Cite as: Wang, Y., Solnosky, R., & Clariana, R. B. (2025). A comparison of full and focused structural feedback on students' knowledge structure and learning. *Technology, Knowledge and Learning,* http://link.springer.com/article/10.1007/s10758-025-09892-0

The effectiveness of full and focused structural feedback on students' knowledge structure and learning

Abstract

Most STEM classrooms overlook the intrinsic conceptual structure of domain content, strategies for improving students' conceptual structure have promise for improving STEM learning outcomes. This experimental investigation continues the development of the web-based tool *Graphical Interface of Knowledge Structure (GIKS)* that provides immediate formative feedback as a network of concepts in the student's essays alongside an expert referent network for comparison and reflection. What should this feedback network look like, especially, should it be more inclusive or small and focused? And is preexisting domain knowledge important for type of network feedback effectiveness? Undergraduate students in a second year Architecture Engineering course, after completing a 2-weeks long lesson on Building with Wood, were randomly assigned to a summary writing task with either Full feedback (a network with 14 central and 12 peripheral terms) or Focused feedback (a network with only the 14 central terms), and then immediately completed a knowledge structure survey. Two weeks later, they completed an End-of-Unit posttest that consisted of a Central-items and a Peripheral-items subtest. A significant interaction of feedback and domain knowledge was observed for post knowledge structure, the low domain knowledge students in the Focus feedback group had the most central link-agreement with the expert and the least peripheral links agreement. On the End-of-Unit declarative knowledge posttest, there was no difference for the Full or Focused feedback interventions, but the high domain knowledge students in both interventions performed significantly better than the low domain knowledge students on the central-items subtest but *not* on the peripheral-items subtest. This investigation shows the need for further research on the role of domain-normative central concepts and pragmatically contributes to the design of essay prompts for STEM classroom use.

(key terms: structural feedback, writing to learn, knowledge structure, essay prompts)

1.0 Introduction

Jonassen et al. (1993) note, “The meaning for any concept or construct is implicit in the pattern of relationships to other concepts or constructs” (p.5). Such patterns of lexical-semantic memory organization continue to develop across the lifespan (Krethlow et al., 2020) and can be shared with others over time and space (Beissner et al., 1994). These concept relationship patterns are referred to here as knowledge structure (KS) that can be represented as network graphs of concepts (i.e., knowledge graphs), in this case, as nouns (Elman, 2004; 2009; Fenker, 1975; Furtner et al., 2009) that offer “a geography of the human mind” (p. 8, Georgakopoulos & Polis, 2018).

KS networks provide a way to represent and visualize the conceptual association patterns of individuals and of group-level language-based artifacts (Clariana, Tang, & Chen, 2022), called local collective knowledge (Teplovs & Scardamalia, 2007). And KS networks can also be used to represent domain knowledge (Shavelson, 1972) and even global knowledge networks of the larger world (Clariana et al., 2022; Lee & Clariana, 2022; Louwerse, 2011).

1.1 Imparting expert conceptual knowledge

Trumpower and Sarwar (2010) propose that “knowledge structures … play a more direct causal [sic mediating] role in enabling good performance” (p.427), so having domain conceptual structure like that of an expert seems like a good idea. Trumpower and Sarwar (2010) coined the term “structural feedback” (e.g., model-based feedback, Ifenthaler, 2011) and demonstrated over several studies that students can use reflection on their own networks to reinforce correct conceptions, to correct misconceptions, and especially to add missing concepts (Sarwar, 2011, 2012; Sarwar & Trumpower, 2015; Trumpower & Sarwar, 2010). They reasoned that when students explicitly compare their knowledge structure to an expert’s structure, they can more readily establish appropriate concept interrelationships and eventually can justify their conceptions.

Some past investigations have provided learners with expert conceptual structure as feedback including as *Pathfinder networks,* as *concept maps*, and as *networks* of essays (Jonassen et al., 1993; Clariana, 2010b). As discussed below, the results across these studies are mixed, providing an expert network as feedback sometimes supports posttest outcomes and sometimes does not, and in some cases negatively impacts posttest outcomes.

1.1.1 Pathfinder networks as feedback when writing to learn

Pathfinder networks used as feedback can be derived from pair-wise ratings of relatedness for a set of concepts, these data are then converted into a linked network using *Pathfinder* software (e.g., Schvaneveldt, 2017). Sarwar and Trumpower (2015) asked 11th grade physics students (n = 133) to rate the relationship of 11 essential physics concepts in lesson units they had just studied (scale 1-5). The following day, each student was given a paper-based handout that contained their individual *Pathfinder* network, the average expert referent network, and individualized instructor-generated feedback comparing the two. The expert referent networks were created by averaging the ratings of multiple instructors in the course to obtain an averaged expert network. The writing task asked students to reflect on specific errors and also to draw corrected links on their own network. Finally, students completed a posttest term-term rating task that was used to generate post KS networks for each student for analysis and comparison to the expert referent. Most notably, the students' written reflections were classified by the researchers as either conceptual, procedural, or declarative in nature. For example,

> In conceptual reflections, students explained how a pair of concepts are related and gave real world examples of the relationship. In procedural reflections, students

> explained how the concepts are related within equations or laws of physics (e.g., directly proportional), but without giving practical examples. In declarative reflections, students may have simply stated that the concepts are related, perhaps also citing a particular equation or law, but without explaining how they are related and without giving any examples. (p. 195, Sarwar & Trumpower, 2015)

Analyses and follow up tests showed that KS significantly improved from pre to post under all three reflection approaches, and that students' post *Pathfinder* network similarity to the expert referent showed significant differences, those with *conceptual* reflections (pre M = .41, post M = 0.78, n = 42) were more like the expert network than those with *procedural* reflections (pre M = .45, post M = 0.69, n = 32) which were more like the expert network than those with *declarative* reflections (pre M = .38, post M = 0.59, n = 59). This shows that when students are given a referent network as feedback along with their own network, they do not all use it in the same way. Students focus on different aspects of the networks resulting in different outcomes, those whose reflections focused on concept pair relationships attained the greatest post similarity to the expert. Since there was not a posttest measure of declarative knowledge, the relationship between post KS and traditional posttest outcomes is not known.

1.1.2 Network maps of essays as feedback

Ifenthaler (2011) asked university students (n = 74) to complete a 27-item declarative knowledge pretest about climate change, use HIMATT to create a concept map of their current understanding of climate change, and then write a text about their understanding of climate change. After a brief recovery period, they received a 1,417-word text on climate change along with one of three forms of automatically generated model-based feedback networks, ether a *cutaway network* (all propositions and links from the student's text), the *expert network*, or the *discrepancy network* (that highlights mismatches between the learner and the expert). Then participants completed the 27-item declarative knowledge posttest. Next, they used the HIMATT system again to construct a second concept map of their understanding of climate change and wrote a second text regarding their understanding of climate change. There was a substantial and significant improvement in climate change declarative knowledge from pretest to posttest, but there was no significant difference for any form of model-based feedback, all were equally effective, although the *cutaway* feedback group (M = 10.8) scored higher than the *discrepancy* group (M = 10.4) and the *expert* group (M = 9.79). The three feedback interventions did not differentially impact post knowledge structures. This is a surprising outcome, because relative to the cutaway network group who did not see the expert network, the final mental models of the discrepancy network and the expert network groups should be more like the expert network. Another surprise is that the written text and the concept maps had different structures and content, even though they wrote the essay immediately after drawing the map with no intervening tasks, drawing the concept map should establish conceptual structures in memory that then show up in the essay.

1.1.3 Network feedback Conceptual mismatch

Taricani and Clariana (2006) asked undergraduates (n = 60) to read a 1,900-word text on the human heart and circulatory system and create a learner-generated concept map. They were randomly assigned to receive either the expert's hierarchical concept map (with 25 terms) or no map (control). The control group outperformed the network feedback group on both the declarative knowledge posttest (ES = .28) and the comprehension posttest (ES = .44). Since there was no post measure of conceptual structure, the relationship between these traditional posttest outcomes and post KS is not known. Similarly, Lee and Nelson (2005) reported that when learners are given a referent concept map "they cannot organize their concept map to make it fit with their internal knowledge structures" (p. 194). Also, Lambiotte and Dansereau (1992) say, "students with more well-established schemas for the circulatory system performed less well when structure was imposed by an outline or a concept map" (p. 198).

Why? "…explicitly manipulating conceptual structure is dangerous because it is powerful…" (p. 30, Clariana, 2025). Visual memory is fundamentally different than verbal memory, the visual nature of networks of expert content when used as feedback will likely perturb verbal memory *when the map structure misaligns with the learner's conceptual structure*. This would be especially true for some students and must so be included during instructional design.

