



A Statistical Study of Concept Mapping Metrics

Dr. Kathryn W. Jablokow, Pennsylvania State University

Dr. Kathryn Jablokow is an Associate Professor of Mechanical Engineering and Engineering Design at Penn State University. A graduate of Ohio State University (Ph.D., Electrical Engineering), Dr. Jablokow's teaching and research interests include problem solving, invention, and creativity in science and engineering, as well as robotics and computational dynamics. In addition to her membership in ASEE, she is a Senior Member of IEEE and a Fellow of ASME. Dr. Jablokow is the architect of a unique 4-course module focused on creativity and problem solving leadership and is currently developing a new methodology for cognition-based design. She is one of three instructors for Penn State's Massive Open Online Course (MOOC) on Creativity, Innovation, and Change, and she is the founding director of the Problem Solving Research Group, whose 50+ collaborating members include faculty and students from several universities, as well as industrial representatives, military leaders, and corporate consultants.

Dr. Joanna F. DeFranco, Pennsylvania State University, Great Valley

Joanna F. DeFranco is Assistant Professor of Software Engineering and a member of the Graduate Faculty at The Pennsylvania State University. Prior to joining Penn State, she held faculty positions at Cabrini College and the New Jersey Institute of Technology. She also held a number of positions in industry and government including an Electronics Engineer for the Naval Air Development Center in Warminster, PA and a Software Engineer at Motorola in Horsham, PA. Dr. DeFranco received her B.S. in Electrical Engineering from Penn State University, M.S. in Computer Engineering from Villanova University, and Ph.D. in Computer and Information Science from the New Jersey Institute of Technology. She is a member of ASEE and has had numerous publications in journals and conference proceedings. She is also on the curriculum advisory board for a local technical high school.

Mrs. Sally Sue Richmond, Penn State Great Valley School of Graduate Professional Studies

A Statistical Study of Concept Mapping Metrics

Abstract

This paper reports on our exploratory statistical study of the relationships among a selection of traditional and holistic concept mapping metrics. Specifically, twelve traditional metrics (nine from the literature and three of our own design) and four holistic metrics (all from the literature) were used to evaluate the concept maps of two sets of engineering students: (i) 73 undergraduate engineers enrolled in first-year engineering design; and (ii) 52 graduate engineers enrolled in a master's level systems engineering course. Our results showed 41 common correlated pairs between these two samples, all at the $p < 0.05$ level of significance. The implications of these correlations for engineering educators are discussed, along with potential directions for future research to expand this area of study.

1. Introduction

The use of concept maps in engineering education research is growing, with applications in the assessment of knowledge mastery and integration within courses, programs, and across multiple disciplines^{2,7,12,13,14,19,21}. Concept maps are also being used to assess the effectiveness of the early stages of engineering problem solving and analysis³. Within these applications, a variety of metrics have been developed for assessing concept maps, including both “traditional” and “holistic” approaches to measuring the accuracy, breadth, and depth of students’ understanding².

In general, *traditional metrics* rely on counting various elements or features of a concept map (e.g., number of concepts, links, or hierarchies) or the computation of map descriptors (e.g., map density, map complexity) as functions of these elements. Due to their dependence on relatively clear-cut characteristics, traditional metrics are generally considered to be quite objective (i.e., different map evaluators are likely to derive the same results). Nevertheless, *holistic metrics* that focus on a more subjective “quality of understanding” represented in a concept map (rather than the “quantity” of specific features) have also emerged. These holistic scoring methods include structural complexity approaches that assess the dominant structural patterns of concept maps (e.g., hub/spoke, tree, network), as well as integrated rubrics based on the organization, comprehensiveness, and correctness of map content^{2,12}.

Despite the wide variety of concept mapping metrics available, very few studies have examined the relationships between them in detail. As a result, we have little sense of which metrics (if any) are truly independent or whether there is significant overlap between them. In addition, without a detailed analysis of the relationships among the available metrics, it is difficult to know which metrics might be best at explaining the variance we observe among students’ concept maps. We need to have a better understanding of these relationships. To address this need, we performed an exploratory statistical analysis to determine if and how the predominant traditional and holistic concept mapping metrics are correlated, with the *future aim* of identifying optimal sets of metrics that are most effective for the evaluation of students’ understanding in specific situations or contexts.

