Reliable Measures of Concept Map Examinations

Darrell L. Butler

Professor Dr.

Dept of Psychological Science, Ball State University, Muncie, IN 47306, USA
dlbutler@bsu.edu

ABSTRACT

The emerging capabilities of online educational materials are creating new challenges for faculty who use examinations to evaluate conceptual knowledge. This report provides an exploration of whether non-traditional concept map exams can not only mitigate some of the challenges, but can encourage students to organize conceptual knowledge, can be created quickly, scored reliably, and have many other desirable characteristics. An analysis of existing concept map scoring approaches reveals some useful variables and some variables with poor reliability and validity. Some new variables with excellent reliability and face validity are suggested. A software testing system called EasyMap was created for this exploration and is available for free to others interested in researching concept map exams.

Keywords:

INTRODUCTION

The emerging capabilities of online educational materials are providing new challenges to pedagogies that use online exams to evaluate conceptual knowledge. For example, some of the new search tools allow students to quickly search multiple educational materials for concepts, names, and phrases. A student taking a multiple choice exam could open an exam in one browser and the search tools in another, allowing them to very quickly look up many answers to typical multiple choice questions. This is not meant as a criticism of the tools that are being added to educational materials or that students should not be encouraged to use them, rather it creates concerns that multiple choice exams may not be ideal measures of conceptual knowledge when such tools exist. Of course, there are various ways to proctor students, but the cost is relatively high and alternatives are desirable.

There are many other attributes of exams that are important besides the concern described above. As a starting point, below is a list of the characteristics of an exam for conceptual information that may be important in pedagogy of online classes in higher education:

1. Enables instructors to efficiently create questions that are appropriate for higher education courses.
2. Requires students to have developed and to use well developed, well organized conceptual systems rather than simply report individual definitions, names, and facts.
3. Emphasizes core material but is sufficiently open-ended that students have the freedom to reveal the ways in which they have gone beyond the core or basic knowledge.
4. Rewards students who learn the technical vocabulary that educated individuals use.
5. Enables instructors to efficiently score exams.
6. Utilizes scoring procedures that students perceive as objective or at least very fair.
7. Produces scores that are reliable and valid measures of conceptual learning.

The literature on exams was evaluated with these criteria in mind. A number of researchers (e.g., Davis, 1993; McKeachie and Svinicki, 2010; Nilson, 1998) have described the strengths and weaknesses of various ways to evaluate conceptual leaning. For example, multiple choice exams have many problems: They are not open ended, good ones are not easy to create, in practice most emphasize basic definitions rather than organized conceptual systems, and many have questionable validity. In contrast, essay exams are very time consuming to grade, students see scores as very subjective, in practice they often do not reveal well developed conceptual systems, and they have questionable reliability. Similar analysis of other traditional testing approaches (e.g., fill-in the blank, matching, and short answer) indicates that none satisfy very many of the criteria listed above. However, a possible non-traditional approach that

may meet these criteria is to have students create concept maps to demonstrate their knowledge and understanding (Jonassen et al., 1993).

CONCEPT MAPPING

Generally speaking, a concept map is a 2-dimensional visual representation of a group of propositions or ideas displayed using simple polygons connected by lines. A proposition consists of two concepts (each displayed in a polygon) and a linking phrase (labeling the line connecting the two polygons). Concept maps contain many nodes and many connections (see Figure 1). Novak and Gowin (1984) proposed concept maps as a way to enable learners to represent concepts and their interrelationships. During the last 30 years, a number of researchers have studied the use of concept maps as a way to promote conceptual learning and have reported positive results. Although not a complete list, the disciplines in which the use of concept maps has been studied include Biology (Heinze-Fry & Novak, 1990; Schmid & Telaro, 1990), Chemistry (Markow & Lonning, 1998; Stensvold & Wilson, 1990), Geosciences (Rye & Rubba, 1998), Mathematics (Williams, 1998; Roberts, 1999), Medicine (Torre et al., 2007; West et al. 2000), Physics (Pankratius, 1987; Shymansky, 1997), and Psychology (Berry & Chew, 2008; Jacobs-Lawon & Hershey, 2002).

