Article

Concept Maps for Improved Science Reasoning and Writing: Complexity Isn't Everything

Jason E. Dowd, Tanya Duncan, and Julie A. Reynolds

Department of Biology, Duke University, Durham, NC 27708

A pervasive notion in the literature is that complex concept maps reflect greater knowledge and/or more expert-like thinking than less complex concept maps. We show that concept maps used to structure scientific writing and clarify scientific reasoning *do not* adhere to this notion. In an undergraduate course for thesis writers, students use concept maps instead of traditional outlines to define the boundaries and scope of their research and to construct an argument for the significance of their research. Students generate maps at the beginning of the semester, revise after peer review, and revise once more at the end of the semester. Although some students revised their maps to make them more complex, a significant proportion of students simplified their maps. We found no correlation between increased complexity and improved scientific reasoning and writing skills, suggesting that sometimes students simplify their understanding as they develop more expert-like thinking. These results suggest that concept maps, when used as an intervention, can meet the varying needs of a diverse population of student writers.

INTRODUCTION

Concept maps are a pedagogical tool that allows learners to construct a visual representation of their understanding of connections between concepts (Novak, 1990). Over recent decades, an increasing amount of scientific literature has focused on concept maps (Nesbit and Adesope, 2006, 2013). Although the applications of concept maps for student learning are extraordinarily wide-ranging, student-generated maps tell us the most about the development of students' reasoning (Novak, 2005).

In science courses, the primary use of student-generated concept maps is to improve content knowledge. The two primary research questions explored in the literature are: 1) Are concept map activities associated with improved

learning outcomes? 2) Do students' concept maps—more specifically, structural aspects of students' concept maps—correspond to learning outcomes? In response to the first question, many studies show that students' creation of maps that represent their understanding is associated with improved learning (Novak *et al.*, 1983; Canas *et al.*, 2003; Nesbit and Adesope, 2006; Wu *et al.*, 2012; Lee *et al.*, 2013; Bramwell-Lalor and Rainford, 2014). In response to the second question, although there is no consensus about how to "best" assess concept maps, the prevailing message in the literature is that *structural complexity increases with expertise* (Novak *et al.*, 1983; Wallace and Mintzes, 1990; Markham *et al.*, 1994; McClure *et al.*, 1999; Mintzes and Quinn, 2006; Srinivasan *et al.*, 2008). Markham *et al.* (1994) show that more senior students use more sophisticated concepts in their maps, suggesting that complexity and sophistication in understanding may be related.

Beyond the sciences, many studies suggest that student participation in concept-mapping activities is associated with improved writing (Cliburn, 1990; Reynolds and Hart, 1990; Floden, 1991; Hyerle, 1995; Anderson-Inman and Horney, 1996; Osman-Jouchoux, 1997; Crane, 1998; Brodney *et al.*, 1999; Osmundson *et al.*, 1999; Gouli *et al.*, 2003; Vanides *et al.*, 2005; Conklin, 2007). Concept mapping has also been studied as a productive intervention to improve critical thinking when reading and learning a second language (Khodaday and Ghanizadeh, 2011) and as a form of prewriting in the liberal arts (Murray, 1978; Giger, 1995;

Yin and Shavelson, 2008). However, despite a number of studies that have focused on the role of concept mapping in improving writing skills, we find only one study in which students' concept maps were directly assessed and correlated with writing assessment in science (Conklin, 2007). In Conklin's study, both concept maps and writing were assessed holistically using rubrics to rate concepts and interrelatedness in concept maps and content and organization in writing. Students' content and concept scores, as well as their interrelatedness and organization scores, were positively correlated, though such holistic assessment is different from some of the more widely discussed structural approaches for evaluating concept maps in the sciences. Specifically, structural approaches tend to involve counting various features (propositions, hierarchies, cross-links, etc.) on concept maps, as opposed to making a judgment about the map as a whole (e.g., using a Likert scale).