2. Background and Literature Review

We began our study with a literature review to identify the most commonly used concept mapping metrics within educational settings. Previous reviews of concept mapping theory and applications include Ruiz-Primo and Shavelson's investigation of concept map tasks¹⁷, response formats, and scoring systems, and Besterfield-Sacre et al.'s overview of concept map terminology, scoring approaches, and mapping applications in engineering²; both were excellent resources. We also examined Bayram's weighted scoring system based on a map's hierarchical levels, propositions, and branches¹, Ruiz-Primo and Shavelson's work in assessing declarative knowledge¹⁶, and Turns et al.'s exploration of the breadth, depth, and connectedness of concept maps¹⁹. From our review, it was evident that while concept mapping is fairly standardized as an activity, the metrics used to assess concept maps are less so. The results of our search are summarized in Appendix A, where both traditional and holistic approaches to measuring the accuracy, breadth, and depth of students' understanding are shown, along with selected references for each metric listed.

As shown in Appendix A, *traditional metrics* enumerate the features of a concept map, such as the number of concepts, propositions, cross-links, concept examples, hierarchies, or the highest level of hierarchy achieved – counted separately or in combination, with equal or variable weighting^{2,5,7,14,15,16,19}. Other traditional metrics involve the computation of various map descriptors (e.g., map density, map complexity, overlap, link similarity) as functions of the individual element counts^{4,6,14}. Still others focus on organizational and semantic assessments based on connectedness, surface structure, concept usage, and vertex matching, among other map characteristics^{8,11}.

Among the *holistic metrics* for concept maps, two main scoring methods have surfaced in the past decade: (1) Kinchin and Hay's structural complexity approach^{9,10,12,22}, which assesses the dominant structural patterns of concepts and links found in a map; and (2) Besterfield-Sacre et al.'s integrated rubric based on the comprehensiveness, organization, and correctness of map content². Other holistic scoring methods based on the evaluation of subject understanding on a simple scale from, say, 1 to 10 have also been presented^{2,7,14}.

Based on our literature review, we began by choosing nine (9) traditional metrics and four (4) holistic metrics to evaluate for each concept map. In addition, based on our desire to investigate students' responses to task structure (a topic that will be addressed in a future publication), we defined three more traditional metrics, namely: (i) the number of concepts used from the list of concepts provided in the task instructions; (ii) the number of concepts added to the list (by a subject); and (iii) the number of concepts unused from the given list. These 16 metrics are listed in Table 1, along with the accompanying equations for any computed metrics (e.g., map density).

TABLE 1. Concept map metrics used in this study

<i>Traditional Metrics</i>	<i>Holistic Metrics</i>
Total number of concepts (<i>C</i>)	Dominant structural pattern: Linear, Chain,
Total number of links (<i>L</i>)	Hub/Spoke, Tree, or Network
Total number of hierarchies	Comprehensiveness
Highest level of hierarchy constructed	Organization
Number of cross-links between hierarchies	Correctness
Map density = $L \div [C \times (C - 1)]$	
Map complexity = $L \div C$	
Number of matching links (with expert map reference) (<i>M</i>)	
Link similarity = $M \div (L + E - M)$; <i>E</i> = total # of expert links	
*Concepts used from given list	
*Concepts unused from given list	
*Concepts added to given list	

*=New metric defined by the authors.

3. Experimental Methods

3.1 Concept Mapping Tasks

As Gregoriades et al.⁷ have described, there is a range of directedness involved in providing information to subjects during concept mapping exercises; this varies from *high-directed* to *low-directed*. High-directed concept map tasks pre-define the specific concepts, connecting lines, linking phrases, and map structure involved, while low-directed tasks leave subjects free to decide which/how many concepts to include in their maps and how those concepts are related (including the propositions used to label each link). In our study, concept mapping tasks were *moderately-directed*, meaning that students were given a list of specific concepts related to the map topic, but they were also allowed to add or subtract concepts from that list. No connecting lines, linking phrases, or map structures were provided, so students were free to organize the concepts in whatever configuration they felt was most appropriate. A sample set of instructions (for a systems thinking mapping task) is shown in Figure 1.

Map #2: Systems Thinking Name _____

Use the terms listed below to create a concept map of your current understanding of Systems Thinking. You may add and/or delete terms from the list in creating your map. Don't forget to label the links between concepts! Draw the map on this page.

