Concept Maps: Making Learning Meaningful
Proc. of Fourth Int. Conference on Concept Mapping
Viña del Mar, Chile, 2010

BEYOND INDIVIDUAL CLASSROOMS: HOW VALID ARE CONCEPT MAPS FOR LARGE SCALE ASSESSMENT?

Sumitra Himangshu, Macon State College, Amy Cassata-Widera, Center for Elementary Math and Science Education, University of Chicago, United States

Email: sumitra.himangshu@maconstate.edu

Abstract. Reliability and validity are still important issues for considering the use of concept maps as assessment tools. The presence of different concept mapping formats/techniques, different forms of traditional testing methods, and a plethora of concept map scoring systems, add complexity to the act of linking student concept map scores to meaningful learning outcomes. The present study provides a descriptive analysis of recent research on the reported concurrent validity of concept map assessments in the K-16 science classroom with more traditional assessment measures. Findings suggest that various fill-in-the-map techniques correlated well with standardized assessments for all learners. In addition, evaluator training for cncept map assessment, and a tight match of content used for both conventional test and concept maps, appeared to have a positive effect on concurrent validity. The results suggest that in addition to the use of fill-in-the-map formats for large scale assessment in science education, a combination of both techniques select-and-fill-in and create-a-map techniques must be used in concert as formative and summative assessment tools, providing complementary information about the content and quality of student learning in science.

1 Introduction

Even after 40 years of use in science education there remains an ambiguity among researchers as to the power of concept mapping as an assessment tool, with little attention paid to the reliability and validity of variations in concept mapping techniques."Before concept maps are used for assessment and before map scores are reported to teachers, students, the public, and policy makers, research needs to provide reliability and validity information on the effect of different mapping techniques" (Ruiz-Primo & Shavelson, 1996, p. 569). Since this statement was published, scant empirical data has been reported supporting reliability and validity of concept maps as assessment tools. This leads to our question: Based on the extant literature, to what degree and under what specific conditions if any do concept maps correlate with summative measures of learning in science education? This paper provides a descriptive analysis of recent research on factors associated with the concurrent validity of concept maps used as outcome assessment tools in the K-16 science classroom. Specifically, this review illuminates challenges in applying consistent and reliable scoring methods to the various concept mapping tasks, highlights the variability in types of information about the learner gained using different types of concept mapping tasks, and emphasizes the need to align the content and processes measured by both concept maps and alternate assessments in order to making inferences about concept maps as valid, reliable measures of science learning.

2 Background

2.1 Scoring Concept Maps for Conceptual Understanding: Issues of Reliability

2.1.1 Assessment of expert knowledge in science

The current emphasis on contextual forms of assessment in science education provides a platform for concept mapping to move beyond assessing learning in individual classrooms and into the broader arena of large-scale assessment. However, a survey of the literature suggests that there are certain logistical and scoring reliability considerations that need to be addressed before the concept mapping community can move forward (Ruiz-Primo & Shavelson, 1996). This is not to say that large-scale assessment using concept mapping have not been previously attempted. Some existing research, including the authors' previous work, provides evidence for large-scale concept mapping, within certain given constraints of evaluator training, and similarity of instructional strategies and requirements. Even though

concept maps are able to demonstrate robust learning effects both for instruction and assessment, the main issues with using concept maps for large-scale assessment are the problems of maintaining content and scoring reliability across the board (Ruiz-Primo, Schultz, Li, & Shavelson, 2001).

An important aspect of contextual learning in science education is helping students assimilate Problem Based Learning techniques by combining the ability to think creatively (to generate ideas) and apply critical thinking (to evaluate ideas). Both modes of thinking are essential for a well-rounded understanding and application of science content. Content and a contextual framework are both important in order to evaluate the generation and ability to evaluate ideas. In this context, the ability to think about thinking as visualized by a concept map then becomes a powerful means of procuring students' understanding. Applied to science education, content and context refer to domain-specific knowledge and true-to-life application of this knowledge, respectively. Although disciplinary knowledge (content) is readily procurable (through a variety of sources), consensus by experts as to the ordinate and sub-ordinate features of such knowledge for teaching purposes is hard to come by. Maps by individual experts and instructors can be highly variable in both the generalities and specifics of the map. In addition, differences in instructional strategies and classroom requirements impact and influence what is learned and how it is learnt (context of learning). These factors in turn impact the degree of concept understanding for learners in the classroom and can result in reliability issues when scoring concept maps (Hollenbeck, Twyman & Tindal, 2006; Rice, Ryan & Samson,1998).