This begs the question: When concept map activities are used to improve writing in the sciences, are the traditional means of assessment informative? Does structural complexity in maps relate to students' learning outcomes? In this study, we address this question by directly assessing students' scientific reasoning in writing undergraduate honors theses as the primary learning outcome and relating that assessment to structural aspects of students' concept maps generated to facilitate writing. Additionally, we investigate how students' concept maps change as they develop more expertise in scientific reasoning and writing through the semester, further exploring the link between structural complexity in maps and learning outcomes. One might predict that students who make more connections among concepts and include more concepts in their maps have a more thorough understanding of their research subject and are better equipped to argue for the significance of their work, interpret their results, and discuss the implications. This would result in a positive correlation between structural features and assessment of learning outcomes related to expertise, as well as positive changes in structural features over the semester. On the other hand, one might predict that students who have a more thorough understanding of their research subject are better able to simplify their concept maps and focus more precisely on the most important connections and concepts. This would result in a negative correlation between structural features and assessment of learning outcomes, as well as negative changes in structural features over the semester. If both explanations are valid for different students to varying extents, then we may find no (or weak) correlations between structural features and assessment of learning outcomes, as well as no (or weak) changes in structural features over the semester.

To answer this question, we directly assessed concept maps, looked for changes in concept maps over time, and looked for correlations between those changes and an external measure of scientific reasoning and writing skills.

Concept Maps for Thesis Writing

The concept map activity is one of the primary activities in Writing in Biology, a writing-intensive course designed for advanced undergraduates (typically seniors) who are working on an honors thesis or major research paper in the Department of Biology at Duke University (Reynolds *et al.*, 2009; Reynolds and Thompson, 2011).[1] Prior studies show that students enrolled in this course demonstrated significantly higher scientific reasoning skills than a statistically indistinguishable comparison group (i.e., students at the same level who wrote theses without taking the course), which suggests that the course directly targets and improves expertise in scientific reasoning.

In this course, instead of constructing traditional outlines, students construct concept maps to describe and contextualize their research. We reason that the complex structure of scientific writing makes conventional outlines a less useful learning tool, since outlines require students to communicate in a linear manner. Concept maps, on the other hand, are less constrained; students can literally construct their understanding of the interrelatedness of concepts and see how the relationships are interconnected.

At the beginning of the semester, students construct concept maps based on their own research in such a way that 1) individual concepts (not multiple or nested concepts) are represented in each box, 2) linking phrases between concepts do not include concepts, and 3) another student can look at the finished map and explain the primary research question in context. The instructor models this with a sample concept map before students work individually. In a subsequent class, some students' maps are selected for review in a whole-class workshop followed by peer review in small groups. Most feedback is generated as students attempt to explain one another's concept maps and ask follow-up questions. Reviewers tend to emphasize the use of jargon and the need for more or less elaboration in parts of the concept map. With this additional feedback, students generate revised concept maps. Instructors then review these revised maps, addressing the boundaries of the topic as delineated in the concept map (not so broad that that a reader might struggle to discern what information is most relevant but not so narrow that insufficient information is provided), the research question under consideration, and the general appropriateness of the propositions in the map. Instructors do not explicitly grade the concept maps in the course; instead, the maps are used as part of ongoing formative assessment. At the end of the semester, when students have completed their theses, they revisit their concept maps for final revision and reflection.

Our ultimate goal in this study is to expand on the relatively unexplored question of directly assessing concept maps as indicators of science reasoning in writing by leveraging our expertise in writing assessment. We would like to know whether the message from other studies—complexity increases with expertise—holds true in this context.

METHODS

Study Sample

Our sample consists of 49 undergraduates who were enrolled in the course (30 students in 2013, 19 students in 2014), all of whom were actively engaged in independent capstone research projects in biology or biology-related subjects. None of the analyses presented here were conducted until after

[1]One of the authors (J.A.R.) was a coinstructor for the course.

the course was complete, and students gave permission for us to use their work as part of this research using a consent form that was approved by the Duke University Institutional Review Board.

Assessment of Complexity of Concept Maps

Hierarchical concept maps—maps in which concepts tend to subsume or constitute related concepts—are often assessed based on the accuracy of the content, the structural characteristics (such as the number of concepts, linkages or propositions, and branches), and the quality of each specific connection (Novak *et al.*, 1983; Markham *et al.*, 1994; McClure *et al.*, 1999; Canas *et al.*, 2003). However, not all concept maps are hierarchical (Ruiz-Primo and Shavelson, 1996; Derbentseva *et al.*, 2007), particularly when generated for writing. Therefore, as we cannot identify hierarchies or crossovers (links between hierarchies) in students' maps, we focus our structural assessment on the number of concepts, the number of propositions (relationships involving two concepts and the link between them), and the number of branching points (concepts with at least three links connected to them) in each map. Each of the 49 students in our sample submitted three concept maps: the first draft version, the revised version, and the final version. We counted the number of concepts, propositions, and branching points in each of these versions for all students in the sample.

