

# Concept Mapping in Biology Education: A Systematic Review and Meta-Analysis

Aaron Wenger

2023-09-21

This is where the abstract will go

Concept mapping (CM) is an instructional activity invented in the 1970s by Novak and his Cornell University research team (Novak & Cañas, 2007). A concept map consists of conceptual nodes with connecting verbal links (see Figure 1). Each node-link-node connection forms a proposition. CM has been applied in a variety of ways and for a variety of instructional purposes such as collaborative learning, group discussion, directed reading, and formative assessment. With the move of classes to online settings during the COVID-19 pandemic, CM has received even more attention in science education as a tool for active learning in virtual settings (Choe, 2020; Gerber Hornink & Costa, 2021).

Over the last 40 years, classroom intervention studies with CM have proliferated and their results vary widely, often conflicting. Based on the results of recent meta-analyses and reviews, it is apparent that there is unexplained variance in the observed effects of CM in instructional contexts. It is an open question whether some of this variance is due to the influence of factors which affect the efficacy of CM, of which we know little (Kinchin, 2014; Pudenko et al., 2012; Schroeder et al., 2018), or other study-specific factors (e.g., study design, comparison condition).

The present work is a systematic review and meta-analysis with the primary purpose of learning more about the factors that influence the effect size of CM interventions in empirical studies. As the use of CM in education is a large, diverse, and complex topic, this project focuses on CM use in biology education. The results are used to draw conclusions on the efficacy of CM in biology education and to provide guidance for future research.

## Concept Mapping in Context

Proponents of various “concept mapping”, “mind mapping”, and “knowledge mapping” activities have promoted their innovations – each with some success – to the education profession

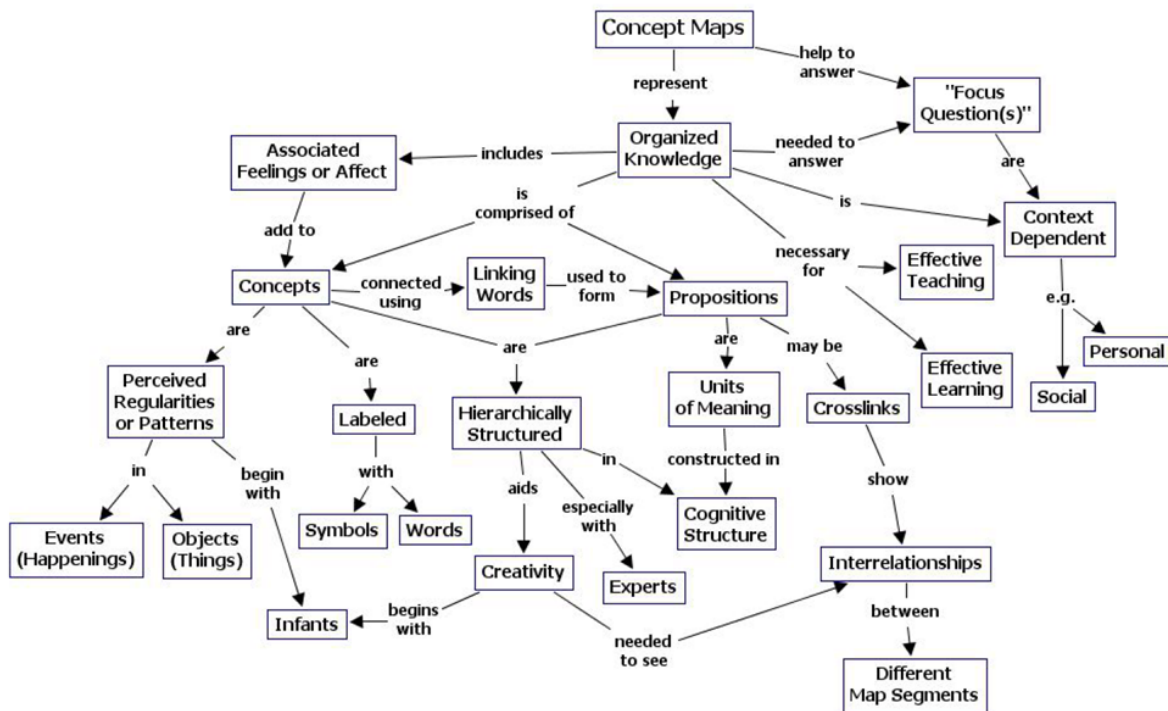


Figure 1: An example concept map with its key features, from Novak & Cañas (2007)

and the general public. As a result, a confusion of terms has set in which creates difficulties for research on CM (Åhlberg, 2004; Kinchin, 2014). The brief description of CM given in the introduction – as a diagram of conceptual nodes with connecting verbal links – serves as the definition of CM for this project. In contrast, mind mapping is a simple association method without a definitive structure or the use of verbal links (Åhlberg, 2004). Knowledge mapping is very similar to CM except that it uses a definite set of terms for links between nodes (O'Donnell et al., 2001).