1.2 Network feedback form

Given the mixed findings, much work remains to understand how structural feedback as networks can influence conceptual change (Ntshalintshali & Clariana, 2020). One promising avenue for research is to consider the form of the expert referent network. Applying similar design issues from the concept map literature can help. Krieglstein, Schneider, Beege, and Rey (2022) note, "the research on how to design concept maps as conducive to learning as possible is still rare, in particular, the salience of spatial arrangement of thematically related concepts within the map as well as the complexity of the map…" (p. 1). They defined *salience* as the spatial closeness of related concepts and *complexity* as the number of nodes in the network (i.e., the raw size of the network). Note that Krieglstein's et al. (2022) complex map is relatively large, with 80 nodes (concepts) compared to 45 nodes in the control map. It was hypothesized that "In terms of disorientation, a lower number of nodes facilitates learning since additional integrating processes of related nodes are reduced" (p. 103). In their study, complexity did not significantly impact performance on multiple-choice or open-ended response comprehension posttests, perhaps because even the smaller map (45 nodes) was too complex. For example, Clariana and Taricani (2010) used concept maps with 16, 26, and 36 terms, 16 terms attained the best concurrent validity with the declarative knowledge tests (Fanella, 2015).

As noted by Krieglstein et al. (2022), extensive concept map research has rarely considered map complexity (size), and this is even rarer in research on structural feedback. So how large should a network be? To address this question, research with a writing-to-learn computer-based tool called *Graphical Interface of Knowledge Structure* (*GIKS*) is presented next.

1.3 Writing to learn with structural network feedback

A meta-analysis by Bangert-Drowns et al. (2004) across 47 studies considered the effects of writing to learn with feedback, feedback was more effective than no feedback for academic achievement (ES = .32) and this difference was even greater when asked to reflect on the feedback (ES = .44). To write a summary essay, the student must recall the related concepts, distinguish the concepts based on the level of importance, and reassemble the concepts in a coherent way. Thus, writing to learn is a way for students to recall, reorganize, and build conceptual knowledge as well as correct misconceptions (Eryilmaz, 2002; Finkenstaedt-Quinn et al., 2021; Moon et al., 2018).

Writing a summary of an informational text is difficult for many reasons, but it is especially difficult to summarize if the information is unfamiliar because it may be unclear what is important. For example, Yeari and Lantin (2021) report that less-skilled readers show a centrality deficit, defined as poor recall of central ideas. An individual's pre-existing knowledge of that content must substantially matter when writing due to the importance of chunking on memory and on preferential attachment, where concepts that are already well connected in their semantic networks are better at acquiring new links (Mak & Twitchell, 2020), the "rich get richer effect" (Bogaerds-Hazenberg et al., 2020; De Jong & Ferguson-Hessler, 1986; Witherby & Carpenter, 2021).

So, will learners with greater content knowledge benefit most from *larger* networks (i.e., thus more complex) that include both central and peripheral terms, while learners with lesser content knowledge benefit most from smaller *focused* networks that include only central terms? Kim, Clariana, and Kim (2019) compared three structural feedback approaches designed to support learning through writing and revision. High school physics students (n = 180) read three separate lesson texts and then wrote a summary essay during one of the three treatment strategies (targeted multiple-choice questions, video-delivered information*, and GIKS* using small, focused networks with 11 key term), and then also wrote a summary essay as the posttest. Posttest essay scores significantly improved from lesson-to-posttest for all three forms of feedback, but *GIKS* obtained the greatest gains, with *increase in relevant and decrease in irrelevant relations* (these findings exactly align with the findings from Sarwar, 2012, p. 85). An interaction was observed. The central concept relations improved with all three forms of writing feedback (pre-post Cohen's *d* effect size: *GIKS* = 1.4, MC = 0.5, Video = 0.8, all significant) but the *central concept relations improved most with GIKS*, while viewing the multimedia feedback showed the greatest pre-to-post increase in the less important peripheral concepts. Thus, providing structural feedback as small, focused networks (11 terms) strongly highlighted the central concepts in the text.

1.4 Purpose

Only a handful of studies have considered network structural feedback when writing to learn and possibly none have considered feedback network complexity in terms of map size and composition (Krieglstein et al., 2022). This experimental investigation begins to address this gap and asks the questions, "How large should a KS network used as structural feedback be?" And "Will learners with lesser content knowledge benefit most from smaller focused networks that include only central terms?"

Specifically, the effectiveness of full versus focused network feedback (i.e., 26 terms versus 14 terms) are compared in a real classroom using the required course material under ordinary conditions (for generalizability). Outcome measures are initial lesson essays, post-writing knowledge structure measures, and the existing End-of-Unit test of declarative knowledge. Descriptive data of essay term frequencies are presented as an indicator of preferential attachment (Clariana & Solnosky, 2024). Conceptual structure is central to this investigation, so the findings of this investigation can support both the design of structural feedback network forms as well as contribute to a developing theory of knowledge structure.

2.0 Method

2.1 Participants

Undergraduate engineering students (n = 110) were recruited in an Architecture Engineering course (AE 222 *Building Documentation and Modeling*) at a large land-grant university in the Northeastern United States. Students in the course were mostly 2nd year students (sophomores) who had covered some of the basic fundamental engineering theories in previous courses. Current college-wide demographics of the Architectural Engineering undergraduate program area (total n = 327) are 32.7% reporting as female, 6.7% international, and 6.2%

underrepresented (includes American Indian/Alaska Native, Black/African American, Hispanic/Latino, Native Hawaiian/Other Pacific Islander). Only 80 volunteered to participate in this research investigation, and there were only 70 of these with complete data for analysis. All data were collected under the University's institutional review board (study #00014420).

Participants were categorized as either low or high domain knowledge (DK) based on median-split of a concurrent predictor (Westerberg et al., 2021) of domain knowledge (i.e., the Neither items subtest, described below). Although median split has fallen into disfavor, Iacobucci et al. (2015) used simulation studies to show that median split is robust and "letting a median split serve as a factor is completely legitimate" (p. 690). In addition, median split is preferred by many researchers "due to the beauty of their parsimony and the ease with which results may be communicated." (p. 691).

2.2 Materials

Materials consist of the GIKS tool and GIKS writing prompts, the expert's referent network used as structural feedback, the knowledge structure survey posttest, and the End-of-Unit declarative knowledge posttest. These are presented next.

2.2.1 The expert network central and peripheral terms

To develop the network referent to be used as structural feedback, the course instructor was given a large network of the textbook chapter generated by *ALA-Reader* software (Clariana, 2010a) and then was asked to create a network of the lesson topic with about 20 to 30 terms. This approximate network size was determined in a dissertation by Fanella (2015), using networks that ranged from 10 to 50 terms, he found that the optimal range of key terms in an essay summary was about 25 to 30 words. Also, the research with networks as structural feedback cited above used referenet networks that ranged from 11 to 80 terms. After review and revision by the instructor, the terms in the expert's network were then categorized as central or peripheral for use in the two writing prompts and to create the two forms of network feedback (see Figure 1).

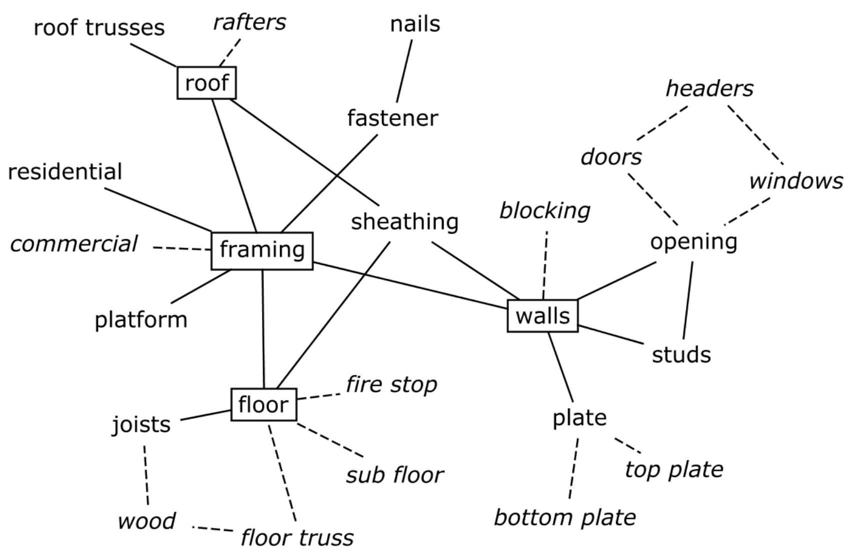


Figure 1. The Full expert referent network with 14 central terms (15 links) and 12 peripheral terms (14 links), periphereal terms are shown in italics with dashed link lines.