Assessment of Honors Theses

We assessed honors theses using the Biology Thesis Assessment Protocol (BioTAP; Reynolds *et al.*, 2009; Reynolds and Thompson, 2011), an assessment tool for valid and reliable assessment of student writing. BioTAP's rubric addresses 13 distinct dimensions of students' writing, including writing skills, scientific reasoning, and accuracy and appropriateness of research. Here, we focused exclusively on the nine dimensions related to writing and reasoning: appropriateness for target audience, argument for significance of research, articulation of goals, interpretation of results, implications of findings, organization, absence of writing errors, consistent and professional citations, and effective use of tables and figures. In focusing on these nine dimensions, we use BioTAP to directly measure primary learning outcomes of undergraduate research and thesis writing.

The procedure for rating students' theses using BioTAP was identical to that used and described in prior studies (Reynolds and Thompson, 2011; Dowd *et al.*, 2015a,b). Each dimension was rated on a scale of 1–5. A rating of 1 indicates the dimension under consideration is either missing, incomplete, or below the minimum acceptable standards. A rating of 3 indicates the dimension is adequate, but the work does not exhibit mastery. A rating of 5 indicates the dimension is excellent and the work exhibits mastery. As different parts of the thesis might fall into different categories, intermediate ratings of 2 and 4 may be appropriate.

Theses were assessed by a group of graduate student and postdoctoral associates, trained and supervised by J.A.R. For all assessments, each rater completed training in the use of the BioTAP rubric, which included examination of samples of students' writings that illustrated inadequate, adequate, and masterful levels of all nine dimensions being assessed. Raters then assessed sample theses that were not part of the

data set, discussed them, and established consensus scores as a means of calibration.

Each thesis in our sample was read by two raters who assessed the theses independently, subsequently discussed discrepancies in their ratings, and finally established a consensus score (Reynolds and Thompson, 2011; Dowd *et al.*, 2015a,b). The consensus score is not the simple average of the scores given by the two raters; rather, it is a discussion-based final score agreed upon by both raters. The Pearson correlation coefficient between raters' independent, prediscussion scores is 0.52 for total thesis scores; raters postdiscussion consensus scores are 100% in agreement. While the prediscussion value may seem low, we note that scores on each dimension are within one point of each other in ~80% of cases. In other words, the low prediscussion correlation seems to exacerbate relatively small differences between raters in this data set. Consensus scores were used in all analyses.

Specifically, we investigated each of the following aspects of BioTAP as possible variables of interest: each of the nine individual dimensions of BioTAP, the sum of scores on dimensions one through five (which are more related to reasoning about one's research), the sum of scores on dimensions six through nine (which are more related to organization and presentation), and the total sum of scores across all dimensions. Total BioTAP scores for students in our sample ranged from 32 to 44 (out a maximum possible score of 45).

Methods for Analysis

To assess changes in concept map features over time and to make comparisons between groups (e.g., students who simplify their maps at some point compared with students who do not), we used Student's *t* tests. To quantify relationships between structural features and the learning outcomes measured using BioTAP, we used Pearson's correlation values. As we considered a number of possible relationships among variables, we accounted for multiple comparisons using the Holm-Bonferroni method for controlling the family-wise error rate, which is both simple and more powerful than the Bonferroni method (Holm, 1979).

RESULTS

In our assessment of the complexity of concept maps, the number of concepts increases, on average, from the first draft to the revision and from the revision to the final map (Figure 1; $p = 0.003$ and 0.02, respectively). The number of propositions and the number of branches increase, on average, from the first draft to the revision and from the first draft to the final, but the change is not statistically significant from revision to final map ($p = 0.006$ and 0.07, respectively, for propositions; $p = 0.01$ and 0.26, respectively, for branches).