CM was developed by researchers as a means of representing children's science knowledge (Novak & Cañas, 2007) and has been used by researchers in other contexts and applications as an instrument or method. For example, CM has been used to measure conceptual change (Wallace & Mintzes, 1990) and as a method of planning program evaluations (Trochim, 1989) and dissertation projects (Donnelly, 2017). Early proponents of CM soon applied CM to instructional purposes as a means of promoting Ausubel's meaningful learning construct via the process of assimilation – integration of new concepts into existing conceptual frameworks (Novak & Cañas, 2007).

The instructional uses of CM are varied. From the earliest years, CM was recommended as a learning and formative assessment tool (J. Novak, 1981). Others applied CM as an advance organizer – an overview of and bridge between the student's prior knowledge and content to be learned (Willerman & Mac Harg, 1991). CM has been used in both a collaborative group setting and with individual learners (Gao et al., 2007), with students constructing their own maps or studying teacher-constructed maps (Dyer et al., 2015). Lastly, CM has been applied extensively in higher education (Kinchin, 2014), but also in K-12 schooling (Schroeder et al., 2018).

While CM has been used as an instructive tool in many school subjects, its first use was in biology education. The first article promoting the instructional use of CM was published in *The American Biology Teacher* (Stewart et al., 1979). Some researchers believe that CM can help students summarize course content using the large vocabulary required in introductory biology courses, thus promoting meaningful rather than rote learning (Javonillo & Martin-Dunlop, 2019). As stated by Schmid & Telaro (1990), "Concept mapping ... appears to be ideally suited to address biology content." (78-79).

## **Theoretical Framework for the Present Study**

Novak worked closely with David Ausubel and later with Bob Gowin (an American philosopher of education at Cornell) coauthoring a book with each (Ausubel et al., 1978; Novak & Gowin, 1984). Ausubel proposed the theory of meaningful learning which would later be closely associated with schema theory and cognitive information processing (Driscoll, 2005). Gowin invented the Vee heuristic as a means of making knowledge construction explicit, starting with objects/events and applying concepts, theories, and practices to build up to knowledge claims (Novak & Gowin, 1984).

Concept mapping has either been framed as an implementation of either cognitive information processing models or constructivist theories of learning. Novak adopted his own subtype of constructivism, “human constructivism”, which was his attempt at unifying phenomena in psychology and epistemology. In his first proposal of human constructivism, Novak includes CM as a means of representing cognitive structures and of demonstrating the constructivist process of learning (Novak, 1993). In contrast, the work of Karpicke and colleagues has considered CM as an elaborative encoding tool which may be effective to the extent that memory is retrieved, processed in the mapping task, and re-encoded (Blunt & Karpicke, 2014; Karpicke & Blunt, 2011).

The competing explanatory frameworks of Karpicke and Novak demonstrates that there is room for differing interpretations of the efficacy of CM and of relevant moderating factors. The present study adopts the perspective of constructivism as this is the most prevalent theoretical framework in CM intervention research. Novak and his colleagues consider human constructivism to be a moderate epistemological position between logical-positivism and social or radical constructivism (Mintzes et al., 1998). They summarized it in three statements:

- Human beings are meaning makers.
- The goal of education is to construct shared meanings.
- The construction of shared meanings may be facilitated by well-prepared teachers.

## Previous Meta-Analyses

A series of systematic reviews have examined the CM literature each with a focus on a particular aspect or application of CM [e.g., Hartmeyer et al. (2018), Machado & Carvalho (2020), and Martin et al. (2015)]. A number of meta-analyses have also been conducted on the effects of CM interventions. Most recently, Barta et al. (2022) synthesized the metacognitive (i.e., critical thinking) and affective effects of CM interventions across 21 studies. Yue et al. (2017) similarly synthesized metacognitive effects but limited their scope to studies in medical education. Erdoğan (2016) synthesized cognitive outcomes (i.e., learning gains or academic achievement) but limited their scope to include only studies in Turkey.

