Clariana, R. B., & Solnosky, R. L. (2025). *Instructor-led Structural Knowledge Reflection and Conceptual Structure of Summary Essays*. A presentation at the Annual Meeting of the American Educational Research Association (AERA) conference, Denver, CO, April 2025.

Instructor-led Structural Knowledge Reflection and Conceptual Structure of Summary Essays

<u>Objectives / purposes</u>

Observing conceptual structure transitions from lesson to posttest provides another way to describe learning in terms of a transition towards domain normative knowledge (expertise). This quasi-experimental investigation considers the influence of an instructor-led discussion of structural knowledge on the conceptual structure of summary essays from lesson to posttest. Undergraduate architectural engineering students, after completing the lecture portions on the topic *Sustainability and Green Design*, during lab time composed a 300-word summary essay using the online tool Graphical Interface of Knowledge Structure (GIKS, Authors, 2024, see Figure 1), then immediately one lab section participated in an instructor-led discussion of their group-average essay structure to note correct conceptions as well as common misconceptions, while the other two sections also wrote but did not have this discussion. Posttest essays were collected the following week.

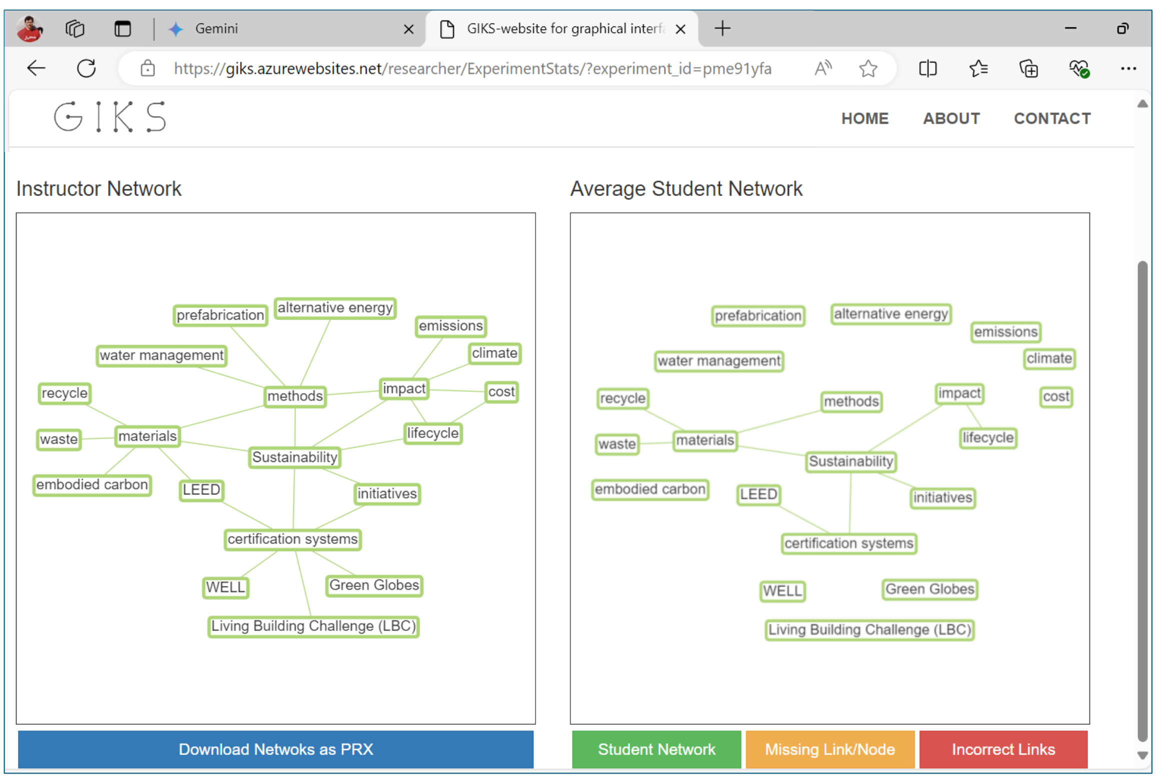

Figure 1. Screen display used by the instructor during the average network discussion.