1. Parts	8. Constraints	11. Inputs
2. Interactions	9. Information	12. Processes
3. Systems	10. Energy	13. Constraints
4. Hierarchy	11. Materials	14. Cycles
5. Functions	12. Feedback	15. Control

Figure 1. Task instructions for a systems thinking concept map

3.2 Study Participants

Our study participants included 73 undergraduate engineering students enrolled in a first-year engineering design course (3 sections) and 52 graduate engineering students enrolled in a master's level systems engineering course (2 sections) at Penn State University. Students volunteered to participate based on a description of our research project and received no compensation for their participation. Each student completed a concept map of a course-related topic as a class exercise mid-way through each course; the topics were systems thinking (undergraduates) and creativity (graduates), respectively. All students were provided with brief instructions about concept mapping and performed at least one "practice map" before completing the maps of interest; they were given approximately 30 minutes to complete each mapping task.

3.3 Map Evaluation

Each map was scored by two independent evaluators, with at least one expert in the subject matter; the maps were not graded for formal course credit. When it was required for the evaluation of a particular map metric (e.g., matching links), an expert map was provided by the course instructor. In cases where evaluators disagreed on a particular metric's value, a consensus opinion was sought through discussion²⁰; when a consensus opinion could not be reached, the evaluators' individual metrics were averaged. The evaluators assessed each map using the metrics listed in Table 1; ranges of possible values for these metrics were assigned according to original definitions from the appropriate scholarly sources (see Appendix A). As noted earlier, we added three traditional metrics to those identified in the literature to assess the number of concepts used, unused, and added to the list of concepts provided for each task; the ranges of possible values for these metrics were a direct function of the number of concepts provided in those lists. Our aim in including these new metrics was based on our interest in how students respond to the structure of the mapping task itself, a topic we will take up in a future publication. Pearson correlations (95% confidence limit) were performed using Minitab[®] and Matlab[®] software to determine if and how the map metrics were correlated among and between the traditional and holistic subsets. These results are discussed in detail below.

4. Results

Our analyses revealed 73 statistically significant correlations ($p < 0.05$) for the metrics assessed using the undergraduates' systems thinking concept maps and 60 statistically significant correlations ($p < 0.05$) for the metrics assessed using the graduates' creativity concept maps. Of these correlated metrics, 41 pairs were common to both data sets (see Tables 2-4); our discussion here will focus on these common correlations. Of these common pairs, 20 were correlations among the traditional metrics, 5 were correlations among the holistic metrics, and 16 were correlations between traditional and holistic metrics.

Beginning with some general observations: all correlations in Tables 2-4 are of the same sign across the two samples, even when their magnitudes are noticeably different. In addition, most of the correlations are moderate ($0.3 < |r| < 0.7$) to strong ($|r| \geq 0.7$), with all $|r| > 0.23$. In most cases ($30/41 = 73\%$), the correlations are higher for the metrics obtained from the graduate creativity maps than for the undergraduate systems thinking maps. In particular, of the twenty examples where $|r| \geq 0.7$, fourteen of these were identified from the creativity maps, while six occurred for

the systems thinking maps. We also note that while the strongest correlations occurred between traditional metrics, strong correlations were also observed between holistic metrics and between traditional and holistic metrics.

TABLE 2. Common correlations (creativity and systems thinking maps) among traditional metrics ($p < 0.05$)

<i>Traditional-Traditional Correlations</i>		Creativity		Systems Thinking	
		<i>r</i>	<i>p</i> -value	<i>r</i>	<i>p</i> -value
Concepts Added	Link Similarity	-0.396	0.004	-0.236	0.046
Concepts from list	Unused Concepts	-1.000	0.000	-1.000	0.000
	Map Density	-0.781	0.000	-0.627	0.000
	Highest Level	0.410	0.003	0.314	0.007
	Link Similarity	0.609	0.000	0.301	0.010
	Matching links (expert map)	0.683	0.000	0.410	0.000
	Total Links	0.813	0.000	0.672	0.000
Map Complexity	Total Cross-links	0.559	0.000	0.691	0.000
Map Density	Highest Level	-0.372	0.007	-0.242	0.041
Matching links (expert map)	Link Similarity	0.980	0.000	0.977	0.000
Total Concepts	Unused Concepts	-0.957	0.000	-0.888	0.000
	Map Density	-0.846	0.000	-0.698	0.000
	Highest Level	0.428	0.002	0.344	0.003
	Matching links (expert map)	0.595	0.000	0.254	0.031
	Total Links	0.825	0.000	0.661	0.000
	Concepts from list	0.957	0.000	0.888	0.000
Total Links	Total Cross-links	0.387	0.005	0.564	0.000
	Highest Level	0.444	0.001	0.247	0.036
	Map Complexity	0.660	0.000	0.776	0.000
	Matching links (expert map)	0.694	0.000	0.265	0.025