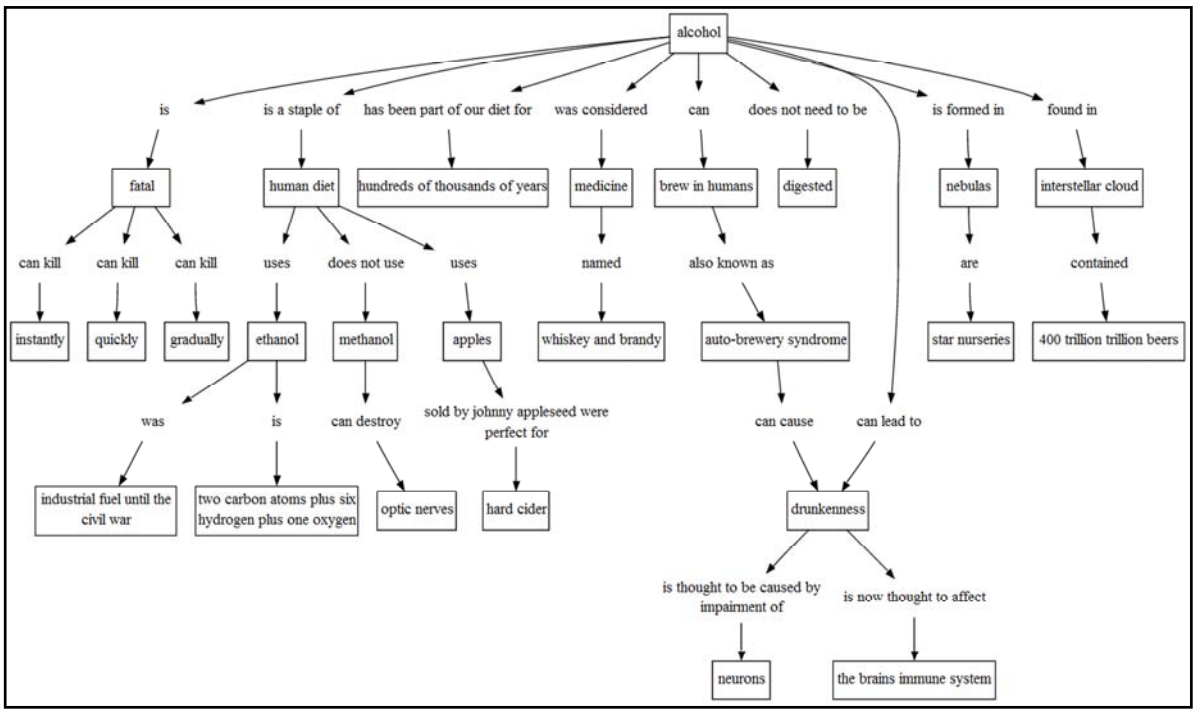

Figure 1: A Simple Concept Map Concerning Alcohol

The prompt for a concept map should be thought of as the topic of the map. It can be written such that a concept map exam would satisfy several of the goals (numbers below refer to the list of goals above):

- Prompts can be relatively simple, like those for an essay exam, and thus can be created relatively quickly. (#1)
- Prompts can explicitly encourage students to demonstrate their organization of ideas and concept maps appear to encourage at least some kind of organization. (#2)
- Prompts can encourage students to emphasize core ideas and still be open-ended. (#3)

A review of literature and some preliminary studies suggest that concept map exams may be able to satisfy the other goals listed above . Below, is a brief summary of the published research on concept map exams, along with a critique of some of the scoring systems used, and some suggestions for more reliable, valid scoring approaches.

SCORING CONCEPT MAPS

Surprisingly, preliminary research indicates that students should type ideas, not draw a map! Most of my colleagues who have had students draw maps have given up trying to grade them because they could not decipher the students' maps. In the literature, I did not find any studies with inter-scorer reliability coefficients above .60 if the

students created maps by drawing them (see review by Ruiz-Primo & Shavelson, 1996).

Concept mapping software presents one solution to the scoring problem. Some concept mapping software permits students to type a sentence in the form of concept | relation | concept. The software then adds the sentence to the map. Scoring maps created this way is much more reliable. In one preliminary study involving 152 students and 6 scorers, the reliability (measured as proportion of sentences scored the same) was .94 and in another study involving 24 students and two judges the reliability was perfect. Butler and Ring (2011) reported a reliability of .91 for an exam involving 35 students and three judges.