Nesbit & Adesope (2006) have conducted the largest and most comprehensive meta-analysis of cognitive and affective outcomes which was updated in 2018 to include over 142 independent effect sizes (ES) from 118 studies (Schroeder et al., 2018). They included all experimental and quasi-experimental studies that contrasted CM with other learning activities regardless of pedagogical setting or academic discipline. Their analysis yielded an overall mean ES of 0.58 with an I<sup>2</sup> of 87.5%, after adjusting the value of two ES identified as outliers. They conducted moderator analyses with five categorical variables using the full set of independent effect sizes. They then divided the dataset by CM type (constructed or studied) and conducted moderator analyses by the four remaining variables plus three additional categorical variables.

The largest differences in CM effects were seen for different comparison conditions, region, and level of student interaction.

No other statistically significant and potentially meaningful differences were seen except for duration of the intervention. The results of the moderator analyses are shown in Table 1.

Table 1  
Moderator Analysis Results by CM Activity Type

Moderator		Constructed		Studied	
		k	g	k	g
NA	Overall	75	0.72	67	0.43
Comparison condition	Discussion/lecture	32	1.05	5	1.09
	Studied or constructed lists	0	-	13	0.43
	Studied or constructed outlines	6	0.4	2	0.72
	Studied text	5	0.33	39	0.29
	Constructed text	10	0.48	3	0.1
	Other	22	0.47	5	0.98
Academic discipline	STEM	64	0.73	54	0.44
	non-STEM	11	0.62	12	0.41
	Not reported	0	-	1	0.05
CM mode	Interactive	8	0.71	16	0.54
	Static	66	0.72	39	0.4
	Animated	0	-	7	0.47
	Mixed	1	0.75	5	0.27
Intervention duration	<1 week	14	0.4	33	0.34
	1-4 weeks	23	0.94	30	0.48
	>4 weeks	37	0.72	4	0.7
	Unknown	1	0.06	0	-
Country	Africa	7	1.44	0	-
	Asia	9	0.78	5	1.04
	Europe	9	0.82	3	0.46
	Middle East	13	0.75	2	0.96
	USA or Canada	33	0.49	51	0.25
	Other/not reported	4	0.62	6	1.29
School level	Intermediate	22	0.68	7	0.82
	Secondary	25	0.74	4	1.24
	Postsecondary and beyond	28	0.73	56	0.32
Student interaction	In groups	14	0.91	10	0.48
	Individual	32	0.55	55	0.41

Mixed	22	0.91	0	-
Other	2	0.95	1	0.75
Unknown	5	0.29	1	0.47

Results from Schroeder et al. (2018)

## Potential Moderators of CM Efficacy

In this section, the variables investigated in the present study are described and related to Novak’s human constructivism and other considerations from the research synthesis literature. Variables were sorted into four groups according to their significance to CM theory and practice. This scheme was derived from the ideas of Lipsey (2019) and the MUTOS framework of Becker (2017).

### Extrinsic Characteristics

Extrinsic characteristics are those characteristics of included studies which concern the researcher or research process. Differences in mean effect sizes by publication status (i.e. journal article or dissertation/thesis) are known and described as a type of publication bias (Cooper et al., 2019). In an empirical investigation of youth psychotherapy trials, McLeod & Weisz (2004) found that studies reported in dissertations have a lower mean effect size than studies reported in journal articles. Variation in publication year may be related to characteristics of the methodology or intervention as practices have evolved. It is likely that such variation exists given the age of CM research as a research topic and instructional practices and so publication year may moderate effects. The country or region where the study was conducted is included as an extrinsic characteristic. While the country or region where a study was conducted may serve as an indicator of socio-cultural factors, it was judged that it is more likely associated with extrinsic or methodological factors than intervention or sample factors. Thus, region was included as an extrinsic factor.

### Methodological Characteristics

In randomized controlled trials (RCTs) and quasi-experimental designs (QEDs) a control or comparison condition is used to remove the influence of factors unrelated to the treatment. Schroeder et al. (2018) coded a variety of comparison conditions and concluded that larger CM effects are associated with lecture and discussion comparisons as opposed to various review activities such as creating or studying lists or texts. It is possible that CM is both more effective than “business-as-usual” (BAU) methods – such as lecturing – and less effective than other evidence-based methods – such as retrieval practice (see Karpicke & Blunt, 2011).

