and time (pre-intervention, post-intervention)

Treatment	Relevant		Irrelevant	
	Pre-KS	Post-KS	Pre-KS	Post-KS
Reflections (<i>GIKS</i>)	3.43 (1.20)	5.25 (1.25)	1.33 (1.01)	0.67 (1.12)
MC questions	3.31 (1.21)	3.98 (1.05)	1.24 (0.95)	1.15 (1.08)
Video instruction	3.24 (1.26)	4.64 (1.33)	1.12 (0.91)	1.01 (1.11)

Standard deviation shown in parenthesis

Fig. 7 The two-way interaction of central concept relevant links (top) and irrelevant links (bottom)

of feedback, type of link, and time, $F(1, 54) = 7.78$, $p = 0.017$, partial $\eta^2 = 0.540$ (see Fig. 7).

There was a significant *increase* of the number of mean relevant links around central concepts in students' KS from pre-to-post intervention for reflections, $F(1, 54) = 124.92$, $p = 0.000$, partial $\eta^2 = 0.558$ (increased from 3.43 to 5.25), for MC questions, $F(1, 54) = 21.12$, $p = 0.001$, partial $\eta^2 = 0.521$ (increased from 3.31 to 3.98), and for video instruction, $F(1, 54) = 112.11$, $p = 0.000$, partial $\eta^2 = 0.957$ (increased from 3.24 to 5.2). Regarding irrelevant links, there was a significant *decrease* of the number of mean irrelevant links around central concepts in students' KS from pre-to-post intervention only for reflections, $F(1, 54) = 189.414$, $p = 0.025$, partial $\eta^2 = 0.708$ (decreased from 1.33 to 0.67), but not for MC question, $p = 0.092$ or for Video instruction, $p = 0.055$. Pre-to-post students' essay network similarity with the lesson texts' network subgraphs reported as Cohen effect sizes for the *Central concept subnetworks* for *relevant links* are (larger is better): solve problems (0.55) < view video (1.11) < KS reflection (1.52) and for *irrelevant links* are (smaller is better): KS reflection (-0.65) < view video (-0.12) < solve problems (-0.09).

Discussion

This exploratory and development-oriented study compared the *GIKS* online writing support tool with reflection to worked problems as multiple-choice questions and to video demonstration in order to observe the pre-to-post change in students' essay network

structures for these three diverse treatments. Analysis focused on central and peripheral concepts, and on the relevant and irrelevant links to central concepts. All three treatments improved the similarity of students' KS to the text referent central concepts subgraph from pre to post (see Table 2) but KS reflection was best, followed by video instruction, and lastly multiple-choice questions, exactly confirming the findings reported by Sarwar (2012).

The findings of this present study show that the automated and immediate KS representation by the *GIKS* system can help students to improve their individual KS through writing with reflection; reflection can be done well or poorly, but it is always a productive experience (Spector and Koszalka 2004). The relatively different influence of reflection observed here for both central and peripheral concepts could be attributed to the holistic visual nature of the network graphs that were used for reflection. These visual network representations are generative (Osborne and Wittrock 1985) since this comparison and reflection require cognitive activity simultaneously with multiple concepts of the entire nomological network at once.

Solving problems as multiple-choice questions were designed to address specific central principles in the lesson materials. Sarwar (2012) comments that “the ability to solve written exercises involves application of concepts and procedure, surely one would hope that improved problem-solving ability translates into better understanding of concepts” (p. 38). On the other hand, the results observed here for video instruction may be due to “the lack of focus of video instruction on just the specific key concepts” (Sarwar 2012, p. 113), video demonstrations by necessity need to include all of the involved concepts, both central and peripheral. In addition, video is a combined oral/visual medium that tends to be additive rather than accommodative (Ong 1982) and so tends to not disassociate pre-existing concept associations, unless intentionally designed to do so, such as by a refutation text approach (Tippett 2010). Thus with this video treatment, students could obtain relationships beyond the targeted concepts which can account for the increase in both central and peripheral concepts.