2.1.2 Variations in concept map scoring systems

Different scoring techniques measure different aspects of conceptual organization and/or understanding. Some of the scoring systems are based on scoring the accuracy and position of propositions while others are comparisons to a criterion map (Ruiz -Primo, Schultz, Li,&Shavelson, 2001). A large variety of scoring systems exist that add to the complexity of reliability, with some researchers in favor of not scoring maps to others designing elaborate computer-based scoring systems (Rice, Ryan,&Samson, 1998; Ruiz-Primo et al., 2001) Current research suggests that estimates of the reliability of scores are affected by the nature of the task, and the scoring scheme that make up the concept-map assessment. The more directed the task the easier the assessment is to grade and thus reliability and task directedness increase proportionally (Debrentseva et al., 2007; Hoeft, Jentsch, Harper, Evans, Bowers, & Salas, 2003; Kaya & Kilic, 2004; Ruiz-Primo et al., 2001; Schmidt, 2006; Shavelson et al., 2008).

The selection of concepts used in the mapping task can likewise impact both the restrictiveness and demands placed on the student. For example, a list of concepts that are highly related will pose different demands and constraints on student responses than a list of loosely associated concepts. One way to capture student knowledge structure without considering the spatial layout of the map is to consider two propositional attributes proposition choice/importance and proposition accuracy. If students are to pair concepts that have hierarchical relationships, then this would be a criterion for appropriate proposition choice. If the essential relationship between the two concepts is hierarchical in nature, then students would be expected to express a hierarchical relationship in the linking phrase in order for the proposition to be considered accurate.

Few studies prior to 1995 accounted for the reliability of mapping scores, instead, most studies reported inter-rater agreement (Ruiz-Primo et al., 1996). Not much has changed since their review of assessment literature with few studies considering other sources of measurement error such as, equivalence of test forms (Ruiz-Primo et al., 2001), internal consistency (Yin & Shavelson, 2005), score stability across test occasion, (Yin & Shavelson, 2005), and such as the consistency of student rankings between raters (e.g., McClure et al., 1999). Very few studies then have compared the reliability estimates of different scoring systems. Much of the work on the reliability estimates of scoring have involved comparison of two different or similar mapping formats, such as comparison of open-ended map construction with fill-in-the-map format (Ruiz-Primo et al., 2001) or comparison of two type of fill-in the map formats (Yin & Shavelson, 2005). McClure and colleagues (1999) reported a significant piece of research comparing six different types of scoring methods. Six pairs of independent raters used one of the following scoring methods: (a) holistic, (b) holistic with master map, (c) relational, (d) relational with master map, (e) structural, and (f) structural with master map. The holistic method involved rating student overall understanding on a scale from 1 to 10. The relational method used a three point scale to measure the accuracy of each proposition used by the student. Reliability coefficients for the six methods ranged from .23 to .76. The lowest correlations were reported from the structural scoring method using a master map for comparison and the highest correlations were reported by raters applying the relational scoring method

using a master map for comparison. According to the researchers some factors that might have served as sources of assessment error could include: (a) variations in assessor content knowledge/expertise, (b) differences in students' ability to follow concept-mapping conventions, and (c) the consistency of concept map evaluation (McClure et al., 1999). This last factor is largely dependent on the scoring method. In order to reach reliability of scoring among different scoring systems raters need to be able to balance content with the purpose of the assessment.