Researchers have reported that one important and highly reliable indicator of map quality is the number of non-redundant, correct propositions or ideas in the map (e.g., McClure & Bell, 1990; McClure, Sonak, & Suen, 1999, Novak & Gowin, 1984). These researchers report that the number of correct propositions is much more reliable than scores based on comparing a student map to an expert map or holistic ratings of maps. Some researchers have counted the number of concepts instead of propositions (e.g., Markham et al., 1994), but from a cognitive perspective, this level of analysis is too narrow as a measure of overall map quality. Kintsch (e.g., Kintsch, 1988, 1998; Kintsch & VanDijk, 1978; Van Dijk & Kintsch, 1983), whose work is the foundation of most comprehension models of reading comprehension, argues that the meaning of text is based on propositions—that is, simple sentences—not individual words.

Counting some words in addition to the number of correct propositions can provide a useful, reliable measure. The fourth criterion for exams listed above was to reward students who use the technical vocabulary that educated individuals use. Counting the number of important vocabulary words in a concept map not only encourages students to learn that vocabulary but it also encourages them to emphasize core ideas in their concept maps.

In their foundational work on concept maps in education, Novak and Gowin (1984) argued that in addition to the number of correct propositions, scores should include a measure of the depth of the hierarchical structure and the number of crosslinks in the map. However, there are problems with both of these measures. The problem with depth measures is that while hierarchies are common in many disciplines, some conceptual systems may be represented by students to include cycles (e.g., Krebs cycle) and or strong linearity (e.g., causal events) and such maps may not be very deep. Not surprisingly, research has revealed that scoring systems using depth of hierarchy are often associated with low validity (see Ruiz-Primo & Shavelson, 1996).

Cross-links are connections between branches of the hierarchy and are measured by counting the number of lines in the graph that cross one another. However, graph theory in mathematics shows that cross-links are not a valid scoring method. For example, Figure 2 shows three graphs, each of which has 6 nodes with each node having three links to other nodes. In the mathematics of graph theory, these maps are equivalent; that is, each could represent a concept map of 9 propositions. However, note that the number of cross-links differs greatly among graphs. Cross-links are a very arbitrary characteristic of a graph and when included in scores of concept maps, the result is low validity.

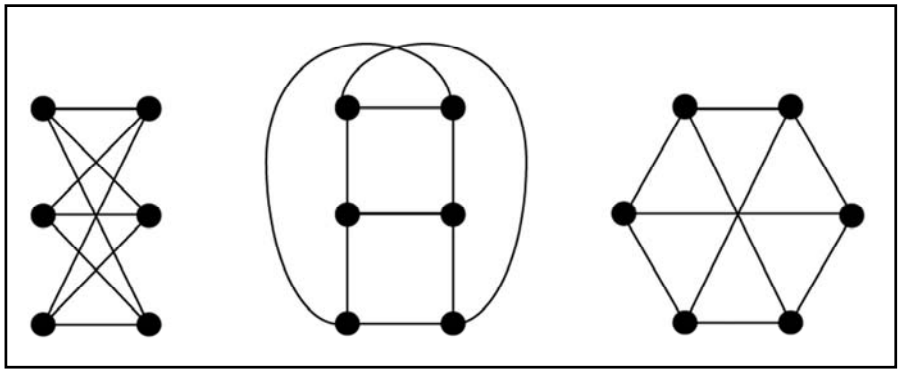

Figure 2: Diagram Illustrating Why Cross-links are not Valid Measures of Concept Maps

So what other characteristics of concept maps should scoring take into account? Given the list of exam criteria presented above, some measure of organization is needed. Furthermore, researchers have reported that under some circumstances, measures of concept maps that involve structure or organization can be more sensitive to learning than counts of propositions (e.g., Srinvasan et. al. 2008; West et al. 2002).