This investigation compares lesson and posttest essay conceptual structures to determine the influence of an instructor-led discussion of conceptual structure on students' posttest essays. In addition, group-level descriptive data describes the conceptual structure of the lesson and posttest essay networks, the referent networks, and essays generated by Google Gemini as a form of validity of the conceptual structure measures.

Theoretical framework

Conceptual structure is a fundamental aspect of memory organization and thus of learning (Robins et al., 2022). Students attain structure implicitly from conceptual structure inherent in language artifacts such as texts, dialog, and discussions. Trumpower (Trumpower & Goldsmith, 2004; Trumpower & Vanapalli, 2023) proposed structural assessment of knowledge (SAK) to elicit and represent how one organizes knowledge in order to develop strategies to explicitly influence students' conceptual structures.

Reading and writing are intimately tied to conceptual structure (Emig, 1977; Authors, 2022). For example, summarizing lesson content helps students to improve and refine their thinking about complex phenomena (Bereiter & Scardamalia, 1987; Hidi & Anderson, 1986) and helps students to grasp concepts in a related fashion rather than as discrete sets of ideas (Gaskins & Guthrie, 1994; Glynn & Muth, 1994; Guthrie et al., 2004). Writing about scientific topics helps students to understand common disciplinary conceptions and to participate in scientific discursive communities (Wallace, 2004). Additionally, Mason and Boscolo (2004) have identified writing as a way to foster conceptual change, especially for the correction of misconceptions, by encouraging students to develop more elaborated explanations of scientific phenomena (Halim et al., 2018; Moon et al., 2018). Summary writing becomes even more effective with formative feedback and reflection (Bangert-Drowns et al., 2004; National Academies of Sciences, Engineering, and Medicine, 2016).

Measuring the conceptual structure of essays as knowledge graphs is one approach for measuring the content of essays (Authors, 2022; Klebanov & Madnani, 2020; Ramesh & Sanampudi, 2022). This investigation uses the analyses of lexical aggregates approach (ALA-Reader) of Authors (2004; 2024) to provide conceptual structure measures of lesson and of posttest essays by converting the essays into pathfinder networks and then calculating the similarity between the students' essay networks and some referent networks (an expert, a textbook chapter, a Power point lecture).

Methods

Undergraduate students (N = 73) enrolled in the course AE 222 *Building Modeling and Documentation* completed a lesson on sustainability that included two class lecture/discussions, *Sustainability Primer Overview* (Wednesday, Apr. 17) and *Embodied Carbon* (Monday, Apr. 22), and a lab session where students attended one of three sections based on their schedules (see Figure 2). During lab time, all students wrote a summary essay using the GIKS browser-based tool, the essay prompt was:

> *Reflect back on the lessons in the past weeks on "Sustainability and Green Design", then write a 300-word summary of the important ideas from the lectures*

and readings. Here are seven key terms that you could include in your summary: Sustainability and green design, materials, methods, impact, lifecycle, initiatives, and certification systems

Students in the Tuesday lab section (Apr. 23, n = 30) in addition to writing with GIKS participated in a 10-minute instructor-led discussion. The instructor's network and the group's essays average network were displayed side-by-side on the large screen while the instructor pointed out the correct conceptions and the misconceptions in the group's group-average network. The Wednesday and Thursday lab sections (Apr. 24 & 25, ns = 23 & 20) wrote with GIKS but did not see the group-average network and did not discuss it.

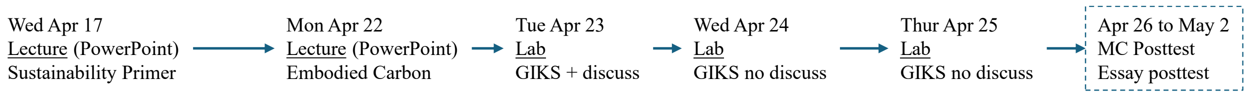

Figure 2. Flow of the investigation.

Beginning on Friday (Apr. 26) for one week outside of class, students were required to complete the end-of-unit test in the Canvas Learning Management System that included rewriting the summary essay on sustainability (word-for-word the same writing prompt as the GIKS lesson essay).