However, in spite of the average increases in these structural features, not all students make their maps more complex. In fact, a substantial number of students reduce the number of propositions in their maps at some point (Figure 2). Twenty-two percent of students simplify propositions in concept maps from the first draft to the revision, and 29% of students simplify from the revision to the final version. Interestingly, 49% of students simplify propositions in concept maps at one stage or another through the semester. We find very similar patterns in numbers of concepts and branching points.
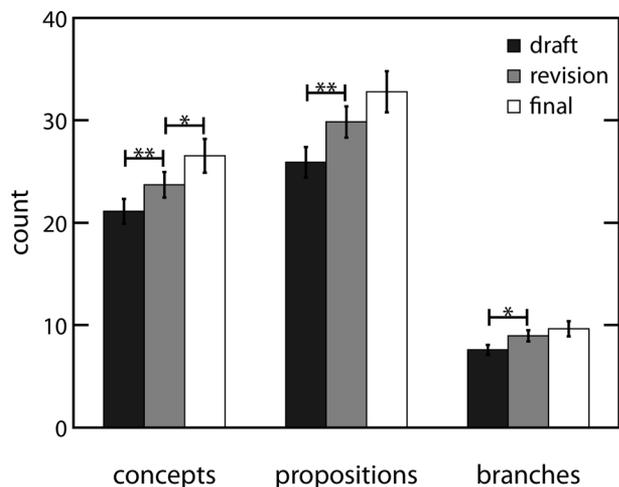
Figure 1. The average values of the number of concepts, the number of propositions, and the number of branches identified in students' initial drafts, revised drafts, and final drafts of their concept maps are shown. The number of concepts increases, on average, from the first draft to the revision and from the revision to the final map. The number of propositions and the number of branches increase, on average, from the first draft to the revision, but the change is not statistically significant from revision to final map. Statistically significant differences are represented by the horizontal bars and associated asterisks (*, $p < 0.05$; **, $p < 0.01$).

Most importantly, when we account for multiple comparisons, we do not find any statistically significant relationships between any of the structural features of concept maps assessed and the aspects of science reasoning in writing assessed using BioTAP. There are no statistically significant relationships between the number of concepts, propositions, or branches on the first drafts, revisions, or final versions and any of the BioTAP variables (individual dimension scores, partial sums, or total sum). Specifically, p values range from 0.03 to 0.99 for the 108 relationships tested; the one correlation for which $p < 0.05$ ($r = 0.31$, $p = 0.03$) is not significant using the Holm-Bonferroni method to correct for the multiple comparisons. Additionally, there are no statistically significant relationships between the *change* in number of concepts, propositions, or branches from first draft to revision, from revision to final version, or from first draft to final version and any of the BioTAP variables. Here, p values range from 0.02 to 1.00 for the 108 relationships tested; there are two correlations for which $p < 0.05$ ($r = 0.33$, $p = 0.02$; and $r = -0.32$, $p = 0.02$), but neither is statistically significant using the Holm-Bonferroni method. In Table 1, we highlight the nonsignificant correlations between the total sum of BioTAP and each of the variables related to concept maps discussed here. A complete table including other BioTAP variables is included in Supplemental Material, Table A. Even when we compare two groups (students who simplify at some point versus students who do not), differences in BioTAP variables are not significantly different; p values range from 0.10 to 0.95 for the 12 relationships tested (see Supplemental Material, Table B). In short, regardless of initial complexity, final complexity, or whether students make concept maps simpler or more complex throughout the semester, we find no relationships to science reasoning exhibited in thesis writing.

DISCUSSION

At first, it may seem that students' concept maps become more elaborate throughout the semester, in keeping with the prevailing idea that complexity is associated with expertise (Figure 1). However, although the *average* numbers of concepts, propositions, and branches increase, we find that a substantial percent of students actually simplify their concept maps throughout the semester (Figure 2). This suggests that students vary in how they engage with concept maps in the context of scientific writing and that increased expertise is not necessarily associated with more elaborate maps.