Different essay feedback approaches influence what is actually revised on the follow-on essays. For example, Mørch et al. (2017) reported that “The students in the target class (sic using the *EssayCritic* tool) included more ideas (content) in their essays, whereas the students in the comparison class (sic peer feedback on their essays) put more emphasis on the organization of their ideas.” (p. 213). Here the *GIKS* structural feedback network subgraph approach influenced both content addition and organization (a reduction in peripheral information and irrelevant links relative to the other two treatments). Further research studies are needed on these subtle but important influences of different feedback approaches.

We acknowledge the ongoing decades-long debate regarding the impracticable nature of media comparison research (the Clark–Kozma media debate, Clark 1994; Kozma 1994). Solving problems as multiple-choice items and viewing videos were selected here from Sarwar (2012) to represent broad categories of strategies that are actually used in real classrooms and that can be delivered online. All three treatments were designed specifically to influence knowledge structure and were made as equivalent to each other as possible for this specific domain knowledge while maintaining and leveraging their unique attributes. These three types of treatments engage different aspects of learning due to the different cognitive processes involved (e.g., Ozuru et al. 2013). The intent of this investigation was not to show that one treatment is superior to another, but rather that *different strategies lead to different knowledge structures in a regular way*. The findings of this

investigation open a much broader question of how any instructional strategy uniquely influences knowledge structure as a mediating influence on learning.

Development and testing of the *ALA-Reader* algorithm is ongoing. In previous investigations and as applied here embedded in *GIKS*, the algorithm generates a symmetric $n \times n$ link array from natural language. Using central concept network subgraphs as the referent network in this investigation overcomes an inherent limitation of the algorithm to over-connect terms, for example by spanning sentence and paragraph boundaries. Future development and concomitant research should consider whether the algorithm should or should not connect the networks across sentences and paragraphs. Further, given that almost all texts are intended to be read in a linear, sequential, and directional way, then asymmetric link arrays that form directed network graphs may be better than the symmetric arrays used in this investigation (i.e., that form undirected network graphs), and so further study should determine which is better under what circumstances.

Closing thoughts

Information about students' understanding and misunderstanding as KS can promote students' active engagement in the development of their KS during learning (Spector and Koszalka 2004), can help teachers to use improved and individualized instructional strategies (Treagust and Duit 2008), and can also assist instructional designers in creating materials that can guide and help learners overcome their misconceptions and transition toward a desired state or level of increasing expertise (DiCerbo 2007). However, assessment of KS is not widely employed to measure students' learning in schools perhaps due to the lack of familiarity, research base, and especially easy-to-use ways to collect it. As evidenced in this investigation, reflection on KS can be a useful way to improve individual students' knowledge of physics by reducing misconceptions as incorrect links and misplaced emphasis on peripheral ideas (Sarwar and Trumpower 2015). This confirms previous studies that report that mental representation is motivational and instructionally effective across a wide range of interventions, and is especially effective for learners with low ability or low domain knowledge (Nesbit and Adesope 2006).

The *GIKS* tool running in an internet browser can generate real time and at-a-glance KS graphic representation inherent in texts and essays that can be used by students for metacognitive reflection as in this investigation, and also probably by teachers for visual progress monitoring. If fully implemented online and further validated, the *GIKS* system could have wide-ranging application across many content areas and delivery approaches; for example, the *GIKS* has recently been applied to the text generated in online discussion boards (Tawfik et al. 2018) and in an introductory statistics course (Zimmerman et al. 2018). Further iterative development of this approach is warranted.