TABLE 3. Common correlations (creativity and systems thinking maps) among holistic metrics ($p < 0.05$)

<i>Holistic-Holistic Correlations</i>		Creativity		Systems Thinking	
		<i>r</i>	<i>p</i> -value	<i>r</i>	<i>p</i> -value
Comprehensiveness	Organization	0.742	0.000	0.349	0.003
Hub pattern	Tree pattern	-0.723	0.000	-0.436	0.000
Linear pattern	Tree pattern	-0.312	0.025	-0.267	0.023
Organization	Network pattern	0.696	0.000	0.581	0.000
Tree pattern	Network pattern	-0.334	0.015	-0.537	0.000

TABLE 4. Common correlations (creativity and systems thinking maps) between traditional and holistic metrics ($p < 0.05$)

<i>Traditional-Holistic Correlations</i>		Creativity		Systems Thinking	
		<i>r</i>	<i>p</i> -value	<i>r</i>	<i>p</i> -value
Concepts from list	Organization	0.591	0.000	0.234	0.048
	Comprehensiveness	0.789	0.000	0.377	0.001
Highest Level	Hub pattern	-0.305	0.028	-0.274	0.020
Link Similarity	Correctness	0.348	0.012	0.639	0.000
Map Complexity	Linear pattern	-0.453	0.001	-0.257	0.029
	Network pattern	0.516	0.000	0.584	0.000
	Organization	0.587	0.000	0.489	0.000
Matching links (expert map)	Correctness	0.340	0.014	0.681	0.000
	Organization	0.493	0.000	0.250	0.034
Total Concepts	Comprehensiveness	0.806	0.000	0.411	0.000
Total Cross-links	Network pattern	0.365	0.008	0.729	0.000
	Organization	0.458	0.001	0.624	0.000
Total Links	Network pattern	0.418	0.002	0.509	0.000
	Linear pattern	-0.359	0.009	-0.275	0.019
	Organization	0.739	0.000	0.496	0.000
	Comprehensiveness	0.835	0.000	0.379	0.001

4.1 Discussion

A detailed discussion of every correlation and its implications would be impractical here, so we will comment on those that are the most “obvious” and those that are most “interesting” (in our view). Beginning with the strong correlations ($|r| \geq 0.7$) from Table 2, we note first the obvious negative correlation ($r = -1.0$) between ‘concepts from list’ and unused concepts (both traditional metrics). As one is equal to the total number of concepts provided to the students for a particular topic (a constant) minus the other, this relationship is no surprise. Other correlations involving computed metrics and one of their factors include:

- Matching links and link similarity (strong positive correlations)
- Total concepts and map density (moderate to strong negative correlations)
- Total links and map complexity (moderate to strong positive correlations)

In each of these cases, the direction/sign of the correlation is as expected; for example, as the number of total concepts increases, we expect map density ($L \div [C \times (C - 1)]$) to decrease, as confirmed by the negative correlation we observed.

Several interesting correlations can also be seen in Table 2, including (a) the moderate positive correlations between map complexity and the total number of cross-links, and (b) the strong positive correlation between the total number of concepts and the concepts used from the list. In the latter case, it is clear that most students relied on the given list of concepts to complete their maps, and they used most of those concepts as well.

Moving to correlations involving two holistic metrics (Table 3), the negative correlations among the various structural patterns (hub, tree, linear, etc.) are somewhat obvious; if a map is assessed as reflecting one pattern, it is less likely to reflect another. We also note the following interesting relationships in Table 3:

- Comprehensiveness and organization (moderate to strong positive correlations)
- Organization and network pattern (moderate positive correlations)

As summarized in Appendix A, comprehensiveness is related to a student's breadth and depth of knowledge of the map topic, while organization reflects their systematic arrangement of concepts and the hierarchy of concept placement. From our results, it appears that a more systematic configuration of concepts and a more comprehensive definition of the subject area may go hand-in-hand among our students. In addition, network patterns may be perceived as "more organized" than the other structural patterns (e.g., hub, tree, etc.).