A particular concern that some scholars have with studies using experimental designs is that important contextual factors will be controlled or ignored, thus leaving the researcher ignorant of their importance to the CM intervention (Kinchin, 2014). Some studies take place in a controlled laboratory setting (e.g., Karpicke & Blunt, 2011) while others assign whole course sections to each condition. In line with general constructivist thought, CM proponents expect that findings from laboratory settings will not generalize to classroom instruction, perhaps underestimating the efficacy of CM (Mintzes et al., 2011; J. D. Novak & Gowin, 1984).

## **Intervention Characteristics**

CM interventions may require students to construct their own maps or study map(s) constructed by the interventionist. Also, CM interventions may involve interactive or animated maps facilitated by a digital CM tool such as CmapTools (Cañas et al., 2004) or static maps created and presented digitally or on a writing surface. While student construction of concept maps is associated with larger effects, the mode of implementation (interactive, animated, or static) is not associated with different effects (Schroeder et al., 2018).

CM interventions where students construct maps may involve an initial training period where students become familiar with the process as recommended by CM proponents (Kinchin, 2011; Mintzes et al., 2011; J. D. Novak & Gowin, 1984). However, the extent of training may vary from minutes to weeks (Karpicke & Blunt, 2011b). It is not known what effect, if any, that training has in CM efficacy.

The duration of CM interventions vary considerably from less than one week to more than four weeks. In line with the importance of CM training, some believe that CM interventions of a longer duration will have more pronounced or long-lasting effects (Mintzes et al., 2011). Schroeder et al. (2018) concluded that longer intervention durations are associated with larger effects.

Different strands of constructivism differ on the importance of student social interaction in learning. It appears uncommon to directly examine a CM condition with and without cooperative student interaction as Okebuloa (1992) did or to state a clear rationale for designing an intervention. Schroeder et al. (2018) examined the level of collaboration between learners as a moderator of CM efficacy and found substantial differences between coded levels, though this was not statistically significant.

## **Sample Characteristics**

As mentioned above, CM has been applied at different grade levels, across school subjects, and in varying contexts. A large majority of CM research appears to have been conducted in STEM subjects rather than non-STEM subjects, though there is no statistically significant or meaningful difference between the two. Likewise, CM research has been conducted across different grade levels. Per the conclusions of Schroeder et al. (2018), there doesn't seem to be

a substantial or statistically significant difference from intermediate to postsecondary grade levels.

## Rationale and Research Questions

The work of Schroeder et al. (2018) is excellent for its contextualization of CM within educational theories and in its discussion of theory vs. evaluation-orientated research. However, it falls short in its meta-analytic methods in important points. First, in calculating effect sizes (ES) for studies with multiple CM conditions, they calculated synthetic ESs which averaged the means and standard deviations of each CM condition. In cases where outcome measures are reported at multiple time points, Schroeder et al. deferred to the last measurement to calculate the ES. Both ES extraction decisions avoid combining dependent ESs in the meta-analysis but at the cost of lower statistical power, a lesser ability to explain heterogeneous effects, and the potential for bias (COOPER, 2019). They can be avoided using multilevel and/or multivariate meta-analyses (Tipton et al., 2019; Van den Noortgate et al., 2015).

Second, Schroeder et al. do not distinguish between cognitive and affective (referred to as motivational) outcomes in their analysis. Instead, they include both outcomes in a univariate meta-analytic model which makes it difficult if not impossible to make meaningful interpretations of their results. It is conceivable, if not probable, that different socio-cultural and cognitive factors are important for each outcome type. Also, the measures for each outcome type can be expected to differ in important ways (e.g., recall or problem-solving behavior vs. self-report likert items). If these outcome types are to be included in the same meta-analysis, each outcome could be analyzed simultaneously using a multivariate model.

Third, Schroeder et al. identified outliers using a formal method, but one based on unweighted ES which does not take variances or the effect of moderators into account. Smaller studies with larger variances may, by chance, have large ES estimates and have little effect on estimation of the meta-analytic model. Schroeder et al. also did not examine these identified outliers or conduct sensitivity analyses to determine the robustness of their findings. Viechtbauer and Cheung (2010) recommend the use of leave-1-out diagnostic statistics (such as studentized deleted residuals or an analog of Cook's distances) to identify outliers or influential cases and the use of sensitivity analyses in all meta-analytic procedures.

Fourth, Schroeder et al. used subgroup analysis to investigate all moderators of CM effectiveness using only one variable (CM type) in multi-way analyses, reexamining all other moderators. This causes three problems: 1) inflated family-wise type I error, 2) no estimation of residual or unexplained heterogeneity, and 3) an inability to conduct statistical comparisons between moderators. It has long been known that meta-regression using categorical covariates is equivalent to subgroup analysis but also allows for residual heterogeneity to be estimated and for statistical comparisons to be made (Thompson & Higgins, 2002). As recommended by Tipton, Pustejovsky, and Ahmadi (2019), meta-regression models can include many moderator variables controlling family-wise type I error and partialling out all moderator effects.