The obvious four most central terms are: framing (7 links), surrounded by three subnetworks with floor (6), wall (6), and roof (4). To even out the number of terms categorized as focus or full, additional terms were included as central that are one level out from these four most central terms including sheathing (3 links), openings (4), studs (2), plates (3), and fasteners (2). In addition, for the purpose of how the feedback network appears to the students in the focus feedback teratment, it was decided that these central terms should be linked to at least one peripheral term to visually emphasize the central terms centrality.

2.2.2 The GIKS tool – combining writing to learn with immediate structural feedback

Summary writing with immediate network feedback is combined in a browser-based tool *Graphical Interface of Knowledge Structure* (*GIKS*). *GIKS* is based on the *ALA-Reader* algorithm that converts text to arrays (Clariana & Wallace, 2007; Zhang, Kim, & Clariana, 2017). *The ALA-Reader* algorithm uses text pattern-matching in a forward pass through the text to find pre-identified key terms, disregarding both the distance between the terms and sentence breaks. This term array can be analyzed using *Pathfinder* network analysis software (Schvaneveldt, 1990, 2017) and other approaches such as multidimensional scaling. The term arrays derived from the essays, called proximity files, are converted to *Pathfinder* networks using the data reduction parameters of Q = n - 1 and Minkowski r = infinity. Several investigations have shown the interrater reliability of essays scored by human raters and *ALA-Reader* scores (i.e., Tawfik, Law, Ge, Xing, & Kim, 2018; Zimmerman et al., 2017).

To use *GIKS*, an instructor prepares a writing task by (1) entering a question or prompt, (2) entering a list of key words with their synonyms and metonyms, and (3) adding an expert referent network. The writing prompt can be any combination of text, images, and videos. To use GIKS in class, the instructor provides the URL to students along with a unique ID code. Students log in to review the writing task, compose a response, and then submit it. Immediately an interactive network graph of the essay is displayed along with the referent expert network (see Figure 2). Rather than seeing a different random force-generated network graph each time, the student's term locations in their networks are spatially aligned to the terms in the expert network referent, thus the student views a network structure of their own essay for the first time laid out in a domain-normative way, where spatial contiguity (i.e., term closeness in the subnetwork clusters, floor, wall, roof) reflects that of the expert (i.e., called structural salience, Krieglstein et al., 2022; Schneider, Krieglstein, Beege, & Rey, 2021).

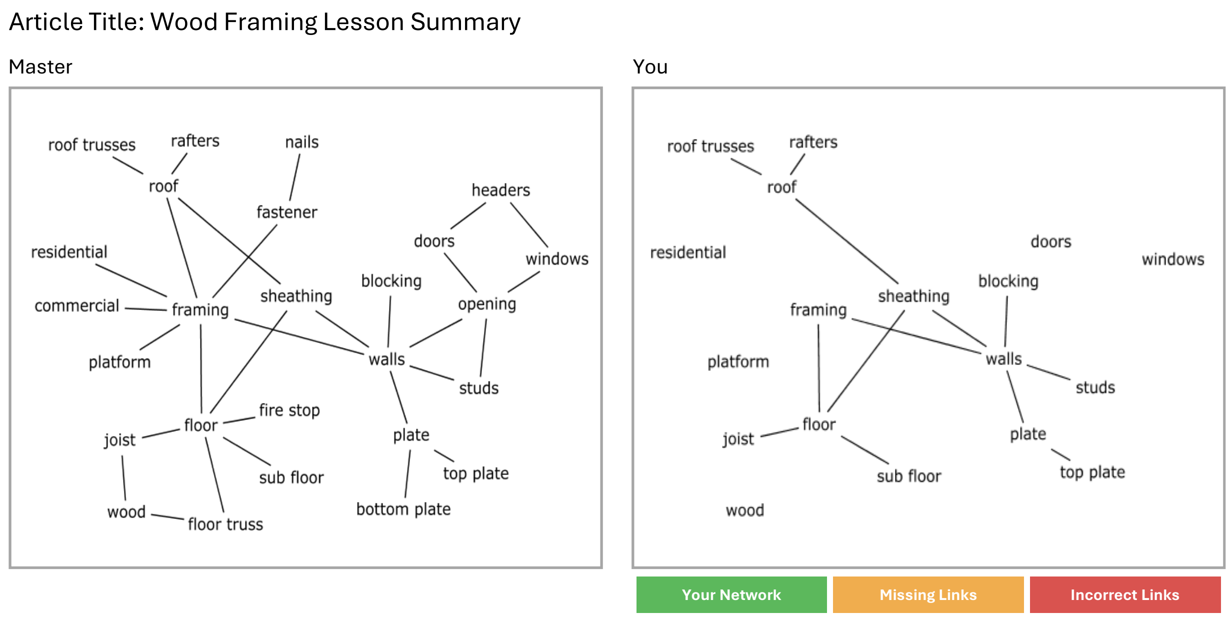


Figure 2. GIKS Full network feedback screen, the expert network is shown on the left side and the student's essay network is on the right.

Clicking on a term in either the expert or student network highlights that term along with its links and term associates in both networks. Dragging any term in either network moves the same term in the other network. These interactive features allow the students to explore the sometimes-complicated networks in a term-by-term way. Also, there are control buttons under the student's network, clicking the green "Your Network" button shows the student's essay network links, clicking the orange "Missing Link/Node" button adds the missing terms and missing links, while clicking the red "Incorrect Links" button shows the incorrect links (see Figure 2).

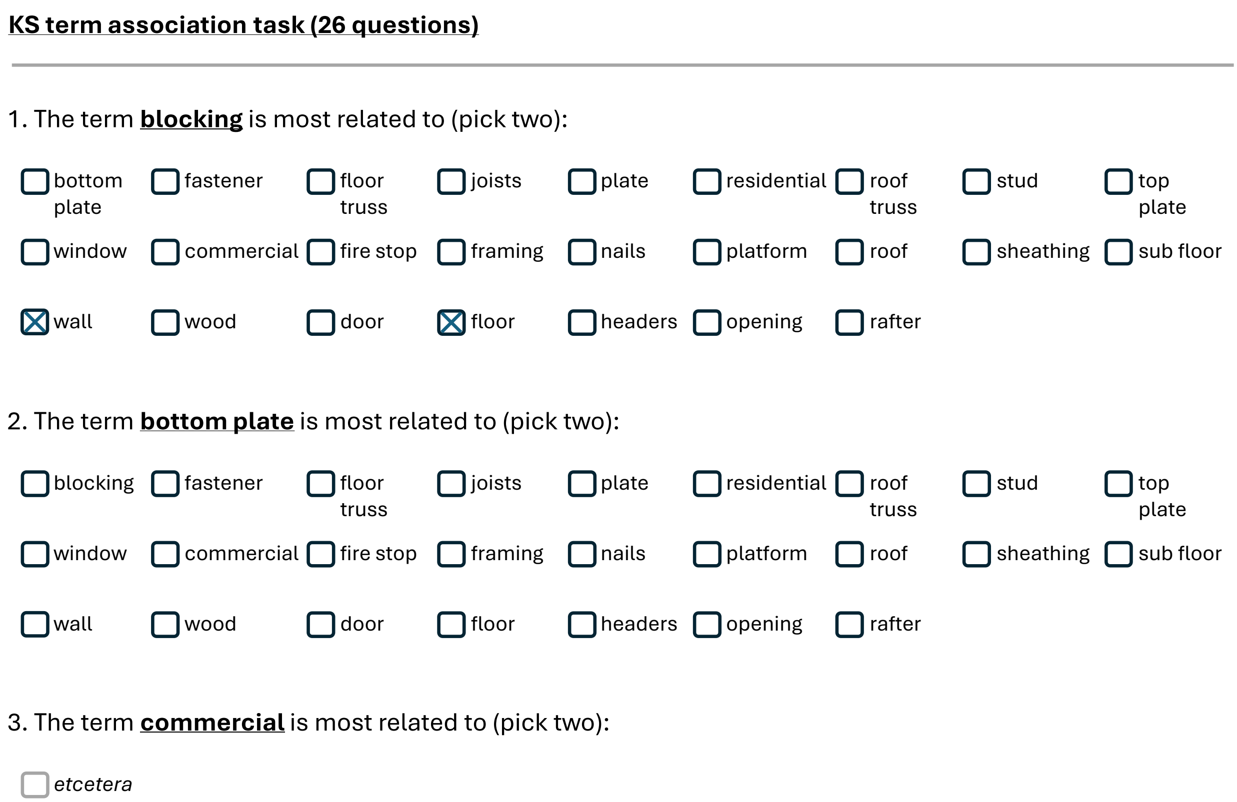
2.2.3 The GIKS writing prompts

The two *GIKS* essay prompts used identical instructions except one prompt included a list of 26 terms (full prompt) and the other had a list of 14 terms (focus prompt). These terms aligned with the structural network feedback that each group received, either Full or Focused. The writing prompt stated, “Given the keywords below, write a 300-word summary of the lesson this week, you may use any terms that you like.” The 14 or 26 keywords were listed in alphabetical order below the instructions.