Finally, moving to correlations involving traditional and holistic metrics (Table 4), the moderate to strong positive correlations between 'concepts from the given list' and comprehensiveness seem fairly obvious, although the difference in *magnitude* between the two samples (graduates and undergraduates) is intriguing. Similar observations hold true for the moderate positive correlations between link similarity (also matching links) and correctness, and for the moderate to strong positive correlations between 'total number of concepts' and comprehensiveness. In general, the difference in *r* values for the graduate and undergraduate samples clearly requires further investigation with larger sample sizes to determine the underlying reasons for these differences. We might speculate, for example, that they are a function of age/maturity, or they may be a result of instructional variations between the two courses. Of course, it is also possible that these differences will effectively disappear when larger sample sizes are used.

In Table 4, we also found the following correlations of interest:

- Total links and organization (moderate to strong positive correlations)
- Total links and comprehensiveness (moderate to strong positive correlations)

In each case, a greater number of links is associated with desirable holistic qualities, i.e., a more systematic arrangement of concepts (higher degree of organization) and a more complete definition of the subject area (higher level of comprehensiveness). A practical implication of this particular finding could be the development of techniques that help students make more connections between concepts in their maps, rather than focusing only on the correctness of the connections made.

As a further step in our analysis, we grouped our findings based on related correlations that, when taken together, lead to several additional observations about students' concept mapping (and map metrics) which may be of interest and/or which deserve subsequent investigation. These related correlations and the resulting observations are presented in Table 5. While the preliminary nature of this study requires us to remain somewhat tentative in our remarks, the exercise is still instructive in providing insights into the ways in which metrics might be combined to yield information about student learning via concept maps.

TABLE 5. Related correlations and resulting general observations

Related Correlations (Correlation Strength)	General Observations
<i>Traditional-Traditional Pairs</i>	
Concepts added – link similarity (mod. neg.) Concepts from list – link similarity (mod. pos.) Concepts from list – matching links (mod. pos.)	As concepts are added to a student’s map, the similarity between the student and expert maps decreases, since the expert only utilizes concepts from the list.
Concepts from list – map density (strong neg.) Concepts from list – total links (strong pos.)	The number of links and the number of concepts both affect map density. As concepts are added, the number of links increases as well, but not always enough to increase map density proportionately.
Map complexity – total cross-links (mod. pos.) Total Concepts – highest level (mod. pos.)	More links indicates a more complex map, which would include cross-links. Given the moderate correlations here, this may indicate that the students’ maps were not highly complex.
Total concepts – unused concepts (strong neg.) Total concepts – concepts from list (strong pos.) Total links – map complexity (mod. to strong pos.)	These correlations show that very few students left out concepts from the given list.
<i>Traditional-Holistic Pairs</i>	
Concepts from list – organization (weak to mod. pos.) Concepts from list – comprehensiveness (weak to strong pos.)	These positive correlations indicate that as students used more concepts from the given list, the maps appeared to have better organization and greater comprehensiveness.
Map complexity – linear pattern (weak to mod. neg.) Map complexity – network pattern (mod. pos.) Map complexity – organization (mod. pos.)	These correlations make sense since map complexity is increased by more links than concepts. More links are required for the network pattern, while fewer are needed for the linear pattern. Organization requires the integration of branches, which implies a network pattern.
Matching links – correctness (weak to mod. pos.) Total cross-links – network pattern (weak to strong pos.) Total cross links – organization (mod. pos.) Total links – network pattern (mod. pos.)	Having more cross-links leads to more networks; organization is related to integration as well. Increased links and increased network patterns, coupled with strong correctness and matching links, could imply an understanding of the topic.
Total concepts – comprehensiveness (mod. to strong pos.) Total links – organization (mod. to strong pos.) Total links – comprehensiveness (weak to strong pos.)	The correlations for total concepts-comprehensiveness and total links-comprehensiveness are similar. As total links increase, so do organization and comprehensiveness.

Note: The following abbreviations are used within Table 5 for convenience: mod = moderate, neg = negative, pos = positive.

5. Limitations, Implications, and Future Work

The primary aim of this exploratory study was to investigate the statistical relationships between concept mapping metrics as a preliminary step toward identifying optimal sets of metrics that might be effective in evaluating our students' understanding in specific situations or contexts. Clearly, we are far from our final goal, but we have learned a good deal. In particular, the large number of common, statistically significant correlations between metrics from our two samples (with all of them in the same direction, and many of similar magnitude) tells us that the metrics typically used to evaluate concept maps have some interdependencies. While the small sample sizes used in this study limit our conclusions about those interdependencies (which limits, in turn, our ability to recommend the use of certain metrics over others), the results are still encouraging and suggest an expanded study with larger sample sizes. In light of these findings, we also note that markedly different ways of interpreting concept maps may be needed to see if other dimensions of meaning can be extracted from them.