<u>Data sources</u>

The lesson and posttest essays (final n = 62, mortality includes 7 missing lesson essays and 4 missing posttest essays) were scored by the lab assistant and also were converted into Pathfinder Networks (Schvaneveldt et al., 1989; Schvaneveldt, 2024) using *ALA-Reader* software that carries out a forward pass through the essay to pattern match 20 pre-selected key terms in the instructor's expert network that include: *alternative energy, certification systems, climate, cost, embodied carbon, emissions, Green Globes, impact, initiatives, LEED, lifecycle, Living building challenge (LBC), materials, methods, prefabrication, recycle, Sustainability and green design, waste, water management, WELL*. As these key terms are found, a "1" is added to the appropriate cell in a term x term array, thus the essay is rendered into a 20 x 20 array containing "1s" or "0s".

The 20 x 20 arrays are proximity files (.prx file format) that are analyzed using *JPathfinder* software (Schvaneveldt, 2024) in order to both create the networks and then compare the networks. In addition, *JPathfinder* allows the researcher to average raw proximity files in order to create group-average networks. Group-average networks reduce individual idiosyncratic error and so have been shown to have probative value for analysis (Chen et al., 2022; Wei et al., 2024).

Additional data consists of summary essays collected in this course in the previous year (i.e., 2023) using the same essay prompt as used in this current investigation, but some were composed in GIKS and the rest were composed in a word processor (i.e., saved as a PDF file). These 2023 essays were converted to networks to compare to the current lesson essays.

Results

The posttest essays were scored for Quality and for Content by a lab assistant using a 0-to-10 scale. Kruskal-Wallis tests were conducted to examine the rater's scores for the Discussion group (n = 28) and No Discussion group (n = 41) posttest essays. No significant difference was found for Essay Content scores (Chi square = 0.36, p = .55, df = 1) but a *significant difference was found for Essay Quality scores* (Chi square = 4.51, p = .03, df = 1) with a mean rank score of 40.96 for Discussion and 30.93 for no Discussion (i.e., Discussion M = 7.53, SD = 1.10, no Discussion M = 7.00, SD = 1.11).

Next, the similarity between networks is calculated here as *network percent overlap* that is calculated as links in common between two networks divided by the average number of links in the two networks. First, the essay group-average lesson networks for the 2023 essays (N = 75) were compared to those of 2024 (see Figure 3). These essay group-average networks are significantly and substantially alike, with more than 50% similarity in network conceptual structure (e.g., $p < .05$ for values greater than .23).

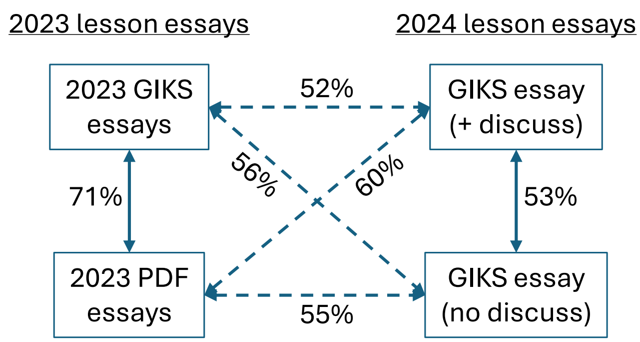

Figure 3. Network percent overlap of the 2023 and 2024 lesson essay group-average networks.

Next the 2024 essay group-average lesson networks and posttest network are compared to each other and to the three referent networks including the Instructor's expert network and the networks of the two PowerPoint (ppt) presentations (see Figure 4). As above, all four of the essay group-average networks are significantly and substantially alike with more than 50% similarity in network conceptual structure, but the lesson-to-posttest essay network percent overlap is greater for the no discussion group (68%) than for the discussion group (63%).

The influence of the *Embedded Carbon lecture/discussion* (EC ppt, Apr. 22) that happened the day before the GIKS lesson essay was NOT observed in any of the four essay group-average networks (values below .24 not significant), however the *Sustainability Primer lecture/discussion* (SP ppt, Apr. 17) that happened the week before the GIKS lesson essay was significantly related to the four essay group-average networks (values greater than 0.24 are significant), especially at posttest for the no discussion essay average-group networks (i.e., 41% overlap with the SP ppt network).