Graph theory in mathematics suggests a useful approach. If a student is asked to produce a well-developed, well organized concept map, there should only be one connected graph or tree in their concept map; in contrast, students who have trouble connecting ideas will have more than one graph or tree in their map. Thus, a useful measure of organization could be the number of disconnected trees in the concept map. However, it is important to take into account incorrect and inappropriate links because students can force ideas to link in illegitimate ways in order to appear organized. Therefore, a potentially more useful measure of organization is the number of disconnected graphs or trees in a map when incorrect and irrelevant links are removed.

There have been suggestions in the literature for other measures. For example, Kinchin (2013) has argued that different kinds of structures may be included in concept maps, such as chains and cycles and these may be important to kinds of knowledge being represented. However, at the current time, no specific measures have been suggested nor any procedure for reliably scoring these kinds of structures.

CONCEPT MAPPING SOFTWARE

There are many concept mapping software programs available. However, each of them has major drawbacks and none had the ability to score all of the variables described above. Eric Sanner, a programmer at nHarmony, Inc. cooperated with me to create a new Internet-based concept mapping program that would have the characteristics of the kind of testing system needed to conduct research. The software is called EasyMap. Over the past three years, usability testing, laboratory experiments, and field studies have provided some evidence that the system may be very useful in online courses.

CREATING EXAMS IN EASYMAP.

Prompts for concept map tests are quite similar to prompts or questions given to students for essay exams. For example, a simple prompt for a concept map exam could be: “Create a well-organized concept map about skin perception from a Psychological perspective. Your map should not contain more than 80 propositiions.” EasyMap software lets faculty re-use these prompts in subsequent exams.

Research suggests some considerations in writing prompts:

- Both lab and classroom research indicates that the best prompts refer to material covered in about one or two chapters in a typical text book or a couple of weeks of class. When the prompt is too general (e.g., an entire course or text), students report that they do not have enough time to demonstrate what they know and report being overwhelmed by the task. If the prompt is too narrow (e.g., one section in a chapter), students cannot build an elaborate map.
- It is helpful to give some sample prompts at the beginning of a learning module so that students have clear objectives to guide their reading, note taking, and studying. Giving the students clear objectives has been has been considered best practice since Bloom (1956) described the importance of educational objectives.
- It is useful to limit the total number of sentences that students enter to create a concept map and include this limit in the prompt. Several in class studies have found that students who have studied can enter a sentence or proposition about every 35 seconds and create a well-organized map. Thus for a 50 minute test, a limit of 80 sentences reduces the amount of irrelevant material students include in their maps and avoids problems associated with assuming that a larger map is always better (Kinchin, 2013).

Along with the prompt, key words or technical terms can be listed. For the skin perception prompt above, some of the keywords could be meissner, merkel, ruffini, pacinian, thermoreceptors, and nociceptors. The software can automatically count these in a student’s concept map using a fuzzy matching algorithm so students do not have to spell key concepts perfectly. However, because the fuzzy match will not catch all equivalent terms, the software allows a test creator to list synonyms with key terms. Thus the software would count either “TSD” or “Theory of Signal Detection” if they were indicated as synonyms.

In addition to the prompt and the key words, faculty indicate who else can score the concept maps (such as teaching assistants), the password students must use to access the exam, and start and end date/time.

STUDENT INTERFACE

Figure 3 shows the student interface once he or she has accessed a test. Note the prompt is at the top and right below the prompt are three boxes. Students enter propositions in the form of subject phrase | verb phrase | object phrase. Because some students have trouble stating ideas in this format, EasyMap does semantic processing to check what students have entered and the software provides feedback if the entry does not appear to be a sentence of this form. Figure 3 shows three possibilities for the sentence typed in the three boxes. A student can ignore the suggestions of the semantic processor or simply select one of the possibilities listed by the software. The software

adds the sentence to the map, connecting it to the existing map if the sentence subject or object matches an existing concept in the map.

All the sentences entered are added to the concept map (lower right) and to the list of entered sentences (lower left). By selecting icons next to a sentence in the list, students can adjust the relative position of the sentence object with sentence subject (above, below, or to the right) or they can delete sentences. They can search for words in their sentences and the software will display sentences containing those words. By selecting one of the sentences, the map will center on the searched for word. Students report this is useful when working on a relatively small screen. Also, students can control the amount of view space used for map and list and they can set the scale of the map.