Previous research indicates that taking this course is associated with improved scientific reasoning skills (Reynolds and Thompson, 2011). Therefore, although we do not directly measure "change in science reasoning expertise" here, we can assume that such expertise (as measured using BioTAP) improves over the duration of this course. We see a range in BioTAP scores, of course, and some high-scoring students simplified their concept maps over the semester. Thus, the simplification of some students' maps, particularly at the end of the semester, should not be attributed to less expert-like understanding of the thesis topic. Moreover, as we found
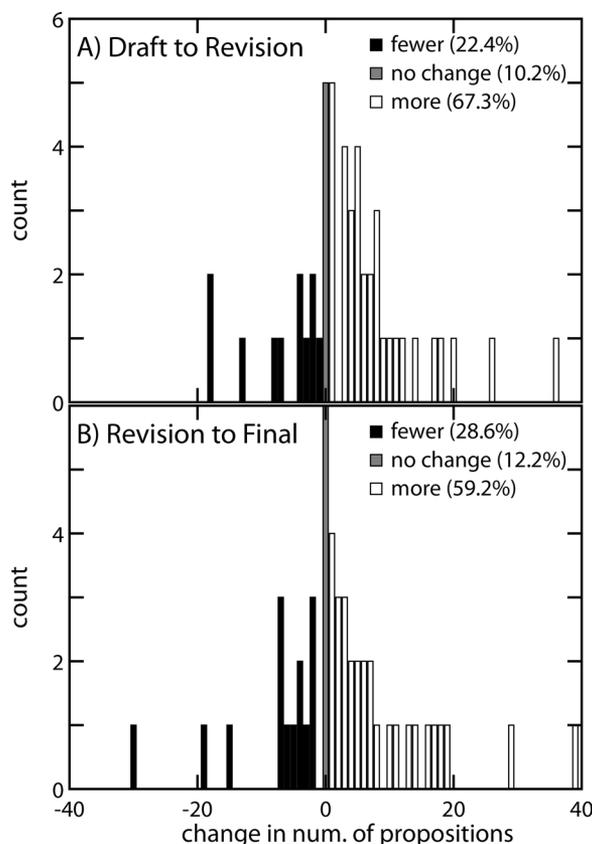
Figure 2. Histograms displaying the number of students who increase the number of propositions in their concept maps, decrease the number of propositions, and do not change the number of propositions are shown, both (A) from initial draft to revised draft and (B) from revised draft to final version. We find very similar patterns in numbers of concepts and branching points; a substantial number of students simplify their maps at one stage or another.

Table 1. Correlations between concept maps and BioTAP ($n = 49$)

		BioTAP total	p Value
Draft	Concepts	−0.18	0.215
	Propositions	−0.09	0.520
	Branches	−0.01	0.963
Revision	Concepts	−0.13	0.369
	Propositions	−0.10	0.489
	Branches	−0.11	0.440
Final	Concepts	−0.12	0.408
	Propositions	−0.09	0.543
	Branches	−0.05	0.728
Δ(Draft to revision)	Concepts	0.06	0.657
	Propositions	−0.01	0.951
	Branches	−0.11	0.460
Δ(Revision to final)	Concepts	−0.03	0.836
	Propositions	−0.02	0.908
	Branches	0.04	0.790
Δ(Draft to final)	Concepts	0.01	0.926
	Propositions	−0.02	0.876
	Branches	−0.05	0.726

no relationships between structural aspects of concept maps and thesis assessment, we suggest that such aspects do not inform instructors about students' science reasoning in their writing. Instead, our findings suggest that students use the concept maps to engage with their research in unique and possibly idiosyncratic ways. Structure-oriented assessments of concept maps simply do not apply when activities are oriented toward scientific writing.

Nonetheless, the changes to the structure of students' concept maps are interesting, because they indicate reorganization of students' ideas and presentation of those ideas. The fact that students' maps are changing suggests that students are engaging with the concept maps to articulate their messages in a dynamic, evolving way. Creating this space for discussion and engagement makes the use of concept maps a valuable pedagogical tool.

Not all students view the benefits of concept maps equally, however. In 2013, approximately two-thirds of the way through the course, instructors administered a survey to the class in which students anonymously indicated their perceived value of various activities. Instructors were interested in documenting students' perceptions of the various course activities. Compared with other key aspects of the course (e.g., individual meetings with instructors, in-class sample writing activities, and peer review), 38% of students ranked concept map activities as either the most or second-most valuable aspect of the course for learning to discuss one's research. It is not surprising that the concept map activity is not the most valuable to everyone, as the other components being compared were specifically designed to be engaging and valuable to students. Instead, we highlight the diversity of perspectives about what is most engaging and productive for students.