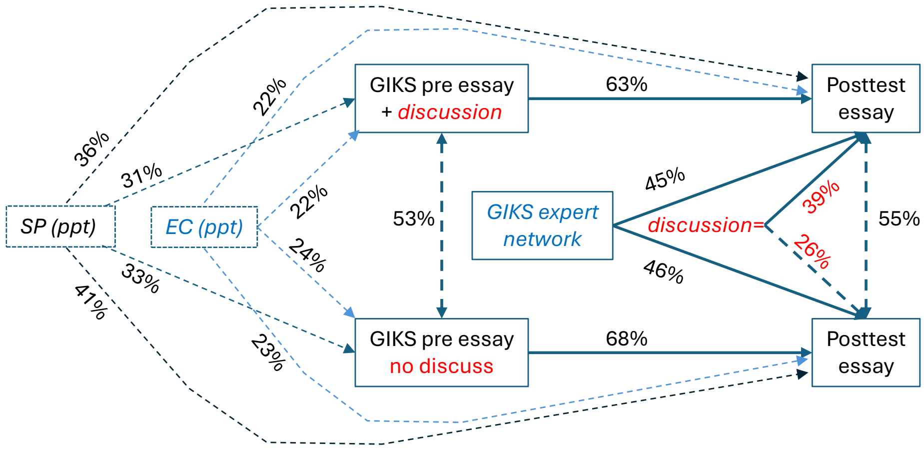

Figure 4. Network percent overlap among the four essay group-average networks and the four referent networks.

And finally, central to this investigation, relative to the GIKS expert network, the posttest essay group-average networks for the discussion and the no discussion groups were not different (45% and 46% percent network overlap with the expert network), the instructor-led discussion apparently did not influence the OVERALL relationship to the expert network; however the discussion group-average network was more like the network of the discussion dialog transcript than was the no discussion group-average network (see Figure 4, 39% vs. 26%). This indicates that the discussion improved a section of the expert network (the central high degree terms) *at the expense of other sections* of the expert's network.

Key term frequency in essays is a bag-of-words approach that is an easily counted metric. Authors (2024) proposed that key term frequency may provide a way to detect plagiarism using AI essay such as Gemini and Open AI (ChatGPT) because these tools non-selectively increase the frequency of terms in the writing prompt, while people usually select key concepts and will often completely disregard some terms in the writing prompt. In this case, in June 2024 Gemini (Goole AI) was used to create 20 essays using the same writing prompt that was given to the students, then the key term frequencies of the 20 key terms in the expert network were used to calculate the frequency of these terms in all of the essays (see Figure 5).

The average term frequencies for the discussion group and no discussion group essays were very similar across all 20 key terms, while the AI essays relative to the students' essays overinflated the use of 5 of the 7 terms that were given in the writing prompt and deflated the use 9 of the 13 terms that were not given in the writing prompt. This difference provides a pattern or "fingerprint" of possible AI essays submitted by students, especially overuse of the terms "impact" and "green design" and the underuse for instance of "WELL" and "prefabrication". Comparison of student essay frequencies to the AI essays based on the 7 key term frequencies data greater that r=.80 *suggests* that 9 students may have used AI during the lesson and 1 may have used AI during the posttest, but there is no way to know this for sure.