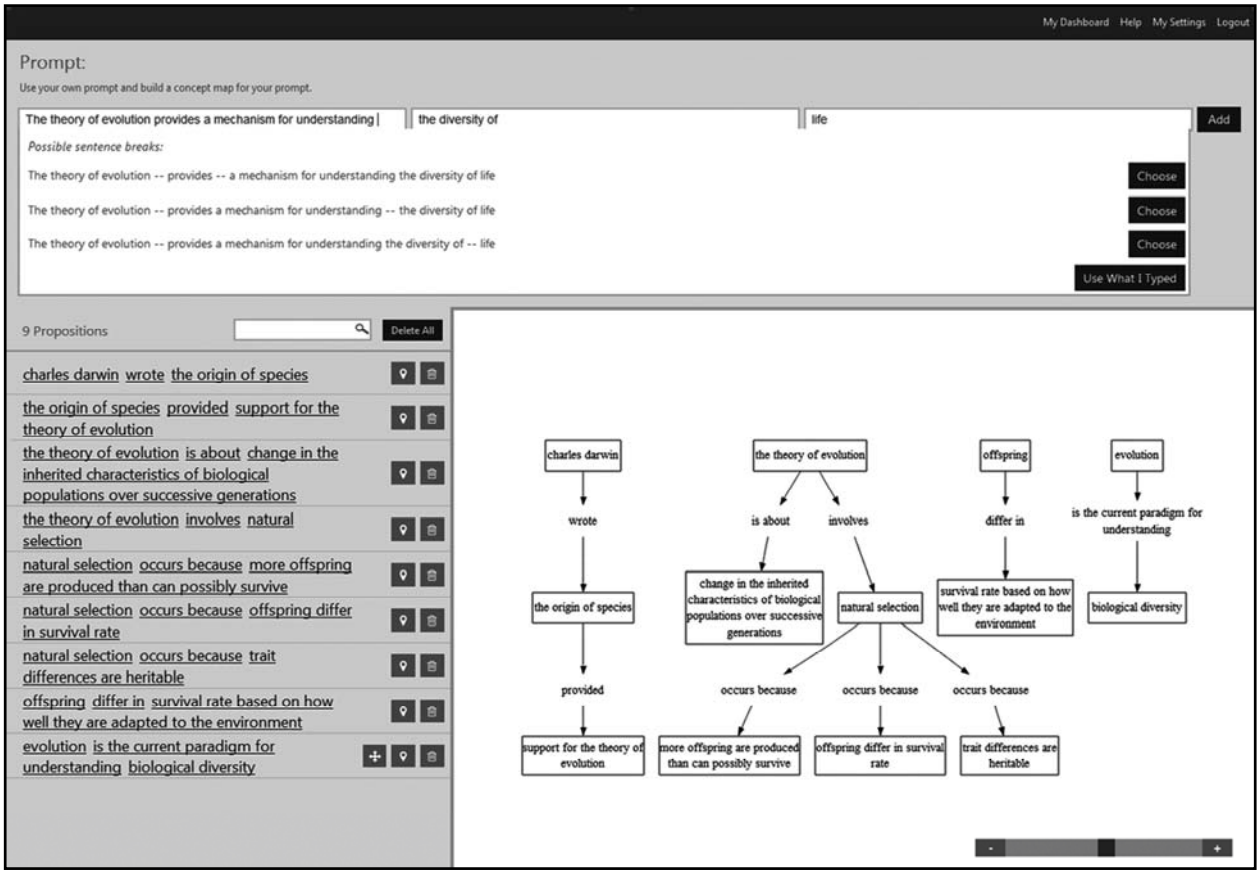

Figure 3: Student Concept Map Interface in EasyMap

When EasyMap was first created, students needed about a 45 minute introduction to use the software. Even with that amount of training, some students really struggled with the interface and many were still a little frustrated after several exams. Usability studies on the interfaces have led to improvements that have dramatically reduced the training time and the problems students reported. In these studies, students are given instructions and a short article from a magazine such as Discover. They are asked to build a map from the information in the article. Students are watched carefully and interviewed about pauses or issues with the software. In the most recent usability testing, students could create an account, go through the online instructions and create a basic map on the article in about 20 minutes. The nine students in the study were given the statement “I feel confident that I can use the software” and asked to indicate their agreement on a 5-point Likert scale ranging from Strongly Disagree (1) to Strongly Agree (5). The mean rating was 4.0 and no students disagreed with the statement. Some students in the usability study did have some difficulty organizing their ideas, but this seems to be more the result of the novelty of the task than an interface issue. When used for course exams, because the software is available on the Internet, students can practice using it outside of class time and quickly overcome its novelty.