2.2.4 The knowledge structure survey and the end-of unit posttests

A knowledge structure survey term-term association task (KS survey) was developed and delivered using *Qualtrics* (see Figure 3). The KS survey consisted of 26 items, one for each term in the Full expert network, each item listed a key term along with an alphabetical list of all terms, students were asked to pick two terms from the list that are most related to the key term. For example, in Figure 3, for item 1, “The term blocking is most related to (pick two):”, the responses “wall” and “floor” are checked, this these two pairs were entered into the 26 x 26 proximity file (i.e., array) as “blocking - wall” and “blocking - floor”.

Following the approach described by Lee and Clariana (2022), the proximity files from the KS Survey for each student were converted to networks using *Pathfinder* software (Schvaneveldt, 1990, 2017) with the data reduction parameters of Q = n - 1 and Minkowski r = infinity. Then the student networks were compared to the expert referent network,


KS term association task (26 questions)

1. The term **blocking** is most related to (pick two):

☐ bottom plate	☐ fastener	☐ floor truss	☐ joists	☐ plate	☐ residential	☐ roof truss	☐ stud	☐ top plate
☐ window	☐ commercial	☐ fire stop	☐ framing	☐ nails	☐ platform	☐ roof	☐ sheathing	☐ sub floor
☒ wall	☐ wood	☐ door	☒ floor	☐ headers	☐ opening	☐ rafter		

2. The term **bottom plate** is most related to (pick two):

☐ blocking	☐ fastener	☐ floor truss	☐ joists	☐ plate	☐ residential	☐ roof truss	☐ stud	☐ top plate
☐ window	☐ commercial	☐ fire stop	☐ framing	☐ nails	☐ platform	☐ roof	☐ sheathing	☐ sub floor
☐ wall	☐ wood	☐ door	☐ floor	☐ headers	☐ opening	☐ rafter		

3. The term **commercial** is most related to (pick two):

☐ *etcetera*

Figure 3. Qualtrics interface of the KS survey website.

The End-of-Unit (EOU) declarative knowledge test given about two weeks after the writing task was designed and modified over several years by the course instructor to cover the separate lessons taught over four weeks and is regularly used in this course each semester. The test consists of multiple choice and true-false items, and for the purposes of this investigation, was subdivided into three subtests based on how each test item relates to the 26 key terms selected from the lesson of interest. The EOU subtests include a Central Items Subtest (M = .85, for items associated with the 14 central terms), a Peripheral Items Subtest (M = .82, for items associated with the 12 peripheral terms), and a Neither Subtest (M = .84, for test items covered in other lessons that are unrelated to the central and peripheral terms in this lesson). For example, a

Central item is “Q12: What kind of framing system is used in the image below?” And a Neither item is “Q04: Lumber is sold in what units?”

The even-odd split-half reliability of the test is r = .73 and the Cronbach Alpha of the End-of-Unit total test is α = .62 (Ursachi et al., 2015). Subtest intercorrelations are another indication of reliability, the Pearson correlations among the three subtests are: Central to Peripheral, r = .85, Central to Neither, r = .70, and Peripheral to Neither, r = .33; all correlations are significant at the $p < .01$ level. The Central and the Peripheral subtests are dependent measures in this investigation, while the Neither subtest provides a concurrent measure of related domain knowledge (Westerberg et al., 2021). See Wang (2021) dissertation for complete details of the End-of-Unit test.

2.3 Procedure

The lesson in this investigation consisted of the textbook reading on *Wood Light Frame Construction* (Chapter 5, Allen & Iano, 2019) and two 45-minute lectures (on Monday and on Wednesday). The lecture was presented in the usual fashion with all of the students together in a large lecture-based classroom. In addition, students met for 2 hours on one day the following week (self-selected, either Wednesday, Thursday, or Friday) in a 35-seat computer lab to work on lesson-related lab activities and also on components of the final course project. The online GIKS writing activity and KS survey were completed during lab time as the final activity in the *Building with Wood* lesson sequence. The End-of-Unit test was completed about two weeks later outside of class time as is the usual practice in this course.

All students were required by the instructor to read the lesson materials before class and attend the lectures. During lab time, the investigation was explained by the researchers, then students were invited to participate, and volunteers completed the IRB form. All students were given the URL of their assigned GIKS task (Full or Focused, see Figure 4) as a required lab activity (note: although all students completed the GIKS task, only data from volunteer participants were used for analysis). Students worked on their assigned *GIKS* treatment for approximately 20 to 30 minutes, receiving the network graph as structural feedback immediately on pressing the submit button. Then participants completed the *Qualtrics* KS survey. Two weeks later, all students completed the End-of-Unit posttest covering the content of several lessons covered over that time period.

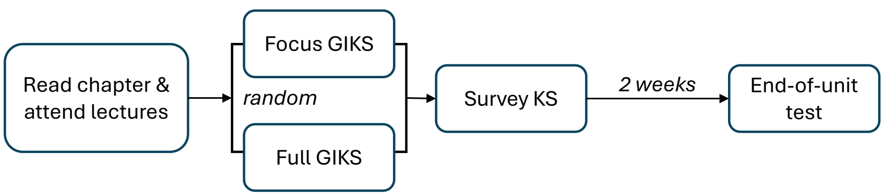


Figure 4. Flow chart of experiment procedure.

3.0 Results

The results section begins with descriptive term frequencies of the students' essays before receiving the network feedback. Then analysis of knowledge structure derived from the KS Survey is provided to consider how GIKS summary writing with immediate network feedback influences knowledge structure of the lesson content downstream. Finally, analysis of performance on the End-of-Unit posttest is provided.

3.1 Descriptive analysis of essays (i.e., before feedback was given)

The descriptive data includes essay length and key word frequency (Clariana & Solnosky, 2024). The Focused group (who were given 14 central terms) and the Full group (who were given 26 terms, both central and peripheral) received slightly different writing prompts when writing their summary. Did including these extra terms in the essay prompt influence the essays in terms of *quantity* measured as total essay length and in terms of central and peripheral term frequencies (Clariana & Solnosky, 2024)?

First, there was a non-significant difference in essay length for the Full and Focus group essays (M = 250 versus 244 words) and also only minor differences in the essay length distributions, $t(78) = .437$, $p = .835$; *Mean diff* = 6.6 words, 95% CI [-22.6, 35.3]. Specifically, the Focus group $M = 250$ words (SD = 65.0, Min = 117, Max = 407, skewness = .074, kurtosis = -.263) and the Full group $M = 244$ words (SD = 64.9, Min = 102, Max = 360, skewness = -.759, kurtosis = -.030).

Descriptive analysis of the frequency counts of central and peripheral terms in the essays (a proxy measure of network node degree) indicates that the term frequencies for the 26 terms for the Full and the Focused treatments were highly correlated, Pearson $r = .96$, indicating that both groups used the 26 terms proportionally the same in their essays (see Figure 5), however the Full group tended to use the five most central terms more in their essays than did the Focus group, while there was no difference between Full and Focused essays for the other 9 central terms nor for all 12 peripheral terms. Thus, when given the full list of 26 terms, only about half of the *most central terms* were used with a relatively greater frequency. This central-terms frequency phenomenon (Zipf's Law) has also been reported by Clariana and Solnosky (2024) who compared summary writing with or without 13 terms in the essay prompt.