Importantly, we are not trying to assess the efficacy of concept mapping as a learning intervention in this study. Given that prior work has shown that this course, which includes concept maps among other elements that collectively function as an intervention, is associated with positive learning outcomes (Reynolds and Thompson, 2011), it would be

arbitrary (and possibly counterproductive) to test individual components of the course without a hypothesis or explicit reason for testing an alternative method. Thus, this study assumes that concept maps are valuable as an activity and is entirely oriented toward the question of whether the concept maps may directly inform instructors about learning outcomes in this context.

Of course, BioTAP is not the only means of assessing students' expertise in declarative knowledge about their research, scientific reasoning ability, and communication skills. It is possible that other measures of expertise (such as comprehensive exams or postgraduate success in writing) could reveal different relationships with structural features of concept maps. We focus on BioTAP because it is a unique and established tool for the assessment of science reasoning in writing at the capstone level.

Our findings suggest that instructors should not draw conclusions from assessment of structural features of concept maps in the context of scientific writing. However, this begs the question: From what, if anything, *should* instructors draw conclusions in students' concept maps? Instructors in this course provide students with discussion-oriented feedback on boundaries of the thesis topic, the research question, and the general appropriateness of the propositions in the concept map. It is possible that, as prewriting activities, such maps are mostly idiosyncratic; direct assessment may inform discussions but does not ultimately relate to learning outcomes. Alternatively, perhaps learning outcome–related direct assessment *is* possible, but it must relate better to the argument and context that students develop in their concept maps. Instructors' feedback in this course is not presently based on a rubric, though holistic rubric-based assessment has been used for concept maps in other contexts, both related to scientific writing and content knowledge (McClure *et al.*, 1999; Conklin, 2007). Thus, the development of a rubric analogous to BioTAP for concept map assessment could be beneficial for instructors, either to facilitate discussion or to connect to learning outcomes.

CONCLUSION

Our primary goal is for students to succeed in their research endeavors and undergraduate thesis-writing experience. If we can identify attributes of concept maps that correlate strongly with these outcomes, we may be able to identify and help those who are struggling much earlier in the course. Here, we asked whether structural features of students' concept maps could play such a helpful role, and we found no evidence to support this. Our findings and our experience with the course lead us to believe that holistic assessment may better inform the development of a rubric for content map assessment than structural attributes. BioTAP will be an invaluable tool for evaluating such a rubric, as it is the primary means of systematically relating scientific reasoning in writing to any measures that might be developed.

ACKNOWLEDGMENTS

REFERENCES

Anderson-Inman L, Horney M (1996). Computer-based concept mapping: enhancing literacy with tools for visual thinking. J Adolesc Adult Literacy *40*, 302–306.

Bramwell-Lalor S, Rainford M (2014). The effects of using concept mapping for improving advanced level biology students' lower- and higher-order cognitive skills. Int J Sci Educ *36*, 839–864.

Brodney B, Reeves C, Kazelskis R (1999). Selected prewriting treatments: effects on expository compositions written by fifth-grade students. J Exp Educ *68*, 5–20.

Canas AJ, Coffey JW, Carnot MJ, Feltovich P, Hoffman RR, Feltovich J, Novak JD (2003). A Summary of Literature Pertaining to the Use of Concept Mapping Techniques and Technologies for Education and Performance Support.

Cliburn JW (1990). Concept maps to promote meaningful learning. J Coll Sci Teach *19*, 212–217.

Conklin E (2007). Concept mapping: impact on content and organization of technical writing in science. Diss Abstr Int Sect A Humanities Soc Sci *68*, 516–516.

Crane M (1998). Write away! Order out of chaos: using concept maps to re-envision student writing. The University of Louisville's Writing Across the Curriculum Newsl *3*(4).

Derbentseva N, Safayeni F, Cañas AJ (2007). Concept maps: experiments on dynamic thinking. J Res Sci Teach *44*, 448–465.

Dowd JE, Connolly MP, Thompson RJ Jr., Reynolds JA (2015a). Improved reasoning in undergraduate writing through structured workshops. J Eco Educ *46*, 14–27.

Dowd JE, Roy CP, Thompson RJ Jr, Reynolds JA (2015b). "On course" for supporting expanded participation and improving scientific reasoning in undergraduate thesis writing. J Chem Educ *92*, 39–45.

Floden R (1991). What teachers need to know about learning. In: Teaching Academic Subjects to Diverse Learners, ed. M Kennedy, New York: Teachers College, 181–202.