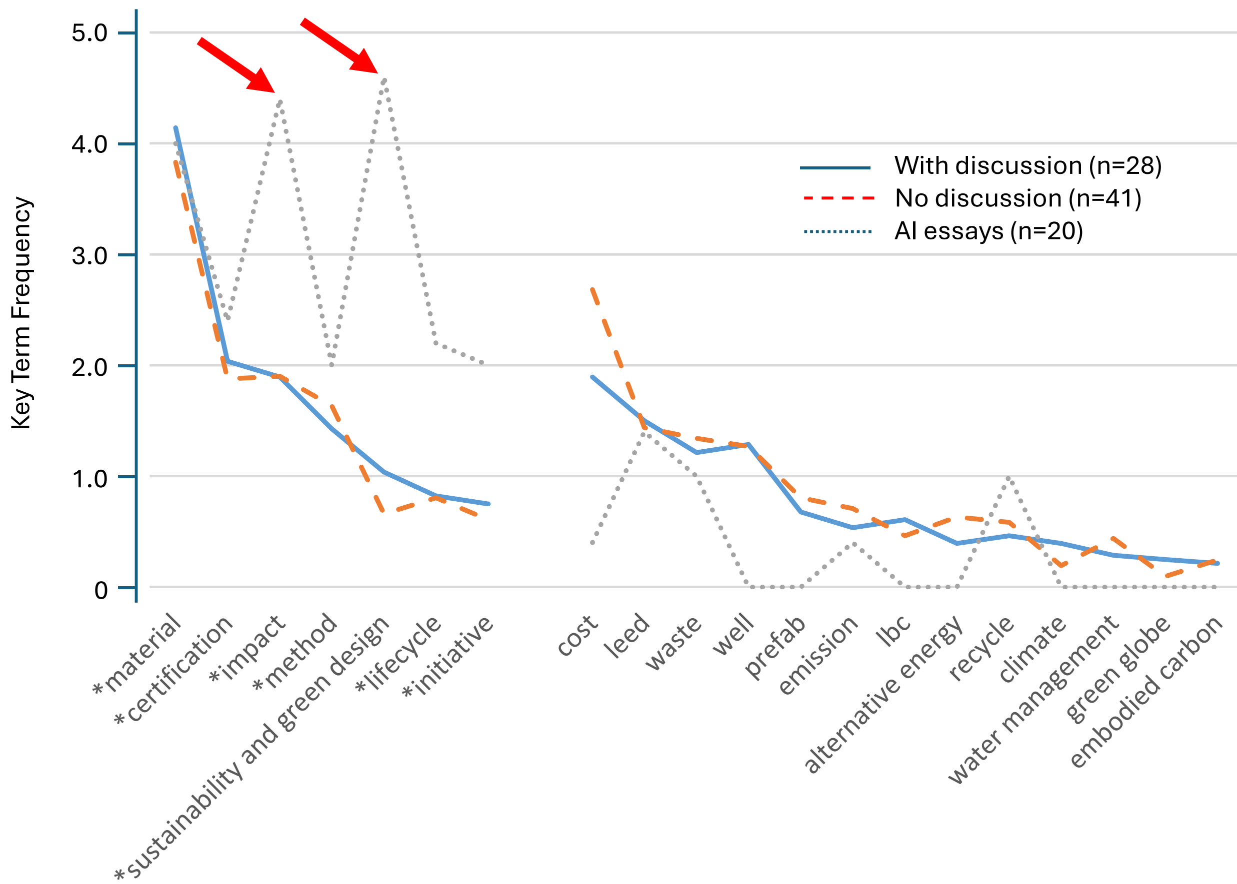

Figure 5. Average posttest essay term frequencies of the 20 expert network terms for the discussion group, the no discussion group, and Gemini AI essays.

Regarding the content conceptual structure quality of the AI essays, the Gemini essay group-average network percent overlap with the expert network is 0.34 (compared to 45% and 46% like the expert for the students' essays), but when the AI prompt is expanded to include all 20 expert terms, the Gemini essay group-average network percent overlap with the expert network is 0.43 which is about the same percent overlap as observed for the students' essay networks.

<u>Scientific or scholarly significance</u>

The instructor-led discussion of the networks <u>did improve</u> posttest essay quality (human rater) relative to no discussion. Also, the data indicates that the discussion did alter students' conceptual structures of the central terms in the expert network, but at the expense of peripheral, unmentioned terms. Therefore instructor-led discussion of content conceptual structure likely does influence students' conceptual knowledge structures, and teachers and instructors must be vigilant in preparing and presenting such a discussion to make sure they appropriately and adequately cover the content.

Theoretically, it is important to note that these group-level measures of conceptual structure are very similar within a group of comparable students, to similar students from previous years, and to Gemini AI essays. This supports the notion of both local and of global collective knowledge structures that are also foundational to large language AI models (Authors, 2022).