Acknowledgements Kyung Kim acknowledges support by the Pennsylvania State University's Center for Online Innovation in Learning (Grant No. 05-042-23 UP10010).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*, 29–58.
- 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., Engelmann, T., & Yu, W. (2013). Using centrality of concept maps as a measure of problem space states in computer-supported collaborative problem solving. *Educational Technology Research and Development, 61*(3), 423–442. <https://doi.org/10.1007/s11423-013-9293-6>.
- Clariana, R. B., Wallace, P. E., & Godshalk, V. M. (2009). Deriving and measuring group knowledge structure from essays: The effects of anaphoric reference. *Educational Technology Research and Development, 57*(6), 725–737. <https://doi.org/10.1007/s11423-009-9115-z>.
- Clariana, R. B., Wolfe, M. B., & Kim, K. (2014). The influence of narrative and expository lesson text structures on knowledge structures: Alternate measures of knowledge structure. *Educational Technology Research and Development, 62*(5), 601–616. <https://doi.org/10.1007/s11423-014-9348-3>.
- Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21–29.
- Coştu, B., & Ayas, A. (2005). Evaporation in different liquids: Secondary students' conceptions. *Research in Science & Technological Education, 23*(1), 75–97.
- DiCerbo, K. E. (2007). Knowledge structures of entering computer networking students and their instructors. *Journal of Information Technology Education, 6*(1), 263–277.
- Draper, D. C. (2013). The instructional effects of knowledge-based community of practice learning environment on student achievement and knowledge convergence. *Performance Improvement Quarterly, 25*(4), 67–89. <https://doi.org/10.1002/piq.21132>.
- Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication, 28*(2), 122–128. <https://doi.org/10.2307/356095>.
- Fesel, S. S., Segers, E., Clariana, R. B., & Verhoeven, L. (2015). Quality of children's knowledge representations in digital text comprehension: Evidence from pathfinder networks. *Computers in Human Behavior, 48*, 135–146.
- Gogus, A. (2013). Evaluating mental models in mathematics: A comparison of methods. *Educational Technology Research and Development, 61*(2), 171–195. <https://doi.org/10.1007/s11423-012-9281-2>.
- Graham, S., & Hebert, M. (2010). *Writing to read: A report from Carnegie Corporation of New York. Evidence for how writing can improve reading*. New York: Carnegie Corporation. https://www.carnegie.org/media/filer_public/9d/e2/9de20604-a055-42da-bc00-77da949b29d7/ccny_report_2010_writing.pdf.
- Ifenthaler, D. (2010). Relational, structural, and semantic analysis of graphical representations and concept maps. *Educational Technology Research and Development, 58*(1), 81–97.
- Ifenthaler, D., Pirnay-Dummer, P., & Seel, N. M. (Eds.). (2010). *Computer-based diagnostics and systematic analysis of knowledge*. New York: Springer. <https://doi.org/10.1007/978-1-4419-5662-0>.
- Johnson-Laird, P. N. (2004). The history of mental models. In K. Manktelow & M. C. Chung (Eds.), *Psychology of reasoning: Theoretical and historical perspectives* (pp. 179–212). New York: Psychology Press.
- Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kim, M. K. (2012). Cross-validation study of methods and technologies to assess mental models in a complex problem solving situation. *Computers in Human Behavior, 28*(2), 703–717. <https://doi.org/10.1016/j.chb.2011.11.018>.
- Kim, K. (2017a). Visualizing first and second language interactions in science reading: A knowledge structure network approach. *Language Assessment Quarterly, 14*, 328–345.
- Kim, K. (2017b). 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*, 301–314. <https://doi.org/10.1007/s10758-017-9320-5>.
- Kim, K., & Clariana, R. B. (2015). Knowledge structure measures of reader's situation models across languages: Translation engenders richer structure. *Technology, Knowledge and Learning, 20*(2), 249–268. <https://doi.org/10.1007/s10758-015-9246-8>.