2.2 Variations in Concept Map Task Demands: Issues of Validity

2.1.2.1 Assessment of rote vs. procedural knowledge

The type of knowledge that is measured by traditional versus alternative assessment tools (concept map) adds to the complexity of using concept maps for assessment: to what extent do concept map assessments actually measure what students understand, or do they over- or under-estimate student understanding? Traditional assessment methods such as Standardized Tests usually examine a mix of rote learning and some procedural knowledge (citation for this?). Concept maps on the other hand, capture knowledge structure at any given point in a learner's trajectory and lay a greater emphasis on process in order to measure deeper understanding (Novak 1990; Novak & Gowin, 1984). Concept mapping assessments are described in the literature as having the potential to differentiate between learning due to rote memorization versus knowledge that is integrated to pre- and new conceptions. The best way to do this would be to assess student understanding using concept maps along a timeline at different points during the semester when rote learning has been forgotten. The lack of standardization of scoring methods among concept mappers and time constraints impacting classroom training, implementation, and assessment is perhaps why this very fundamental factor has not yet been tested to any great degree.

2.1.2.1.2 Differences in the level of student generativity among concept mapping tasks

The task demands inherent in the act of concept map creation, which greatly vary in level learner generativity, number and type of constraints, also need to be explored as potential factors influencing the potential for concept map scores to represent meaningful learning outcomes. The structure of the map, concepts, linking lines, and linking words may be provided within the task or may be learner-generated, with greater generativity by learner increasing the level of task demand. In addition, irrespective of concept map structure, the specific content of the concepts or linking words provided for learners to arrange or "fill in" can place both restrictive constraints and topic-related demands on the student. For example, a concept list that is loosely connected presents a different set of cognitive demands on a student than a tightly connected list of concepts. Consideration of two propositional attributes, proposition choice/importance and accuracy functions are important. Students' choice of hierarchical linking words between hierarchical concept pairs would demonstrate accurate understanding of the hierarchical nature of concept pairs (Schau et al., 2001). Finally, the wording of the question stem has the potential to impact the content and structure of the concept map. Derbentseva and colleagues (2007) report that most maps answer a question of, "what is…X?" which necessitates a description of concept X (for example, identifying component parts, specifying categories, specifying uses or functions). Other questions, such as "what happens when X changes" or "how does X work" requires the learner to think about how the concepts affect each other, eliciting more dynamic linking terms and cyclical map structures.

A review of the current literature indicates that there are multiple ways that concept maps have been used for assessing student learning in the science classroom, varying in both conceptual and structural generativity required by the student. Researchers in the K-16 environments have reported the use of task structures primarily falling into the following three main categories: (i) Create/Construct-a-map technique, (ii) Fill-in-a-map technique, and (iii) Create-a-map technique with either links created by student (C-mapping) or links (S-mapping) provided based on specific criterion (provide citation/s).

The "Create-a-map" technique (C-mapping) involves creating a concept map with unique structure and linking phrases based on concepts and terms that are common to a particular topic. There are several ways in which this can be achieved. One way, especially with young and novice learners, is to provide a 'parking lot' of concepts associated with a particular topic, much in the same manner as a vocabulary list provided in standards based lessons. The students use this concept list to generate a map using linking words and cross-links based on their individual or group understanding of the topic. Another method used in creating-a-map, especially with older and experienced learners is to simply rely on the individual's own knowledge regarding a particular topic and using that knowledge to create a map either

by themselves or by an interviewer translating the individual's ideas into a concept map. The use of unique linking words and cross-links used to create-a-concept map by either of these methods provides an insight into the individual's or groups' understanding of a related set of concepts at a given time. Depending on the nature of the task, this could be a holistic or a specific narrow topic. Although this type of open-ended map construction reveals preconceptions and misconceptions, and a picture of the students' knowledge structure, it is difficult to score unless compared tightly to an expert map following specific parameters. The main problem with scoring open-ended maps is the plethora of possibilities based on proposition number, proposition choice, and structure of concept relatedness that are possible due to individual understandings. This makes it difficult to score the maps and also creates reliability issues especially when compared to traditional assessments (Yin et al., 2005).