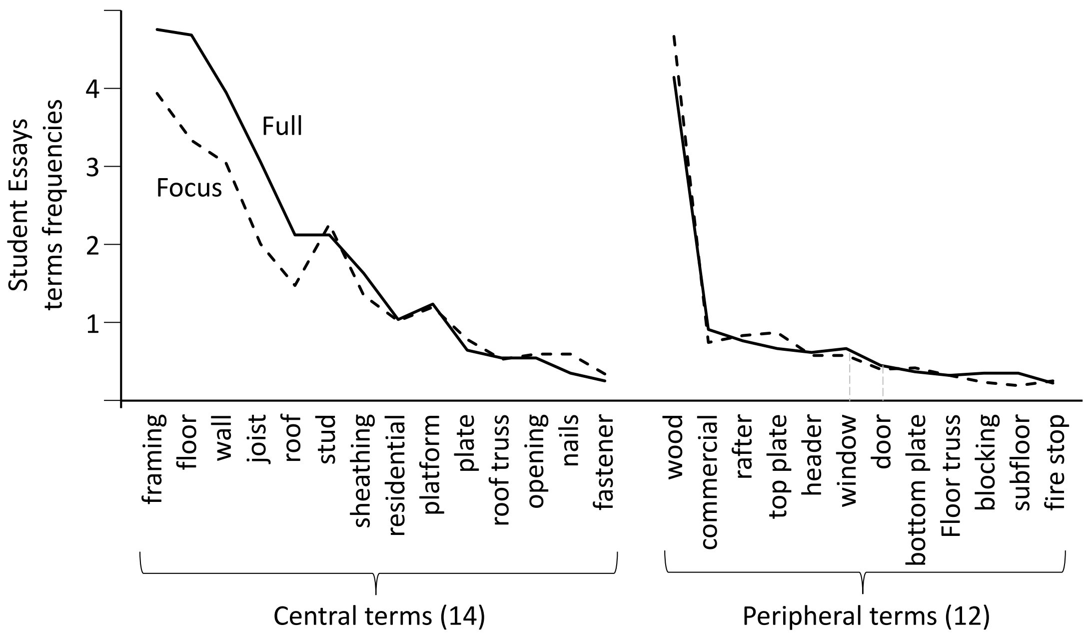


Figure 5. Central and peripheral average term degree frequencies for the students' essays (before feedback) for the Full (solid line) and the Focus (dashed line) groups in descending term frequency order.

3.2 Analysis of essay quality, KS Survey data, and EOU test data

Separate analyses were conducted on three measures including: participants' lesson essay network structure analyzed as network similarity to the expert network, KS survey data analyzed as network similarity to the expert network, and End-of-Unit (EOU) declarative knowledge multiple-choice posttest analyzed as subtest items that aligned with the central or peripheral terms in the expert network.

3.2.1 Inferential analysis of lesson essays (i.e., before feedback was given)

The network core structure of high degree concepts is a potentially important feature of essay concept networks, the term-frequency data reported above for the five most central terms would suggest that the Full group essays (before receiving structural feedback) may be structurally different from the Focus group essays. To consider possible structural differences, the students' essays were converted to networks using *ALA-Reader* (Clariana & Wallace, 2007; Clariana, 2010a) and then essay *quality* was measured as network percent overlap of the students' essay networks with the expert's *full* network.

A 2-between subjects Analysis of Variance (ANOVA) of the lesson essay quality data with the between subjects factors (*Prompt* - full or focused writing prompt) and Domain Knowledge (DK - high or low domain content knowledge) showed no significant main effects for Prompt, $F(1,59) = 0.380$, $p = 0.540$ nor for Domain Knowledge $F(1,59) = 1.644$, $p = 0.205$, also the interaction was not significant $F(1,59) = 0.167$, $p = 0.684$. Students' average essay quality (as networks) in the Full writing prompt treatment (M = .24, SD = .10) were not significantly different from those in the Focus prompt treatment (M = .22, SD = .10). Thus providing 14 terms (focus) or 26 terms (full) in the writing prompt made little difference in essay quality.

3.2.2 *Analysis of KS Survey of post knowledge structure*

The networks derived from the KS survey were compared to the Full network referent, separated into the central-terms subnetwork and the peripheral terms subnetwork (refer back to Figure 1). The KS Survey data are measured as *network percent overlap*, calculated as KS Survey *links in common* with the expert's referents (the expert's central and peripheral subnetworks) divided by the average number of links in the two networks (see Table 1). *Network percent overlap* corrects for the different network sizes in each student's essay network, and is easier to understand by readers unfamiliar with *Pathfinder* similarity measures.

Table 1. KS Survey networks measured as link percent overlap with the expert's Full network referent for the Full and Focus feedback treatments, grouped by low and high domain knowledge (standard deviations shown in parenthesis).

	Central	Peripheral
Full, low DK (n = 16)	.14 (.08)	.18 (.11)
Full, high DK (n = 16)	.15 (.06)	.17 (.07)
Focus, low DK (n = 19)	.16 (.07)	.15 (.04)
Focus, high DK (n = 19)	.13 (.04)	.19 (.04)

The KS survey data network similarity to the expert's Central network (15 links max.) and Peripheral network (14 links max.) were analyzed using Multivariate Analysis of Variance (MANOVA) as a 2 Intervention (*Feedback* - full or focused network feedback) and 2 *Domain Knowledge* (DK - high or low domain content knowledge). The central and peripheral similarity measures were significantly related, Pearson $r = .52$, $p < .001$. There was homogeneity of variances, as assessed by Levene's test for equality of variances, for the Central network ($p = .08$) but not for the Peripheral network ($p = .001$). Box's test to estimate the equality of covariance matrices was significant, Box M = 30.408, $p < .001$, so Wilk's lambda values are reported.

The two main factors *Feedback*, $F(2, 65) = 0.181$, $p = .84$, and *Domain Knowledge*, $F(2, 65) = 1.238$, $p = .30$, were not significant, but there was a significant MANOVA effect observed for the interaction of *Domain Knowledge* and *Feedback* on the combined dependent variables, $F(2, 65) = 6.053$, $p = .004$; Wilks' $\Lambda = .843$, partial eta squared $\eta^2 = .157$ (see the left panel of Figure 6). Separate follow-up ANOVAs for the KS Survey Central network similarity to the expert and for the KS Survey Peripheral network similarity to the expert, each showed *no significant* main or interaction effects.

To further explore this significant MANOVA interaction, a *post hoc* Feedback x Domain Knowledge (DK) ANOVA was conducted on *difference scores* as the dependent variable (*difference scores* calculated as central % similarity minus peripheral % similarity). As in the MANOVA above, the two main factors *Feedback*, $F(1, 66) = 0.364$, $p = .55$, and *Domain Knowledge*, $F(2, 65) = 2.513$, $p = .12$, were not significant, but there was a highly significant interaction of *Domain Knowledge* and *Feedback*, $F(1, 66) = 12.216$, $p < .001$, partial eta squared $\eta^2 = .156$ (see the right panel of Figure 6).

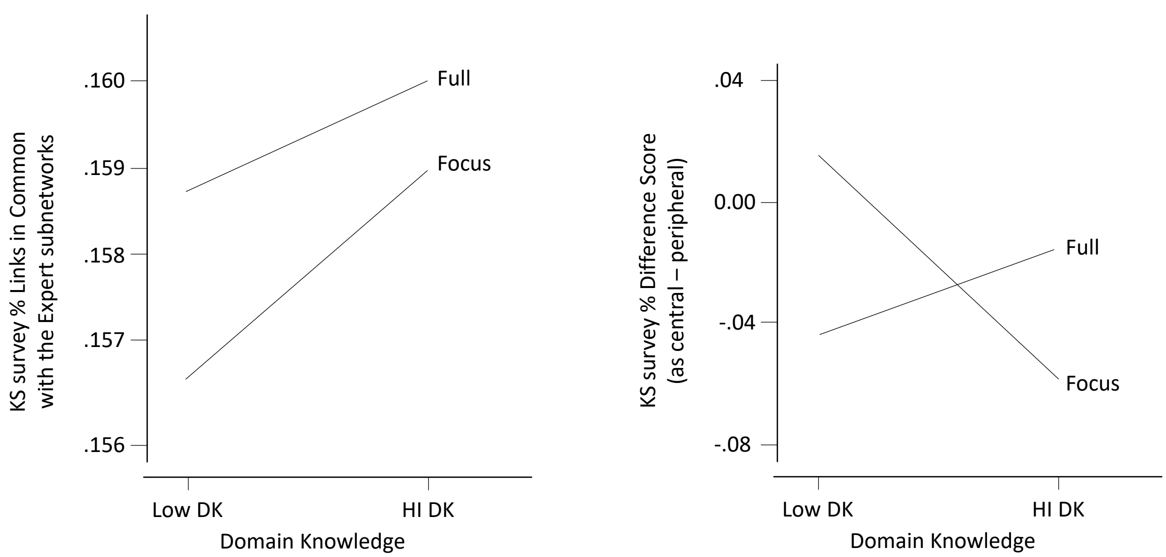


Figure 6. The KS Survey network data interaction of *Feedback* and *Domain Knowledge* (DK) showing the combined estimated marginal means (left panel), and the significant follow-up interaction of difference scores calculated as Central minus Peripheral network scores (right panel).