SCORING EASYMAP EXAMS

Not surprisingly, the fewer options in a scale, the greater its reliability. A scale that has very high reliability (as measured by percent agreement among judges) is:

- Correct
- Correct, but redundant
- Incorrect
- Irrelevant

Adding two additional options to separate quality of correctness (e.g., correct but small problem or vague vs correct and incorrect vs incorrect, but not seriously so) lowered reliability from .94 to .82.

The EasyMap scoring interface is easy to use, facilitates very efficient scoring, and produces scores with very high reliability. Once propositions are scored, the software provides a screen view and exportable spreadsheet of the following variables for each student. All of these variables are machine scored except the correctness of propositions and all are highly reliable measures:

- Quantity
    - Number of Propositions (machine scored; reliability = 1.00)
    - Number of Technical Vocabulary Items Included (machine scored; reliability = 1.00)
- Quality (human score correctness; reliability > .90)
    - Number of Correct Propositions
    - Number of Technical Vocabulary Items Used in Incorrect Propositions
- Organization
    - Number of Disconnected Groups of Propositions Number of Graphs in map) (machine scored; reliability = 1.00)
    - Number of Graphs in map after removing incorrect and irrelevant propositions (machine scored; reliability = 1.00)

The variables can all be downloaded and weighted in a spreadsheet to create total scores for students. For example, it may be appropriate to use a weighted sum such as

+ (.9) * (number of correct proposition/maximum number of propositions allowed in prompt)

+ (.1) * (number of technical vocabulary items/max number of technical vocabulary items of any student in the class) –

- (.01) * (1 - number of graphs in the map after removing incorrect and irrelevant propositions).

Recently, I used EasyMap and this scoring system in a senior-level undergraduate class. Here are some of the comments received from students at the end of the course:

- "This course has been what I always thought college would be. But this is the first time I've experienced it and it is my last semester!"
- "I really prefer the kinds of tests you gave us. They make me study the way I should!"
- "I prefer essay tests over these new tests because I don't have to work as hard."

Interviews with the top students in the class revealed that they either made practice concept maps or elaborate outlines and made sure they could connect their ideas to prepare for all of the exams.

Although all of these variables have high content validity, detailed criterion-related validity studies on all of them have not been completed. Ley at al (2012) reported positive correlations between score on a multiple choice test and the correct ideas in a concept map test. Preliminary data is available for students in an upper division course who were asked to create three maps for each learning module in a class:

1st: a map before reading or attending lectures on a new module (max time 90 min.);

2nd: a map in-class after reading, attending lectures, and studying (max time 90 min.); and

3rd: a map at home with open book and open notes (max time 48 hrs).

A straightforward prediction is that students should be able to create more elaborate concept maps at each step in this process. Students would know the least when creating the first map making it the weakest. Once students

have studied, they should perform better. However, the best performance should occur on the third map because students can check other sources, look up information they are unsure of, and take advantage of a longer performance time. The correlation between the students' number of correct propositions and when they created the map was .79.

CAN CONCEPT MAPS BE SCORED QUICKLY

In a course during the summer of 2013, students completed three essay tests and three concept map tests. Students were given the same amount of time for each. All 6 exams were graded using the same rubric and a timer was used to measure how long each scoring session took. The grading approach involved counting the number of correct, non-redundant ideas (not including examples, trivial details, irrelevant ideas, etc.) and the number of technical vocabulary items used in correct propositions. Although the organizational variables in easymap are machine scored so they do not require a grader, they were not included because it was not clear how to do this for an essay exam. On average, it took five times as long to grade an essay exam as it did to grade a concept map exam. Furthermore, students stated that they thought the concept map grades were more objective than the essay grades (interesting given the same rubric was used to grade both).