In this meta-analysis, these limitations were addressed using current best practice in meta-analysis. The use of meta-regression with multilevel-mixed-effects models containing relevant moderators allowed the following research questions to be investigated:

- RQ1. What is the mean effect size for CM interventions in biology education and to what extent does it vary within and between studies?
- RQ2. After controlling for study-level, extrinsic and methodological characteristics, to what extent does the mean effect size vary within and between studies?
- RQ3. What is the relationship between effect size and characteristics of the intervention and sample for CM interventions in biology education controlling for study-level, extrinsic and methodological characteristics?

## Methods

As with all systematic reviews, this study started with a systematic literature search and screen for relevant records of research as defined by criteria that addressed the research questions. These records and the studies they reported were coded according to a prespecified scheme and statistical information for ES calculation were extracted. The results were then analyzed in multi-level, meta-regression models which took into account the hierarchical structure of the data with multiple effects often reported for each study. Influential cases were identified and sensitivity analyses were conducted to check the robustness of the results to them and to assumptions made during ES calculation. All analyses were run in R () principally using the ‘metafor’ () and ‘clubSandwich’ packages. All analysis code and meta-analysis data are available at ———.

## References

- Åhlberg, M. (2004). Varieties of Concept Mapping. *Proceedings of the First International Conference on Concept Mapping*, 4.
- Ausubel, D. P., Novak, J. D., & Hanesian, H. (1978). *Educational psychology: A cognitive view* (2d ed). Holt, Rinehart and Winston.
- Barta, A., Fodor, L. A., Tamas, B., & Szamoskozi, I. (2022). The development of students critical thinking abilities and dispositions through the concept mapping learning method – A meta-analysis. *Educational Research Review*, 37, 100481. <https://doi.org/10.1016/j.edurev.2022.100481>
- Becker, B. J. (2017). Improving the Design and Use of Meta Analyses in Evidence-Based Practice. In J. P. Sampson, E. Bullock-Yowell, V. C. Dozier, D. S. Osborn, & J. G. Lenz (Eds.), *Integrating Theory, Research, and Practice in Vocational Psychology: Current Status and Future Directions*. Florida State University Libraries. <https://doi.org/10.17125/svp2016.ch10>

- Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, 106(3), 849–858. <https://doi.org/10.1037/a0035934>
- Choe, K. P. (2020). Synchronous and collaborative online concept mapping of membrane transport. *Biochemistry and Molecular Biology Education*, 48(5), 516–517. <https://doi.org/10.1002/bmb.21430>
- Cooper, H. M., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *Handbook of research synthesis and meta-analysis* (3rd Edition). Russell Sage Foundation.
- Donnelly, J. P. (2017). A systematic review of concept mapping dissertations. *Evaluation and Program Planning*, 60, 186–193. <https://doi.org/10.1016/j.evalprogplan.2016.08.010>
- Driscoll, M. P. (2005). *Psychology of learning for instruction* (3. ed). Pearson Allyn and Bacon.
- Dyer, J.-O., Hudon, A., Montpetit-Tourangeau, K., Charlin, B., Mamede, S., & Van Gog, T. (2015). Example-based learning: Comparing the effects of additionally providing three different integrative learning activities on physiotherapy intervention knowledge. *BMC Medical Education*, 15(1), 37. <https://doi.org/10.1186/s12909-015-0308-3>
- Erdoğan, Y. (2016). An Investigation of the Effectiveness of Concept Mapping on Turkish Students' Academic Success. *Journal of Education and Training Studies*, 4(6), 1–9. <https://doi.org/10.11114/jets.v4i6.1380>
- Gao, H., Shen, E., Losh, S., & Turner, J. (2007). *A Review of Studies on Collaborative Concept Mapping: What Have We Learned About the Technique and What Is Next?* 14.
- Gerber Hornink, G., & Costa, M. J. (2021). Constructing online concept maps in CMap Cloud collaboratively: Connecting pathways in case scenarios. *Biochemistry and Molecular Biology Education*, 49(1), 29–31. <https://doi.org/10.1002/bmb.21481>
- Hartmeyer, R., Stevenson, M. P., & Bentsen, P. (2018). A systematic review of concept mapping-based formative assessment processes in primary and secondary science education. *Assessment in Education: Principles, Policy & Practice*, 25(6), 598–619. <https://doi.org/10.1080/0969594X.2017.1377685>
- Javonillo, R., & Martin-Dunlop, C. (2019). Linking Phrases for Concept Mapping in Introductory College Biology. *Journal of College Biology Teaching*, 45(3), 5.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. *Science*, 331(6018), 772–775. <https://doi.org/10.1126/science.1199327>
- Kinchin, I. M. (2014). Concept Mapping as a Learning Tool in Higher Education: A Critical Analysis of Recent Reviews. *The Journal of Continuing Higher Education*, 62(1), 39–49. <https://doi.org/10.1080/07377363.2014.872011>
- Lipsey, M. W. (2019). Identifying Potentially Interesting Variables and Analysis Opportunities. In H. M. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *Handbook of research synthesis and meta-analysis* (3rd Edition). Russell Sage Foundation.
- Machado, C. T., & Carvalho, A. A. (2020). Concept Mapping: Benefits and Challenges in Higher Education. *The Journal of Continuing Higher Education*, 68(1), 38–53. <https://doi.org/10.1080/07377363.2020.1712579>
- Martin, L. G., Martin, F. A., & Southworth, E. (2015). A Critical Review of Concept Mapping Research Literature: Informing Teaching and Learning Practices in GED Preparation