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.

Bereiter, C., & Scardamalia, M. (1987). An attainable version of high literacy: Approaches to teaching higher-order skills in reading and writing. *Curriculum Inquiry*, *17*(1), 9-30.

Chen, X., Li, Z., et al. (2022). The influence of the conceptual structure of external representations when relearning history content. *Educational Technology Research and Development*, *71*, 415–439. https://doi.org/10.1007/s11423-022-10176-y

Emig, J. (1977). Writing as a mode of learning. *College Composition and Communication, 28*(2), 122-128.

Gaskins, I. W., Guthrie, J. T., Satlow, E., Ostertag, J., Six, L., Byrne, J., & Connor, B. (1994). Integrating instruction of science, reading, and writing: Goals, teacher development, and assessment. *Journal of Research in Science Teaching*, *31*(9), 1039-1056.

Glynn, S. M., & Muth, K. D. (1994). Reading and writing to learn science: achieving scientific literacy. *Journal of Research in Science Teaching*, *31*(9), 1057-1073.

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., & Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology*, *96*(3), 403-423.

Halim, A. S., Finkenstaedt-Quinn,S. A., LauOlsen, L. J., Gere, A. R., & Shultz, G. V. (2018). Identifying and remediating student misconceptions in introductory Biology via Writing-to-Learn assignments and peer review. *Life Sciences Education, 17* (28), 1-12.

Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research, 56* (4), 473-493.

Klebanov, B. B., & Madnani, N. (2020). Automated evaluation of writing – 50 years and counting. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 7796–7810). July 5 - 10, 2020. https://aclanthology.org/2020.acl-main.697/

Mason, L., & Boscolo, P. (2004). Role of epistemological understanding and interest in interpreting a controversy and in topic-specific belief change. *Contemporary Educational Psychology*, *29*(2), 103-128.

Moon, A., Zotos, E., Finkenstaedt-Quinn, S., Gere, A. R., & Shultz. G. (2018). Investigation of the role of writing-to-learn in promoting student understanding of light-matter interactions. *Chemistry Education Research and Practice, 19*, 807-818.

National Academies of Sciences, Engineering, and Medicine. (2016). *Barriers and Opportunities for 2-Year and 4-Year STEM Degrees: Systemic Change to Support Diverse Student Pathways*. Washington, DC: The National Academies Press. doi: 10. 17226/21739.

Ramesh, D., & Sanampudi, S.K. (2022). An automated essay scoring systems: a systematic literature review. Artificial Intelligence Review, 55, 2495-2527. https://doi.org/10.1007/s10462-021-10068-2

Robins, S., Aronowitz, S. & Stolk, A. (2022). Memory structure and cognitive maps (Chapter 11). In F. De Brigard & W. Sinnott-Armstrong (Eds.), *Neuroscience and Philosophy* (Chapter 11). MIT Press. https://www.ncbi.nlm.nih.gov/books/NBK583710/

Schvaneveldt, R. W. (2024). JPathfinder software. Downloaded form https://research-collective.com/PFWeb/Download.html

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). *Network structures in proximity data. In Psychology of learning and motivation* (Vol. 24, pp. 249-284). Elsevier. https://doi.org/10.1016/S0079-7421(08)60539-3

Trumpower, D. L., & Goldsmith, T. E. (2004). Structural enhancement of learning. *Contemporary Educational Psychology, 29*, 426–446.

Trumpower, D.L., & Vanapalli, A.S. (2023). Structural Assessment of Knowledge as, of, and for Learning. In J.M. Spector, B.B. Lockee, M.D. Childress (Eds.*), Learning, Design, and Technology*. Springer. https://doi.org/10.1007/978-3-319-17461-7_23

Wallace, C. S. (2004). Framing new research in science literacy and language use: Authenticity, multiple discourses, and the “Third Space”. *Science Education*, *88*(6), 901-914.

Wei, Z., Zhang, Y., et al. (2024). The effects of reading prompts and of post-reading generative learning tasks on multiple document integration: evidence from concept network analysis. *Educational Technology Research and Development, 72*, 661-685. https://doi.org/10.1007/s11423-023-10326-w