"Fill-in-a-map" formats use an expert-drawn concept map structure and either omit some concepts or some linking words. Students can fill-in these maps either by generating the words or by selecting from a provided list, a technique referred to as "Select-and-fill-in" (SAFI) (Schau et al., 1997; Schau, 2001). Each of the fill-in items can be a part of one or several inter-connected propositions. Depending on the type of format used, students can either choose a node term or link term both of which add to the visualization of specific domain knowledge. Although fill-in-the-map formats can correlate robustly to traditional assessments such as standardized tests and instruction-directed multiple choice tests, they are constrained by providing the students with the domain structure being assessed (Schau, 2001). This means that fill-in-the-maps might not represent students' individual understanding nor the connectedness of understanding for the domain of knowledge (Schau, 2001). The advantage to fill-in-the-map technique lies in its ability to differentiate between rote learning and conceptual learning, since rote learning is readily forgotten and conceptual understanding (even superficially) is required for connecting propositions. A more pertinent advantage of fill-in-the-map techniques is the reliability with which they can be scored when compared to the expert map.

The third technique, "C- and S-mapping" technique uses the create-a-map technique with either the concepts or linking words prescribed (Yin et al., 2005). Although this sounds similar to the constructing a map from a 'parking lot' of concepts, it is different because the prescribed concepts or linking words are based on a pre-set expert or criterion map. Both C- and S-mapping techniques use reliable scoring methods based on the criterion provided and can be easily adapted to an electronic format. The main disadvantage to these two methods lies in the inability to differentiate scores from a high and low performer (Yin et al., 2005). Unless the quality of propositions? are specified in a C-mapping technique, in addition to quantity, two individuals can receive the same score, one using concise propositions and the other using imprecise or superficial propositions (essentially guessing) that both result in accurate content. Examination of the concept map based on the use of key propositions and the quality of propositions used, in addition to number, would differentiate high performers from low performers. With respect to the S-mapping technique, student language skills can positively or negatively aspect use of linking words and only partially reflect individual understanding. Yin and colleagues (2005), suggest that in designing C- and/or S-mapping formats, expert criterion must consider student language skills, student vocabulary, and student misconceptions in order to provide valid measures of knowledge.

As these examples illustrate, while concept map performance scores may reflect deep, conceptual knowledge, they also are likely to represent the facility to which a student is able to engage in task-specific cognitive processes, such as organization and integration of information, comprehension monitoring, planning next steps, reflecting about conceptual relationships, and expressing knowledge. Identifying the cognitive processes inherent in various concept maps have important implications for establishing reliability and validity. First, in order to make an accurate correlation between concept map scores and scores to alternate measures, the level of generativity on the concept map needs to match that which is required in the alternate assessment. Second, the more directed a task, such as in fill-in-a-map formats, the easier it is to specify a scoring system and thus increase the reliability of the concept mapping assessment process.

The main purpose of the present study was to synthesize the current literature reporting the use of concept maps for assessment in the K-16 science arena, specifically investigating the extent to which concept maps are reported as valid measures of student learning when compared to traditional assessment measures. Specifically, we conducted a review of studies reporting the correlation between scores from student generated concept maps and other measures of achievement (concurrent validity), taking into account both the type of concept map assessment, the type of scoring system, and the match between the two.

3 Method

A review of articles including populations ranging from elementary to post-secondary was conducted. A thorough database search, carried out within PsycInfo, ISI, ERIC, and Google Scholar was implemented to locate all potential articles for review, using the following search criteria:
Topic=("concept map*" OR "knowledge map*" OR "mind map*") AND Topic=(validity OR validate) AND Topic=(science OR scient*).

Articles were selected for review which reported concept mapping tasks used for the purposes of providing information about student knowledge / conceptual understanding in science, and alternate assessments were given concurrently to the concept mapping task. The 25 studies selected ranged from elementary classrooms, middle school and secondary environments, through pre-service teacher preparation to medical school classrooms.

Current research on the use of concept maps for assessment in science education was coded using the ten criteria shown in Table 1.