While this finding indicates that the software is providing better speed of grading than essays, even faster scoring is possible in the future. Machine assisted scoring is feasible because the variability in simple propositions with the same meaning is not great, especially when compared to the variability of sentences in essays. The software can identify sentences that have been graded previously and can handle a certain amount of misspelling. The software could be adapted to handle synonyms and minor grammar variations of simple propositions to increase the number of propositions that could be computer scored. Scoring may be speeded up while raising confidence in the results by using a social network of scorers who teach classes that allow them to use the same prompts. However, it may be feasible to use artificial intelligence (AI). The general idea is to let the AI system build a knowledge base of simple propositions from relevant textbooks or other materials. The AI might also need some training on map scoring. Following training, an AI may be able to accurately score nearly all the propositions.

CONCLUSION

Researchers have reported that concept mapping is generally more effective for attaining knowledge retention and transfer than reading, attending lectures, and participating in class discussions (see meta analysis by Nesbit & Adesope, 2006). However, this research has emphasized using concept maps as a learning strategy rather than as a testing method. The research on using concept maps for exams suggests that concept map exams can be used to fulfill the list of desirable exam characteristics given provided at the beginning of this paper. There is still much to determine. However, more research on the validity of map scoring is needed. Can concept map exams be valid for graduate school students, upper division students, lower division students, and perhaps students at the secondary or younger levels?

The preliminary research is encouraging, but there is still much research to do. If you are interested in doing research with EasyMap, I look forward to communicating with you. The Easymap software is free for faculty.

ACKNOWLEDGEMENTS

The research reported in this paper was approved by the Institutional Review Board. There are no conflicts of interest. The software described in the research is available for free to other faculty researchers. This research was not supported by grants but it would not have been possible without the software provided by nHarmony, Inc as well as a number of students who worked with the author over the last five years. The author thanks the following students: Natalie Strzok, Nathanial Ring, and Matt Ladwig for their assistance with the laboratory studies. The author also thanks Dr. Paul Biner and Dr. Bernard Whitley for comments on drafts of this article.

REFERENCES

Berry, J. and Chew, S. 2008. "Improving Learning Through Interventions of Student-Generated Questions and Concept Maps." *Teaching of Psychology,* 354, 305-312.

Bloom, B. 1956. *Taxonomy of Educational Objectives, The Classification of Educational Goals – Handbook I: Cognitive Domain*. NY: McKay.

Bower, G.H., Clark, M.C., Lesgold, A.M., and Winzenz, D. 1969. "Hierarchical retrieval schemes in recall of categorized word lists." *Journal of Verbal Learning and Verbal Behavior,* 83, 323-343.

Butler, D.L. and Ring, N.S. Nov, 2011. "A New, Non-Traditional, Web-Based Approach to Measuring Student Knowledge in Stem Courses." *Proceedings of the 17th Annual Sloan Consortium International Conference on Online Learning*, Lake Buena Vista, Florida.

Davis, B.G. 1993. *Tools for Teaching*. San Francisco, CA: Jossey-Bass.

Heinze-Fry, A. and Novak, J.D. 1990. "Concept mapping brings long term movement towards meaningful leaning." *Science Education,* 744, 461-472.

Jacobs-Lawson, J.M. and Hershey, D.A. 2002. "Concept maps as an assessment tool in Psychology courses." *Teaching of Psychology*, 291, 25-29.

Jonassen, D.H., Beissner, K., and Yacci, M.A. 1993. *Structural knowledge: Techniques for representing, assessing, and acquiring structural knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kinchin, I. (2013). Concept mapping and the fundamental problem of moving between knowledge structures. *Journal for Educators, Teachers and Trainers*, *4(1),* 96–106.

.Kintsch, W. and Van Dijk, T.A. 1978. "Toward a model of text comprehension and production." *Psychological Review, 85 5,* 363-394.

Kintsch, W. 1988. "The use of knowledge in discourse processing: A construction-integration model." *Psychological Review, 95*, 163-182.

Kintsch, W. 1998. *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.