Separate follow-up tests within DK group compared Full and Focus Difference scores show first for the low group, Focus > Full, $F(1,33) = 6.891$, $p = .013$, ES = .89 and then for the high group, Full = Focus, ($F(1,38) = 3.580$, $p = .066$, ES = .78. This significant difference for the low DK group's KS survey network indicates that receiving *the focused feedback intervention* *increased the number of central terms relative to peripheral terms*. Thus, the focused network feedback intervention had the anticipated effect of improved central knowledge structure for low domain knowledge students.

3.2.3 End-of-Unit Posttest Performance

The End-of-Unit declarative knowledge Central Items and Peripheral Items Subtests data (see Table 2) were analyzed by Multivariate Analysis of Variance (MANOVA; Stevens, 1996) with two factors: *Intervention* (*Feedback* - full or focused network feedback) and *Domain Knowledge* (high or low knowledge DK). There was homogeneity of variance, as assessed by Levene's test for equality of variances for both the Central Items Subtest (p = .35) and the Peripheral Items Subtest (p = .66). Mauchly's test indicated that the assumption of sphericity was not violated. However, Box's test to estimate the equality of covariance matrices was significant, Box M = 60.91, $p < .001$, so the Wilk's lambda values are reported.

Table 2. End-of-Unit subtest means for the Full and Focus treatments for low and high domain knowledge groups (standard deviation shown in parenthesis).

	End-of Unit Subtests	
	Central	Peripheral
Full low (n = 16)	0.77 (0.17)	0.79 (0.19)
Full high (n = 16)	0.91 (0.08)	0.87 (0.13)
Focus low (n = 19)	0.80 (0.11)	0.79 (0.25)
Focus high (n = 19)	0.91 (0.08)	0.85 (0.15)

There was no significant main effect of the feedback intervention, $F(1, 65) = 1.173$, p = .32, the Full or Focused feedback did not differentially influence performance on the EOU subtests, nor interaction of feedback and *Domain Knowledge*, $F(1, 65) = .259$, p = .77. However, there was a statistically significant effect for the *Domain Knowledge* (DK) on the combined dependent variables, $F(1, 65) = 25.936$, $p < .001$; Wilks' $\Lambda = .557$, partial eta squared $\eta^2 = .444$ (high DK > Low DK). As would be expected, the high domain knowledge students outperformed the low domain knowledge students on the EOU measure. However, the separate follow up ANOVAs of each EOU Subtest show a statistically significant effect for the *Domain Knowledge* (DK) factor only for the *Central Items Subtest*, low DK M = .79 < high DK M = .91, $F(1, 66) = 21.947$, $p < .001$, partial eta squared $\eta^2 = .250$, but not for the *Peripheral Items Subtest*, low DK M = .79 ≅ high DK M = .86, $F(1, 66) = 2.457$, $p = .12$, partial eta squared $\eta^2 = .036$. Thus, the high DK group scored higher than the low DK group on the central but *not the peripheral* EOU subtests.

4.0 Discussion

This experimental investigation was designed first to further the research base on structural feedback regarding the complexity (size) of the network provided as feedback (Kim et al., 2019; Kim & McCarthy, 2021; Sarwar, 2012; Sarwar & Trumpower, 2015) and second to establish the likely benefit of a smaller network for low domain knowledge students. The Focus feedback treatment supported low domain knowledge students' acquisition of the central terms network structure and both Focus and Full treatments supported high domain knowledge students' End-of-Unit declarative knowledge posttest scores on the central terms subtest but not on the peripheral terms subtest.

4.1 Elaboration of the results

The term frequency descriptive analysis of the students' *essays* before receiving structural feedback shows that the students in both Full and Focus treatments were sensitive to the four most central terms in the expert's network, *framing*, *floor*, *walls*, and *roof*, but also to *joist* (a peripheral term). During summary writing, students should have read the chapter and attended the lectures on this content, but they had not yet received any *explicit* structure information beyond the full or focused terms list in the writing prompt that might influence their use of the central terms. A similar sensitivity to central concepts was reported by Clariana and Solnosky (2024) for summary writing and also in neurocognitive reading comprehension research, where Swett et al. (2013) in an investigation of the fMRI neural correlates of expository text comprehension says, “The authors reported different patterns for central versus peripheral text concepts, which implies that good readers notice and use the implicit textual KS of the expository text by focusing on the central and peripheral concepts differently” (Hsu et al, 2019, p. 4).

Regarding post knowledge structure elicited as a KS survey, a significant interaction of Feedback and Domain Knowledge was observed, the low domain knowledge students in the Focus feedback group had the most central link agreement with the expert and the least peripheral links relative to the other three groups (see Table 1). Note that Tawfik et al. (2020) reported that students with high domain knowledge benefited more from *GIKS* than those with lower domain knowledge (i.e., “the rich get richer effect”, Witherby & Carpenter, 2021). In contrast, the findings here show that the low domain knowledge students benefitted more as expected from the *Focused* network feedback (i.e., most central terms) but not the Full, while the high domain knowledge students benefitted equally from Full and from Focused network feedback.

Existing domain knowledge is an important factor for most learning (Dochy et al., 2002), the low domain knowledge students probably started with a centrality deficit (Yeari & Lantin, 2021) and so focusing on the central terms structure allowed them to establish a foundation structure of the central terms. The high domain knowledge students probably started out with a better knowledge structure foundation to build upon and to integrate with the new lesson content (Cordova et al., 2014), so it was easier for them to add both central and peripheral terms, as long as the their prior knowledge structure did not conflict with the new content structure to be learned (i.e., a possible expertise reversal effect, Kalyuga et al., 2003). But further research is needed to determine what kinds of structural feedback are appropriate for improving test performance for low (novice) and for high domain knowledge students (Chen et al., 2022; Kim & Tawfik, 2021).

This investigation only considered network *complexity* (size, Krieglstein et al., 2022) of Full or Focused expert networks as structural feedback. Note that Yin et al. (2005) has alternately defined map complexity in terms of its visual form, such as hub, chain, and net. Future research should consider the possible learning benefit not only of network size but also form (e.g., hub, chain, net) and visual salience, such as using greater white space between clusters of terms to create various spatial layouts.

4.2 *Limitations*

Due to the seating capacity of the engineering lab (25-30 participants), the data were collected over three days. This made it easier to collect the signed IRB consent forms, to implement the interventions, and this may have encouraged students to participate. Also, this multi-day approach aligns with most real-world classrooms and so improves real-world generalizability. However, the students on the first day might have informed some of the students scheduled for the second or third day. Since we informed all the participants of the procedures and the intention of the investigation before each lab meeting, passing on information to peers should have less effect. Future studies should consider controlling this kind of student-to-student communication.

Another possible limitation of this investigation are the data analytical tools, including *GIKS*, the *ALA-Reader* algorithm for converting text to array files, and the *Pathfinder* software. These approaches and these analytical tools depend on pre-selected terms that were determined from the instructor's referent network, but there could be a more optimal set of terms. More research is needed regarding how to establish the terms and the expert network referents.

Another limitation is that the reliability of the instructor-made End-of-Unit multiple choice declarative knowledge posttest was only adequate. This has been regularly reported for instructor-made compared to researcher developed tests (Bangert-Drowns et al., 2004; Slavin, 2019). Specifically, there was a ceiling effect, with 20 of the 70 students scoring a 100% score on one of the three subtests of declarative knowledge posttest with 8 of these 20 scoring 100% on all three subtests. Although a significant difference was observed for domain knowledge, a ceiling effect may have obscured differences in the feedback interventions. Using instructor-made tests versus researcher-made tests supports the generalizability of results to other real classrooms, but future investigations should use better measures. For example, a posttest summary essay task may better capture the outcomes of writing interventions and would make it possible to directly compare lesson essays to posttest essays.

This was a brief intervention, the writing task with feedback was only about 30 minutes long. Future research should consider repeated writing over an extended time, for example, writing after each lesson in the course. Also, qualitative data could be added through focus groups to ask participants how they used the feedback, or a qualitative 1:1 think-aloud protocol while using the tool could help interpret how the feedback was perceived and used, and this would help to interpret the outcome measures.

In the present investigation, the instructor had a high expectation for this essay summary activity to improve students' learning and understanding. Thus, there may be a Pygmalion effect in this investigation that may not replicate in other settings (Carreira & Silva, 1998).

Regarding the essay writing prompt, this investigation used an alphabetically ordered list of keywords in the Full and Focus treatments. Further research should include a no terms control group (Clariana & Solnosky, 2024). In addition, adding structure to the list of key terms may also influence the essays, for example, key terms could be listed in order of importance or in the order of occurrence in an expert's summary essay in order to see if the order of terms in the essay prompt matters.