Giger T (1995). Zen and the art of prewriting. Writing Lab Newsl *20*(2), 9.

Gouli E, Gogoulou A, Grigoriadou M, Gouli E, Gogoulou A, Grigoriadou M (2003). A coherent and integrated framework using concept maps for various educational assessment functions. J Inf Technol Educ Res *2*, 215–240.

Holm S (1979). A simple sequentially rejective multiple test procedure. Scand J Stat *6*, 265–70.

Hyerle D (1995). Thinking maps: seeing is understanding. Educ Leadership *53*, 85.

Khodaday E, Ghanizadeh A (2011). The impact of concept mapping on EFL learners' critical thinking ability. Eng Lang Teach *4*, 449–460.

Lee W, Chiang C-H, Liao I-C, Lee M-L, Chen S-L, Liang T (2013). The longitudinal effect of concept map teaching on critical thinking of nursing students. Nurse Educ Today *33*, 1219–1223.

Markham KM, Mintzes JJ, Jones MG (1994). The concept map as a research and evaluation tool: further evidence of validity. J Res Sci Teach *31*, 91–101.

McClure JR, Sonak B, Suen HK (1999). Concept map assessment of classroom learning: reliability, validity, and logistical practicality. J Res Sci Teach *36*, 475–492.

Mintzes J, Quinn HJ (2006). Knowledge restructuring in biology: testing a punctuated model of conceptual change. Int J Sci Math Educ *5*, 281–306.

Murray DM (1978). Write before writing. College Compos Commun *29*, 375–381.

Nesbit JC, Adesope OO (2006). Learning with concept and knowledge maps: a meta-analysis. Rev Educ Res *76*, 413–448.

Nesbit JC, Adesope OO (2013). Concept maps for learning. In: Learning Through Visual Displays, ed. G Schraw, MT McCrudden, and D Robinson, Charlotte, NC: Information Age, 303–328.

Novak JD (1990). Concept mapping: a useful tool for science education. J Res Sci Teach *27*, 937–949.

Novak JD (2005). Results and implications of a 12-year longitudinal study of science concept learning. Res Sci Educ *35*, 23–40.

Novak JD, Bob Gowin D, Johansen GT (1983). The use of concept mapping and knowledge vee mapping with junior high school science students. Sci Educ *67*, 625–645.

Osman-Jouchoux R (1997). Linking reading and writing: concept mapping as an organizing tactic. In: VisionQuest: Journeys toward Visual Literacy. Selected Readings from the 28th Annual Conference of the International Visual Literacy Association, held October 1996, in Cheyenne, WY. http://eric.ed.gov/?id=ED408955.

Osmundson E, Chung G, Herl HE, Klein DC (1999). Knowledge Mapping in the Classroom: A Tool for Examining the Development of Students' Conceptual Understandings, Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing.

Reynolds JA, Smith R, Moskovitz C, Sayle A (2009). BioTAP: a systematic approach to teaching scientific writing and evaluating undergraduate theses. BioScience *59*, 896–903.

Reynolds JA, Thompson RJ (2011). Want to improve undergraduate thesis writing? Engage students and their faculty readers in scientific peer review. CBE Life Sci Educ *10*, 209–215.

Reynolds SB, Hart J (1990). Cognitive mapping and word processing. J Exp Educ *58*, 273–279.

Ruiz-Primo MA, Shavelson RJ (1996). Problems and issues in the use of concept maps in science assessment. J Res Sci Teach *33*, 569–600.

Srinivasan M, McElvany M, Shay JM, Shavelson RJ, West DC (2008). Measuring knowledge structure: reliability of concept mapping assessment in medical education. Acad Med *83*, 1196–1203.

Vanides J, Yin Y, Tomita M, Ruiz-Primo MA (2005). Using concept maps in the science classroom. Sci Scope *28*, 27–31.

Wallace JD, Mintzes JJ (1990). The concept map as a research tool: exploring conceptual change in biology. J Res Sci Teach *27*, 1033–1052.

Wu P-H, Hwang G-J, Milrad M, Ke H-R, Huang Y-M (2012). An innovative concept map approach for improving students' learning performance with an instant feedback mechanism. Br J Educ Technol *43*, 217–232.

Yin Y, Shavelson RJ (2008). Application of generalizability theory to concept map assessment research. Appl Meas Educ *21*, 273–291.