- Programs. *New Horizons in Adult Education and Human Resource Development*, 27(3), 27–45. <https://doi.org/10.1002/nha3.20109>
- McLeod, B. D., & Weisz, J. R. (2004). Using Dissertations to Examine Potential Bias in Child and Adolescent Clinical Trials. *Journal of Consulting and Clinical Psychology*, 72(2), 235–251. <https://doi.org/10.1037/0022-006X.72.2.235>
- Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (1998). *Teaching science for understanding: A human constructivist view*. Academic Press.
- 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>
- Novak, J. (1981). Applying Learning Psychology and Philosophy of Science to Biology Teaching. *The American Biology Teacher*, 43(1), 12–42. <https://doi.org/10.2307/4447108>
- Novak, J. D. (1993). Human constructivism: A unification of psychological and epistemological phenomena in meaning making. *International Journal of Personal Construct Psychology*, 6(2), 167–193. <https://doi.org/10.1080/08936039308404338>
- Novak, J. D., & Cañas, A. J. (2007). Theoretical Origins of Concept Maps, How to Construct Them, and Uses in Education. *Reflecting Education*, 15.
- Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn*. Cambridge University Press.
- O'Donnell, A. M., Dansereau, D. F., & Hall, R. H. (2001). Knowledge Maps as Scaffolds for Cognitive Processing. *Educational Psychology Review*, 16.
- Pudelko, B., Young, M., Vincent-Lamarre, P., & Charlin, B. (2012). Mapping as a learning strategy in health professions education: A critical analysis. *Medical Education*, 46(12), 1215–1225. <https://doi.org/10.1111/medu.12032>
- Schmid, R. F., & Telaro, G. (1990). Concept Mapping as an Instructional Strategy for High School Biology. *The Journal of Educational Research*, 84(2), 78–85. <https://doi.org/10.1080/00220671.1990.10885996>
- Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and Constructing Concept Maps: A Meta-Analysis. *Educational Psychology Review*, 30(2), 431–455. <https://doi.org/10.1007/s10648-017-9403-9>
- Stewart, J., Van Kirk, J., & Rowell, R. (1979). Concept Maps: A Tool for Use in Biology Teaching. *The American Biology Teacher*, 41(3), 171–175. <https://doi.org/10.2307/4446530>
- Trochim, W. M. K. (1989). An Introduction to Concept Mapping for Planning and Evaluation. *Evaluation and Program Planning*, 12, 1–16.
- Wallace, J. D., & Mintzes, J. J. (1990). The concept map as a research tool: Exploring conceptual change in biology. *Journal of Research in Science Teaching*, 27(10), 1033–1052. <https://doi.org/10.1002/tea.3660271010>
- Willerman, M., & Mac Harg, R. A. (1991). The concept map as an advance organizer. *Journal of Research in Science Teaching*, 28(8), 705–711. <https://doi.org/10.1002/tea.3660280807>
- Yue, M., Zhang, M., Zhang, C., & Jin, C. (2017). The effectiveness of concept mapping on development of critical thinking in nursing education: A systematic review and meta-analysis. *Nurse Education Today*, 52, 87–94. <https://doi.org/10.1016/j.nedt.2017.02.018>