4.3 *Theory into practice*

A fundamental premise of this investigation is that the patterns of conceptual structure of learning artifacts, in this case the textbook, lectures, lab activities, and the expert's network as structural feedback, are a critical aspect of the influence and effect of the lesson materials. Gibson et al. (2023) notes, "Another approach of AI [artificial intelligence] assisting assimilation and future accommodation is providing visualized model-based feedback, in which concept maps are offered that are structurally and semantically like expert solutions." (p. 1134). If structure is a fundamental characteristic of domain knowledge (Lehmann et al., 2020), then both domain structure and structural feedback may become more important in learning than previously assumed. Knowledge structure tools such as Highly Integrated Model Assessment Technology and Tools (HIMATT, Ifenthaler, 2011; Pirnay-Dummer et al., 2010), the *ALA-Reader* algorithm, and *GIKS* could be used to expand the explanatory power of future research on learning and for the design of future tools.

Obviously, there are various ways a teacher could use student networks and expert referent networks in their courses. Following Ifenthaler (2011), Trumpower, Filiz, and Sarwar (2014) list three approaches:

> First, simply providing a referent knowledge map that depicts well-structured knowledge to students, similar to providing exemplar essays or correct answers to an exam as feedback, may facilitate further development of their understanding. Second, providing students with both their knowledge map and a referent map for comparison may prove useful. And third, providing additional content, suggested by discrepancies between students and referent knowledge maps as being potentially fruitful areas to study, could be useful. (p. 228)

This current investigation considered the second approach that uses a student map and a referent map. Further research should consider adding additional content, for example, an instructor during class or later as a podcast could present student networks (or a group averaged network) along with a referent network in order to discuss similarities and discrepancies.

5.0 In Conclusion

This investigation was conducted in a typical engineering course; the students were quite focused on hands-on lab activities that take a huge amount of time in their study plan. Trotskovsky et al. (2015) noted that many engineering students misperceive significant engineering concepts needed to solve even simple problems in real-world practice. The present investigation is a step forward in having students pause and conceptually engage with conceptual domain knowledge, both concrete and abstract (i.e., how a masonry veneer wall composite aligns with heat energy principles) that are necessary to solve problems, make predictions, and generate questions. Summary writing with *GIKS* provides a unique way to examine students' conceptual learning which also can trigger self-evaluation, especially for finding and correcting their knowledge gaps, misunderstandings, and misconceptions (Ntshalintshali & Clariana, 2020).

Concept acquisition (Beissner et al., 1994) is more than "having" concepts but rather is about how acquired concepts are associated together as knowledge structure. Today, most teaching and learning in STEM classroom overlooks the inherent structure of domain content, but shifting the classroom focus on knowledge structure from background to foreground has promise for improving STEM learning outcomes (Trumpower & Sarwar, 2010).

6.0 Acknowledgement

This material is based upon work supported by the U.S. National Science Foundation under award No. (2215807). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. National Science Foundation.

The authors have no competing interests.

7.0 References

Allen, E. & Iano, J. (2019). *Fundamentals of building constriction: Materials and methods (7^{th} edition)*. Hoboken, NJ: John Wiley & Sons, Inc.

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.

Beissner, K. L., Jonassen, D. H., & Grabowski, B. L. (1994). Using and selecting graphic techniques to acquire structural knowledge. *Performance Improvement Quarterly, 7*(4), 20-38.

Bogaerds-Hazenberg, S. T. M., Evers-Vermeul, J., & van den Bergh, H. (2020). A meta-analysis on the effects of text structure instruction on reading comprehension in the upper elementary grades. *Reading Research Quarterly, 56*(3), 435-462. doi:10.1002/rrq.311

Carreira, J., & Silva, J. (1998). Computer Science and the Pygmalion Effect. *Computer, 31*(2), 116-117. doi:10.1109/2.652979

Chen, X., Li, Z., Wei, Z., & Clariana, R. B. (2022). The influence of the conceptual structure of external representations when relearning historical content. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-022-10176-y

Clariana, R. B. (2010a). Deriving group knowledge structure from semantic maps and from essays. In D. Ifenthaler, P. Pirnay-Dummer, & N.M. Seel (Eds.), *Computer-Based Diagnostics and Systematic Analysis of Knowledge* (Chapter 7, pp. 117-130). New York, NY: Springer.

Clariana, R.B. (2010b). 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, NY: Springer.

Clariana, R. B. (2025). Framing knowledge as conceptual structure. In P. Isaías & D. Ifenthaler (Eds.), *Computer-Based Diagnostics and Systematic Analysis of Knowledge, 2nd edition* (pp. 9-37). Springer. https://link.springer.com/chapter/10.1007/978-3-031-87740-7_2

Clariana, R. B., & Solnosky, R. (2024). Anchoring concepts – conceptual structure and test performance. In P. Isaías, D, G. Sampson, D. Ifenthaler (Eds.), Artificial Intelligence for Supporting Human Cognition and Exploratory Learning in the Digital Age (pp. 87–100). Springer. https://link.springer.com/chapter/10.1007/978-3-031-66462-5_5

Clariana, R. B., Tang, H., & Chen, X. (2022). Corroborating a sorting task measure of individual and of local collective knowledge structure. *Educational Technology Research and Development*, *70*, 1195-1219. https://doi.org/10.1007/s11423-022-10123-x

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), 163-173.

Clariana, R. B., & Wallace, P. E. (2007). A computer-based approach for deriving and measuring individual and team knowledge structure from essay questions. *Journal of Educational Computing Research, 37* (3), 209-225.

Cordova, J. R., Sinatra, G. M., Jones, S. H., Taasoobshirazi, G., & Lombardi, D. (2014). Confidence in prior knowledge, self-efficacy, interest and prior knowledge: Influences on conceptual change. *Contemporary Educational Psychology*, *39*(2), 164-174. doi:10.1016/j.cedpsych.2014.03.006

De Jong, T., & Ferguson-Hessler, M. G. (1986). Cognitive structures of good and poor novice problem solvers in physics. *Journal of Educational Psychology, 78*(4), 279-288. https://doi.org/10.1037/0022-0663.78.4.279

Dochy, F., De Rijdt, C., & Dyck, W. (2002). Cognitive prerequisites and learning. *Active Learning in Higher Education*, *3*(3), 265-284. doi:10.1177/1469787402003003006

Elman, J. L. (2004). An alternative view of the mental lexicon. *TRENDS in Cognitive Sciences, 8*(7), 301-306.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: lexical knowledge without a lexicon. *Cognitive Science, 33*, 547-582.

Eryilmaz, A. (2002). Effects of conceptual assignments and conceptual change discussions on students' misconceptions and achievement regarding force and motion. *Journal of Research in Science Teaching, 39*(10), 1001-1015. doi:10.1002/tea.10054

Fanella, D. F. (2015). *The effects of changing the number of terms used to create proximity files on the predictive ability of scoring essay-derived network graphs via the ALA-Reader approach* (Doctoral dissertation, Pennsylvania State University). Electronic Thesis and Dissertations. https://etda.libraries.psu.edu/files/final_submissions/11192

Fenker, R. M. (1975). The organization of conceptual materials: A methodology for measuring ideal and actual cognitive structures. *Instructional Science, 4*(1), 33-57. doi:10.1007/bf00157069

Finkenstaedt-Quinn, S. A., Petterson, M., Gere, A., & Shultz, G. (2021). Praxis of Writing-to-Learn: A model for the design and propagation of Writing-to-Learn in STEM. *Journal of Chemistry Education, 98*, 1548-1555.

Furtner, M. R., Rauthmann, J. F., & Sachse, P. (2009). Nomen est omen: Investigating the dominance of nouns in word comprehension with eye movement analyses. *Advances in Cognitive Psychology, 5*, 91-104. https://www.ncbi.nlm.nih.gov/pubmed/20523853

Georgakopoulos, T., & Polis, S. (2018). The semantic map model: State of the art and future avenues for linguistic research. *Language and Linguistics Compass, 12*, 1-33. https://doi.org/10.1111/lnc3.12270

Gibson, D., Kovanovic, V., Ifenthaler, D., Dexter, S., & Feng, S. (2023). Learning theories for artificial intelligence promoting learning processes. *British Journal of Educational Technology*, *54*(5), 1125-1146. https://doi.org/10.1111/bjet.13341

Hsu, C-T., Clariana, R. B., Schloss, B., & Li, P. (2019). Neurocognitive signatures of naturalistic reading of scientific texts: A fixation-related fMRI study. *Scientific Reports, 9*, 10678. https://www.nature.com/articles/s41598-019-47176-7

Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015). The median split: Robust, refined, and revived. *Journal of Consumer Psychology, 25*(4), 690-704. https://doi.org/10.1016/j.jcps.2015.06.014

Ifenthaler, D. (2011). Bridging the gap between expert-novice differences: The model-based feedback approach. *Journal of Research on Technology in Education, 43*(2), 103-117.