Population	Sample Size	Prior-experience	Ability Level	Topic	Assessment Timing	Type of Cmap	Alternate Assessment	Scoring Method	Reliability

Table 1. Criteria used for coding research on concept maps as assessment tools

The coding allowed for a comparison of different concept mapping techniques used, denoted by type of concept map, to robustness – high or low to alternate assessment measures. This provided a measure for reliability of concept map techniques to traditional methods of assessment. Some examples are provided in Table 2.

Cmap Technique/Source	Alternate Assessment	Robustness of Correlation
1.Create-a-map from scratch /Kaya & Kilic, 2004	Standardized Test	Low (r = .478)
2.Fill-in-a-map/Schau et al., 1997	Multiple Choice Test	High (r = .75)
3.Create-a-map: C- and/or S-mapping/Yin et al., 2005	Instructor Designed Test	Modearate (r = .5 or .6) r value depends on C- or S-type

Table 2. Robustness of concept map techniques to alternate assessment measures

Those studies that showed a moderate to high measure of correlation between concept maps and alternate assessments were further investigated for logistical components such as: (i) study design, (ii) training of assessors, (iii) ability level of learners, (iv) similarity of instructional strategies between classrooms, and (v) cmap scoring method (McClure et al., 1999). Among others, these five factors can have an integral impact on measurement of classroom learning and are critical to the implementation of concept mapping in large scale assessment. Table 3 shows further analysis of example 3, used in Table 2.

Yin et al., 2005	Study Design	Assessor Training	Learner Ability Level	Similarity of Instruction	Cmap Scoring Method
Create-a-map	Robust – based on field testing	Advanced	Average-High	Controlled – single instructor in all 6 classes	Looked at propositions with respect to expert designed criterion

Table 3. Example of factors that impact use of cmaps in classroom assessment

4 Analysis

Table 4 summarizes the findings for this analysis with respect to concept map technique used to measure student understanding using the 25 selected studies and the degree of reliability based on correlations with alternate assessments. The range of correlations reported in these studies with respect to multiple choice assessments were moderate to high ranging from .475 to .80, with fill-in-the-map techniques being highly correlated to traditional assessment formats

such as multiple choice tests, short answer essays, and achievement tests (Ahlberg & Ahronta, 2008; Henno & Reiska, 2008; Hollenbech et al., 2006; Ingec, 2009; Klein et al., 2002; McClure et al., 1999; Schau et al., 1997; Schau et al., 2001; Yin et al., 2005).

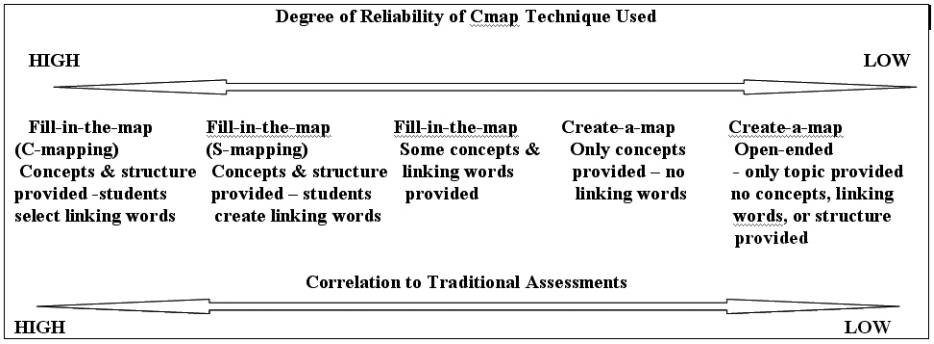

Table 4. Correlation of concept map techniques to traditional assessments as a measure of reliability