The strong correlations observed within each of the three pairings of metrics (i.e., traditional-traditional, holistic-holistic, and traditional-holistic) also suggest that the optimal sets of metrics we seek may be a combination of traditional and holistic metrics, rather than one type or the other. This finding has interesting implications for future research – for example, in the development of new holistic metrics to supplement those already in the literature. At present, the number of traditional metrics is greater than the number of holistic metrics; the potential power of combining the two approaches suggests that this imbalance should be addressed. To support this integrated strategy, we are currently developing a detailed set of guidelines for coding concept maps using the full spectrum of traditional and holistic metrics discussed in this paper that will assist engineering educators in their adoption of a wider range of map assessments.

Even now, an awareness of the interdependencies among map metrics can help educators choose between metrics when assessment efficiency and/or confirmation are important. For example, if particular traditional and holistic metrics are moderately to strongly correlated (e.g., total concepts and comprehensiveness), then an educator could choose to use the traditional metric instead of the holistic metric based on the practical fact that traditional metrics are generally less subjective. Or, if an educator wants to confirm the complexity of a student's knowledge through concept mapping, he/she could calculate map complexity (traditional metric) and compare the result with the structural pattern and organization of the map (both correlated holistic metrics).

Finally, the differences in magnitude of the common correlations between our two samples need to be explored further. As noted earlier, the correlations were higher for the metrics obtained from the graduate creativity maps than for the undergraduate systems thinking maps, but the reasons behind this phenomenon are not clear. We might hypothesize that the graduate students – who, in this case, are all working adults with an average age of (circa) 30 – are more “settled” or “mature” in terms of their thinking patterns, leading to stronger relationships among the map metrics, but this is only speculation. In our future work, which will include larger sample sizes across a wide range of academic levels and age groups, we hope to answer this question with greater certainty.

The next steps in our research will also include an investigation of the impact of cognitive preferences on students' mapping outcomes. In particular, a study of the correlations between concept mapping metrics (both traditional and holistic) and cognitive style is already underway. Understanding the relationships between different concept mapping metrics – and between those metrics and the individual cognitive differences of the students who create the maps themselves – marks the first steps toward helping educators make informed choices about which metrics they use to assess these student outcomes.

References

- [1] Bayram, S. 1995. The Effectiveness of Concept and Software Mapping for Representing Student Data and Process Schema in Science, Ed.D. Dissertation, University of Pittsburgh.
- [2] Besterfield-Sacre, M., Gerchak, J., Lyons, M., Shuman, L. J., and H. Wolfe. 2004. Scoring concept maps: An integrated rubric for assessing engineering education. *Journal of Engineering Education*, 93(2):105–115.
- [3] DeFranco, J., Neill, C., and R. Clariana, R. 2011. A Cognitive Model to Improve Team Performance — A Study of Team Outcomes and Mental Model Sharing. *Systems Engineering Journal* 14(3): 267–278.
- [4] DeFranco, J. F., Neill, C. J. 2011. Problem-Solving Style and its Impact on Engineering Team Effectiveness. *Proceedings of 2011 Annual Conference on Systems Engineering Research*.
- [5] Dorough, D. K. and J. A. Rye. 1997. Mapping for understanding. *The Science Teacher* January, 64(1):36–41.
- [6] Graff, M. 2005. Differences in concept mapping, hypertext architecture, and the analyst-intuition dimension of cognitive style. *Educational Psychology* 25(4):409–422.
- [7] Gregoriades, A., Pampaka, M., and H. Michail. 2009. Assessing students' learning in MIS using concept mapping. *Journal of Information Systems Education* 20(4):419–430.
- [8] Hay, D. B. and I. M. Kinchin. 2006. Using concept maps to reveal conceptual typologies. *Education & Training* 48(2/3):127–142.
- [9] Hay, D., and I. Kinchin. 2008. Using concept mapping to measure learning quality. *Education + Training* 50(2): 167-182.
- [10] Hay, D. B., Wells, H., and I. M. Kinchin. 2008. Quantitative and qualitative measures of student learning at university level. *Higher Education* 56:221-239.
- [11] Ifenthaler, D., Masduki, I., and N. M. Seel. 2011. The mystery of cognitive structure and how we can detect it: Tracking the development of cognitive structures over time. *Instructional Science* 39: 41–61.
- [12] Kinchin, I. M., Hay, D. B., and A. Adams. 2000. How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational Research*, 42(1):43–57.
- [13] McCarthy, J., Abatan, A., Parker, R., and Besterfield-Sacre, M. 2005. “Work in progress: Pillars of chemical engineering.” *Proc. ASEE/IEEE Frontiers in Education Conference*.
- [14] McClure, J. R., Sonak, B., and H. K. Suen. 1999. Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, 36(4): 475–492.
- [15] Novak, J. and D. Gowin. 1984. *Learning How to Learn*. New York: Cambridge University Press.
- [16] Ruiz-Primo, M., Shavelson, R. J., Li, M., and Schultz, S. E. 2001. On the validity of cognitive interpretations of scores from alternative concept mapping techniques. *Educational Assessment* 7(2): 99–141.