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

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist, 38*, 23-31.

Kim, K., Clariana, R. B., & Kim, Y. (2019). Automatic representation of knowledge structure: enhancing learning through knowledge structure reflection in an online course. *Educational Technology Research and Development, 67*, 105-122. doi:10.1007/s11423-018-9626-6

Kim, K., & Tawfik, A. A. (2021). Different approaches to collaborative problem solving between successful versus less successful problem solvers: Tracking changes of knowledge structure. *Journal of Research on Technology in Education*. https://doi.org/10.1080/15391523.2021.2014374

Kim, M. K., & McCarthy, K. S. (2021). Using graph centrality as a global index to assess students' mental model structure development during summary writing. *Educational Technology Research and Development*, *69*(2), 971-1002. doi:10.1007/s11423-021-09942-1

Krethlow, G., Fargier, R., & Laganaro, M. (2020). Age-specific effects of lexical-semantic networks on word production. *Cognitive Science, 44*, e12915.

Krieglstein, F., Schneider, S., Beege, M., & Rey, G. D. (2022). How the design and complexity of concept maps influence cognitive and motivational learning processes. *Educational Technology Research and Development*, *70*, 99-118. https://doi.org/10.1007/s11423-022-10083-2

Lambiotte, J., & Dansereau, D. (1992). Effects of knowledge maps and prior knowledge on recall of science lecture content. *Journal of Experimental Education, 60*, 189-201.

Lee, M. K., & Clariana, R. B. (2022). The influence of external concept structures on an individual's knowledge structures. *Educational Technology Research and Development, 70*, 1657-1674. https://doi.org/10.1007/s11423-022-10144-6

Lee, Y., & Nelson, D. (2005). Viewing or visualizing-which concept map strategy works best on problem solving performance? *British Journal of Educational Technology, 36*(2), 193-203.

Lehmann, T., Pirnay-Dummer, P. & Schmidt-Borcherding, F. (2020). Fostering integrated mental models of different professional knowledge domains: instructional approaches and model-based analyses. *Educational Technology Research and Development, 68*, 905-927. https://doi.org/10.1007/s11423-019-09704-0

Louwerse, M. M. (2011). Symbol interdependence in symbolic and embodied cognition. *Topics in Cognitive Science, 3*, 273-302.

Mak, M. H. C., & Twitchell, H. (2020). Evidence for preferential attachment: Words that are more well connected in semantic networks are better at acquiring new links in paired-associate learning. *Psychonomic Bulletin & Review*, 27, 1059-1069

Moon, A., Gere, A. R., & Shultz, G.V. (2018). Writing in the STEM classroom: Faculty conceptions of writing and its role in the undergraduate classroom. *Science Education, 109*, 1007-1028. https://doi.org/10.1002/sce.21454

Ntshalintshali, G. N., & Clariana, R. B. (2020). Paraphrasing refutation text improved higher knowledge forms and hindered lower knowledge forms: examples from repairing relational database design misconceptions. *Educational Technology Research and Development, 68*, 2165-2183. https://doi.org/10.1007/s11423-020-09758-5

Pirnay-Dummer, P., Ifenthaler, D., & Spector, J. M. (2010). Highly integrated model assessment technology and tools. *Educational Technology Research and Development, 58*, 3-18.

Sarwar, G. S. (2011). *Structural assessment of knowledge for misconceptions: Effectiveness of structural feedback provided by Pathfinder networks in the domain of physics*. Köln, Germany: LAP Lambert Academic Publishing.

Sarwar, G. S. (2012). *Comparing the effect of reflections, written exercises, and multimedia instruction to address learners' misconceptions using structural assessment of knowledge* (Doctoral dissertation, University of Ottawa, Ottawa, MD).

Sarwar, G. S., & Trumpower, 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. doi:10.1007/s11423-015-9368-7

Schneider, S., Krieglstein, F., Beege, M., & Rey, G. D. (2021). How organization highlighting through signaling, spatial contiguity and segmenting can influence learning with concept maps. *Computers and Education Open, 2*, 100040. https://doi.org/10.1016/j.caeo.2021.100040.

Schvaneveldt, R. W. (2017). *JPathfinder software*. Retrieved from https://research-collective.com/PFWeb/

Schvaneveldt, R. W. (Ed.). (1990). *Pathfinder associative networks: Studies in knowledge organization.* Norwood, NJ: Ablex Publishing Corp.

Shavelson, R. J. (1972). Some aspects of the correspondence between content structure and cognitive structure in physics instruction. *Journal of Educational Psychology, 63*(3), 225-234. https://doi.org/10.1037/h0032652

Slavin (2019). Developer- and researcher-made measures. Robert Slavin's Blog. Available from https://robertslavinsblog.wordpress.com/2019/10/24/developer-and-researcher-made-measures/

Stevens, J. P. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Swett, K., Miller, A. C., Burns, S., Hoeft, F., Davis, N., Petrill, S. A., & Cutting, L. E. (2013). Comprehending expository texts: The dynamic neurobiological correlates of building a coherent text representation. *Frontiers in Human Neuroscience, 7*, Article 853. https://doi.org/10.3389/fnhum.2013.00853

Taricani, E. M. & Clariana, R.B. (2006). A technique for automatically scoring open-ended concept maps. *Educational Technology Research and Development, 54*, 61-78.

Tawfik, A. A., Kim, K., & Kim, D. (2020). Effects of case library recommendation system on problem solving and knowledge structure development. *Educational Technology Research and Development*, *68*(3), 1329-1353. doi:10.1007/s11423-020-09737-w

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, 87*, 436-449. https://doi.org/10.1016/j.chb.2018.01.035

Teplovs, C., & Scardamalia, M. (2007). *Visualizations for knowledge building assessment*. AgileViz workshop, CSCL 2007 Convention, New Brunswick, NJ, United States. http://citeseerx.ist.psu.edu/viewdoc/summary?doi = 10.1.1.590.1779

Trotskovsky, E., Sabag, N., & Waks, S. (2015). Students' achievements and misunderstandings when solving problems using electronics models—A case study. *IEEE Transactions on Education, 58*(2), 104-109. doi:10.1109/te.2014.2331918

Trumpower, D. L., Filiz, M., & Sarwar, G. S. (2014). Assessment for learning using digital knowledge maps. In D. Ifenthaler & R. Hanewald (Eds.), *Digital knowledge maps in education: Technology enhanced support for teachers and learners* (pp. 221-238). New York: Springer.

Trumpower, D. L., & Sarwar, G. S. (2010). Effectiveness of structural feedback provided by *Pathfinder* networks. *Journal of Educational Computing Research, 43*(1), 7-24.

Ursachi, G., Horodnic, I. A., & Zait, A. (2015). How reliable are measurement scales? External factors with indirect influence on reliability estimators. *Procedia Economics and Finance, 20*, 679-686.

Wang, Yu (2021). *Comparative Investigation of full and focused network feedback on students' knowledge structure and learn*ing. (Doctoral dissertation, Pennsylvania State

University). Electronic Thesis and Dissertations. https://etda.libraries.psu.edu/files/final_submissions/23749

Westerberg, L., Litkowski, E., Finders, J.K., Gerde, H.K., Duncan, R.J., Schmitt, S.A., & Purpura, D.J. (2021). Concurrent predictors of science core knowledge in preschool. *Cognitive Development, 57*, 2021 100981. https://doi.org/10.1016/j.cogdev.2020.100981.

Witherby, A. E., & Carpenter, S. K. (2021). The rich-get-richer effect: prior knowledge predicts new learning of domain-relevant information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(4), 483-498. https://doi.org/10.1037/xlm0000996

Yeari, M., & Lantin, S. (2021). The origin of centrality deficit in text memory and comprehension by poor comprehenders: a think-aloud study. *Reading and Writing, 34*, 595-625. https://doi.org/10.1007/s11145-020-10083-9

Yin, Y., Vanides, J. Ruiz-Primo, M., Ayala, C. & Shavelson, R. (2005). Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use. *Journal of Research in Science Teaching. 42*, 166-184. https://onlinelibrary.wiley.com/doi/abs/10.1002/tea.20049?msockid=2a738517602261883091917e61bb6030

Zhang, F., Kim, K., & Clariana, R. B. (2017). *Graphical Interface of Knowledge Structure, GIKS*. [Computer software]. Available from https://giks.azurewebsites.net/

Zimmerman, W., Kang, H. B., Kim. K., Gao, M., Johnson, G., Clariana, R. B., & Zhang, F. (2017). Computer-automated approach for scoring short essays in an introductory statistics course. *Journal of Statistics Education, 25*, 40-47.