The correlations of multiple choice test and other conventional tests with create-a-map from scratch reported a range of reliability measures from low (r = .41) to high (r = .77) (Hollenbeck et al., 2006; Schau et al., 1997). The match of content being tested to content used to generate the concept maps appeared to greatly influence the summative analysis in these studies. For example, the Hollenbeck (2006) study reported a low correlation between student-generated concept maps and problem solving essays. The study suggested that there was a low rate of exchangeability between information on student maps and student essays. Although some of this result could be due to small sample size, and sample selection, other studies also reported low correlations between student-generated maps and multiple choice tests (Ingec, 2009; Liu & Henchley, 19961a). In all these cases, the low correlation actually provides validity for the concept mapping measure since there was either a disconnect between the material being tested or it was seen as a content/topic dependent effect i.e. force and solutions (Liu & Henchley, 19961a). By the same token, the stronger correlations are indicative of tighter matches between material used for the tests and the concept maps (Heno & Reiska, 2008; Rice et al., 1998). With respect to fill-in-the-map techniques, the more constrained the map structure, the higher was the reliability correlating to multiple choice tests, r = .74-.77 (Schau et al., 1997). For the SAFI method of mapping, the student map scores were appreciable when compared to multiple choice test scores (r = .61-.83), (Schau et al., 2001). The range of scores observed fell within reported correlation estimates for comparison of concept maps and multiple choice scores. Studies using the SAFI technique suggests that while the C-mapping technique is better for capturing student partial knowledge, the S-mapping technique is easier to score (Yin et al, 2005, Yin & Shavelson, 2008). In addition, a recent report by Shavelson and colleagues (2008) lends validity to using the fill-in-the map technique for formative assessment especially when embedded in the curriculum.

5 Conclusion

Our analysis suggests that "fill-in maps" were the most highly correlated with other, more traditional assessment measures. As an additional advantage, the "fill-in" task format lends itself to a reliable method of scoring for large scale assessments, when compared to open-ended maps constructed from scratch. Three aspects of the "fill-in maps" likely contribute to their scoring reliability: First, the task constraints are communicated through a clear quantification of concepts used and clear task expectations, ensuring that specific forms of knowledge are tapped (Debrentseva et al., 2007). Second,"fill-in maps" can be scored objectively according to an expert or criterion-referenced key (McClure et al., 1999). Third, "fill-in maps" can be easily matched to complement the content and process knowledge covered included within conventional tests.

However, while concurrent validity to the content of traditional exams is gained, perhaps some information regarding other forms of student knowledge is lost; by their nature, "fill in maps" are not designed to detect misconceptions or depict a student's unique structure of knowledge. Depending on the scope of the assessment, "fill-in maps" may not provide the potential for the student to make links to a larger conceptual framework within the science domain, to other academic domains, or to personal experience.

Based on our review, we propose that a combined method using a series of select-and-fill-in-the map (SAFI) technique as formative assessment together with create-a-map technique for summative assessments might be best for large scale assessment. This would require front-end preparation using electronic formats for ease of scoring but can be easily setup using available technology (Chang, Sung, Chang & Lin, 2005; Hoeft et al., 2003; Taricani & Clariana, 2006; Ruiz-Primo et al., 2001; Yin & Shavelson, 2005). The development of a scaffolded trajectory consisting of teacher training and professional development for formative and summative assessment, and designing assessment embedded in curriculum is critical to avoid disconnect and student confusion when tested by traditional and alternate methods of assessment (Ayala et al., 2008; McClure et al., 1999; Shau et al., 1997; Shavelson et al, 2008).

In order to address a timeline of reliable progression, i.e. to be able to differentiate between rote learning and conceptual understanding, it is necessary to measure formatively using fill-in-the-map techniques, where fill-in-the-links is more robust measure of student knowledge than fill-in-the concept technique (Ruiz-Primo, both 2001s; Schau et al.,, 1997), and then use a construct-a-map technique at major intervals to correlate with and supplement the information provided by the formative fill-in-the-map technique. (Shavelson et al., 2008). This approach would address both the reliability and validity of using concept maps as assessment of and for learning.

6 Acknowledgements

To the concept mapping community, teachers, and students, and their ongoing efforts.

References

Ayala, Carlos C., Shavelson, Richard J., Araceli Ruiz-Primo, Maria, Brandon, Paul R., Yin, Yue, Furtak, Erin Marie, Young, Donald B. and Tomita, Miki K.(2008) 'From formal embedded assessments to reflective lessons: The development of formative assessment studies', Applied Measurement in Education, 21: 4, 315 — 334.