Ley, S.L., Krabbe, H., & Fisher, H.E. (2012). Convergent validity: Concept maps and competence test for students diagnosis in physics. Proc. of the 5th International conference on Concept Mapping. (http://cmc.ihmc.us/cmc2012papers/cmc2012-p30.pdf, last accessed on 1/13/2014).

Markham, K.M., Mintzes, J.J., and Jones, M.G. 1994. "The concept map as a research and evaluation tool: Further evidence of validity." *Journal of Research in Science Teaching*, 31, 91-101.

Markow, P. G. and Lonning, R. A. 1998. "Usefulness of concept maps in college chemistry laboratories: Students' perceptions and effects on achievement." *Journal of Research in Science Teaching, 359*, 1015-1029.

McKeachie, W. J., and Svinicki, M. 2010. *McKeachie's Teaching Tips: Strategies, Research, and Theory for College and University Teachers, 13th Ed*. Florence, KY: Wadsworth.

Nesbit, J.C. & Adesope, O.O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research, 76 (3),* 413-448.

Nilson, L.B. 1998. *Teaching at its Best*. Bolton, MA: Anker Publishing.

Novak, J.D., & Gowin, D.R. (1984). *Learning how to learn*. New York: Cambridge Press.

Paivio, A. 1969. "Mental imagery in associative learning and memory." *Psychological Review, 763*, 241-263.

Pankratius, W.J. 1990. "Building an organized knowledge base: Concept mapping and achievement in secondary school physics." *Journal of Research in Science Teaching*, 274, 315-333.

Postman, L. 1972. "A pragmatic view of organization theory." In E. Tulving and W. Donaldson Eds., *Organization of Memory*. NY: Academic Press.

Roberts, L. 1999. "Using concept maps to measure statistical understanding." *International Journal of Mathematical Education in Science and Technology, 30,* 707–717.

Ruiz-Primo, M.A., and Shavelson, R.J. 1996. "Problem and issues in the use of concept maps in science assessment." *Journal of Research in Science Teaching*, 336, 569–600.

Rye, J.A. and Ruba, P.A. 1998. "An exploration of the concept map as an interview tool to facilitate the externalization of students' understandings about global atmospheric change." *Journal of Research in Science Teaching*, 355, 521–546.

Schmid, R.F. and Telaro, G. 1990. "Concept mapping as an instructional strategy for high school biology." *Journal of Educational Research,* 84, 78-85.

Shymansky, J.A., Yore, L.D., Treagust, D.F., Thiele, R.B., Harrison, A, Waldrip, B.G., Stocklmayer, S.M., and Venville, G. 1997. "Examining the construction process: A study of changes in level 10 students' understanding of classical mechanics." *Journal of Research in Science Education,* 346, 571-593.

Srinivasan, M., McElvany, M., Shay, J.M., Shavelson, R.J., and West, D.C. 2008. 'Measuring knowledge structure: Reliability of concept mapping assessment in medical education." *Journal of Medical Education*, 83(12), 1196-1203.

Stensvold, M.S., and Wilson, J.T. 1990. "The interaction of verbal ability with concept mapping in learning from

a chemistry laboratory activity." *Science Education,* 744, 473-480.

Torre, D.M., Dalye, B., Stark-Skchweitzer, T., Siddartha, S., Petkova, J. and Ziebart, M. 2007. "A qualitative evaluation of medical student learning with concept maps." *Medical Teacher,* 29, 949-955.

Tulving, E. 1962. "The effect of alphabetic subjective organization on memorizing unrelated words." *Canadian Journal of Psychology,* 16, 185-191.

West, D.C., Pomeroy, J.R., Park, J.K., Gerstenberger, E.A., and Sandoval, J. 2000. "Critical thinking in graduate medical education: A role for concept mapping assessment?" *JAMA*, 284, 1105–1110.

West, D.C., Park, J.K., Pomeroy, J.R., and Sandoval, J. 2002. "Concept mapping assessment in medical education: A comparison of two scoring systems." *British Journal of Medical Education*, 36(9), 820-826.

Williams, C.G. 1998. "Using concept maps to assess conceptual knowledge of function." *Journal of Research in Mathematical Education,* 29, 414–421.