- Kim, K., & Clariana, R. B. (2017). Text signals influence second language expository text comprehension: Knowledge structure analysis. *Educational Technology Research and Development*, 65, 909–930. <https://doi.org/10.1007/s11423-016-9494-x>.
- Kim, K., & Clariana, R. B. (2018). Applications of Pathfinder Network scaling for identifying an optimal use of first language for second language science reading comprehension. *Educational Technology Research and Development*. <https://doi.org/10.1007/s11423-018-9607-9>.
- Kiuhara, S. A., Graham, S., & Hawken, L. S. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology*, 101(1), 136–160. <https://doi.org/10.1037/a0013097>.
- Koul, R., Clariana, R. B., & Salehi, R. (2005). Comparing several human and computer-based methods for scoring concept maps and essays. *Journal of Educational Computing Research*, 32(3), 261–273.
- Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development*, 42(2), 7–19.
- Li, P., & Clariana, R. B. (2018). Reading comprehension in L1 and L2: An integrative approach. *Journal of Neurolinguistics*, 45. Retrieved from http://blclab.org/wp-content/uploads/2018/04/Li_Clariana_2018.pdf.
- Mørch, A. I., Engeness, I., Cheng, V. C., Cheung, W. K., & Wong, K. C. (2017). EssayCritic: Writing to learn with a knowledge-based design critiquing system. *Educational Technology & Society*, 20(2), 213–223.
- Mun, Y. (2015). *The effect of sorting and writing tasks on knowledge structure measure in bilinguals' reading comprehension*. Masters Thesis. Retrieved from <https://scholarsphere.psu.edu/files/x059c7329>.
- Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, 76(3), 413–448. <https://doi.org/10.3102/00346543076003413>.
- Ong, W. J. (1982). *Orality and literacy: The technologizing of the word*. London: Methuen.
- Osborne, R., & Wittrock, M. (1985). The Generative Learning Model and its implications for science education. *Studies in Science Education*, 12, 59–87.
- Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology*, 67(3), 215–227.
- Sarwar, G. S. (2012). *Comparing the effect of reflections, written exercises, and multimedia instruction to address learners' misconceptions using structural assessment of knowledge*. Doctoral Thesis, University of Ottawa.
- Sarwar, G. S., & Trumppower, D. L. (2015). Effects of conceptual, procedural, and declarative reflection on students' structural knowledge in physics. *Educational Technology Research and Development*, 63(2), 185–201.
- Spector, J., & Koszalka, T. (2004). *The DEEP methodology for assessing learning in complex domains*. Final report to the National Science Foundation Evaluative Research and Evaluation. Syracuse, NY: Syracuse University.
- Su, I.-H., & Hung, Pi.-H. (2010). *Validity study on automatic scoring methods for the summarization of scientific articles*. A paper presented at the 7th conference of the international test commission, 19–21 July, 2010, Hong Kong. Retrieved from https://bib.irb.hr/datoteka/575883.itc_programme_book-final_2.pdf.
- Tang, H., & Clariana, R. (2017). Leveraging a sorting task as a measure of knowledge structure in bilingual settings. *Technology, Knowledge and Learning*, 22(1), 23–35. <https://doi.org/10.1007/s10758-016-9290-z>.
- Tawfik, A. A., Law, V., Ge, X., Xing, W., & Kim, K. (2018). The effect of sustained vs. faded scaffolding on students' argumentation in ill-structured problem solving. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2018.01.035>.
- Tippett, C. D. (2010). Refutation text in science education: a review of two decades of research. *International Journal of Science and Mathematics Education*, 8(6), 951–970.
- Treagust, D. F., & Duit, R. (2008). Conceptual change: a discussion of theoretical, methodological and practical challenges for science education. *Cultural Studies of Science Education*, 3(2), 297–328. <https://doi.org/10.1007/s11422-008-9090-4>.
- Tripto, J., Assaraf, O. B. Z., & Amit, M. (2018). Recurring patterns in the development of high school biology students' system thinking over time. *Instructional Science*. <https://doi.org/10.1007/s11251-018-9447-3>.
- Trumppower, D. L., & Sarwar, G. S. (2010). Effectiveness of structural feedback provided by Pathfinder networks. *Journal of Educational Computing Research*, 43(1), 7–24.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

- 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.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185. <https://doi.org/10.1037/0033-2909.123.2.162>.

Kyung Kim is Assistant Professor of Educational Technology, Research and Assessment at Northern Illinois University. He has worked on the measures and applications of knowledge structure in various learning environments.

Roy B. Clariana is Professor of Learning, Design, and Technology program and Department Head of Learning and Performance Systems at the Pennsylvania State University. His current research focus is on knowledge structure measures.

Yanghee Kim is Professor of Educational Technology, Research and Assessment and LD & Ruth G. Morgridge Endowed Chair in Teacher Education and Technology at Northern Illinois University. Her research expertise lies in human/technology interaction design, with an emphasis of anthropomorphism (e.g., pedagogical agents and robots), learner diversity, learner affect, and instructional design for inclusion.