Chang, K.-E., Sung, Y.-T., Chang, R.-B., & Lin, S.-C. (2005). A new assessment for computer-based concept mapping. Educational Technology & Society, 8 (3), 138-148.

Derbentseva, N., Safeyni, F. & Canas, Alberto, J. (2007). Concept maps experiments on dynamic thinking. Journal of Research in Science Teaching, 44(3), 448-465.

Hollenbeck, K., Twyman, T. & Tindal, G. (2006). Determining the exchangeability of concept map and problem solving essay score. Assessment for Effective Intervention, 31, 51-68.

Ingec, S.K. (2009). Analysing concept maps as an assessment tool in teaching physics and comparison with the achievement test. International Journal of Science Education, 31(14), 1897-1915.

Kaya, O.N. & Kilic, Z. (2004). Student-centered reliabilityconcurrent validity and instructional sensitivity in scoring

of students' concept maps ina university science laboratory. Poster presented at 18th International Conference on Chemical Education "Chemistry Education for the Modern World", İstanbul, TURKEY, 2004.

Klein, Davina, C.D., Chung, Gregory, K.W.K, Osmundson, E. & Herl, Howard, E. (2002). Examining the validity of knowledge mapping as a measure of elementary students' scientific understanding. CSE Technical Report No. 557, California.

Liu, X. & Hinchey, M. (1996). The internal consistency of a concept mapping scoring scheme and its effect on prediction validity. International Journal of Science Education, 18(8), 921-937.

McClure, J.R., Sonak, B. & Suen, H.K. (1999). Concept map assessment of classroom learning: Reliability, validity,

and logistical practicality. Journal of Research in Science Teaching, 36(4), 475-492.

Novak, J.D. (1990). Concept mapping: A useful tool for science education. Journal of Research in Science Teaching, 27, 937–949.

Novak, J.D.&Gowin, D.B. (1984). Learning how to learn. New York: Cambridge UniversityPress.

Rice, D.C., Ryan, J.M. & Samson, S.M. (1998). Using concept maps to assess student learning in the classroom: Must different methods compete? Journal of Research in Science Teaching, 35(10), 1103-1127.

Ruiz-Primo, M.A.& Shavelson, R.J. (1996). Problems and issues in the use of concept maps in science assessment. Journal of Research in Science Teaching, 33(6), 569-600.

Ruiz-Primo, M.A., Schultz, S.E., Li, M. & Shavelson, R.J. (2001). Comparison of the reliability and validity of scores from two concept mapping techniques. Journal of Research in Science Teaching, 38(2), 260-278.

Schau, C., Mattern, N. & Weber, R.W.(1997). Use of fill-in-concept maps to assess middle school students' connected understanding of science. Paper presented at Annual Meeting of the American Educational Research A ssociation, Chicago:IL.

Schau, C., Mattern, N., Zelik, M., Teague, W. & Weber, R.J. (2001). Select and fill-in concept map scores as measure of students' connected understanding of science. Educational and Psychological Measurement, 61, 136-158.

Schmidt, H.J. (2006). Alternative approaches to concept mapping and implications for medical education: Commentary on reliability, validity and future directions. Advances in Health Sciences Education, 11, 69-76.
Shavelson, R.J., Young, D.B., Ayala, C.C., Brandon, P.R., Furtak, E.M., Ruiz-Primo, M.A., Tomita, M.K. & Yin,Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. Applied Measurement in Education, 21(4), 295-314.

Taricani, E.M. & Clariana, R.B.(2006). A technique for automatically scoring open-ended concept maps. Educational Technology, Research and Development, 54(1), 65-82.

Yin, Y., Vanides, J., Ruiz-Primo, M.A., Ayala, C.C. & Shavelson, R.J. (2005) .Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use. Journal of Research in Science Teaching, 42(2), 166-184.

Yin, Y. & Shavelson, R.J. (2008). Application of generalizability Theory to concept map assessment research. Applied Measurement in Education, 21(3), 273-291.

Facilitation of Learning; Knowledge Management; Research Planning; Instructional Design; Brainstorming