- [17] Ruiz-Primo, M., and R. Shavelson. 1996. Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching* 33(6): 569–600.
- [18] Rye, J. A., and P. A. Rubba. 2002. Scoring concept maps: An expert map-based scheme weighted for relationships. *School Science and Mathematics*, January, 102(1):33–44.
- [19] Turns, J., Atman, C., and R. Adams. 2000. Concept maps for engineering education: A cognitively motivated tool supporting varied assessment functions. *IEEE Transactions on Education* 43(2): 164–173.
- [20] Trochim, W. 2001. *The Research Methods Knowledge Base* (2nd ed.). Cincinnati, OH: Atomic Dog Publishing.
- [21] Walker, J.M.T., and King, P. H. 2003. Concept mapping as a form of student assessment and instruction in the domain of bioengineering. *Journal of Engineering Education*, April, 92(2):167-179.
- [22] Yin, Y, Vanides, J. Ruiz-Primo, M.A., Ayala, C. C., and R. J. Shavelson. 2005. Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use. *Journal of Research in Science Teaching*, 42(2):166–184.

APPENDIX A. Concept Map Assessment: Review of Scoring Methods, Metrics, and Rubrics

General Approach	Metric / Rubric Element(s)	Selected Reference(s)
<i>Traditional / Quantitative</i> - Single metrics	Number of (relevant) concepts Number of (valid) propositions Number of (appropriate) total cross-links Number of examples of specific concepts Number of hierarchies Highest level of hierarchy achieved	Novak & Gowin (1984) [†] Dorough & Rye (1997) [†] Besterfield-Sacre et al. (2004) [†]
<i>Traditional / Quantitative</i> - Combined metrics - Computed metrics - Weighted metrics	Structural scoring methods (combinations of, e.g., number of concepts, hierarchies, total cross-links, examples, and/or correct propositions) Set theoretic method: similarity between two maps (participant and expert) Relational scoring methods (e.g., propositional accuracy, proportion of valid propositions, link descriptions, link directions) Weighting of elements counted (e.g., concepts, propositions, levels of hierarchy, etc.) depending on importance of the element and/or aim of the experiment Map density = $(\#links) \div (\#concepts)(\#concepts - 1)$: a measure of concept integration Map complexity = $(\#links) \div (\#concepts)$: a measure of map breadth Visualization of concept usage using pattern analysis Organizational metrics: surface structure, graphical structure, connectedness, ruggedness, avg. degree of vertices, number of cycles, vertices Semantic content metrics: vertex matching, propositional matching Overlap analysis: number of shared concepts across multiple maps Link similarity: $(\#shared\ links\ between\ maps) \div (union\ of\ all\ links)$	McClure et al. (1999) ^{*†} McClure et al. (1999) [*] McClure et al. (1999) ^{*†} Ruiz-Primo & Shavelson (2001) [*] Gregoriades et al. (2009) [*] Turns et al. (2000) ^{*†} Rye & Rubba (2002) [*] Gregoriades et al. (2009) [*] Graff (2005) [†] Gregoriades et al. (2009) [*] Ifenthaler et al. (2011) [†] DeFranco, Neill, and Clariana (2011) [†]

*Holistic /
Qualitative*

General holistic scoring method: overall understanding of the map topic rated on a scale from 1 to X

McClure et al. (1999)*†
Besterfield-Sacre et al.
(2004)†
Gregoriades et al. (2009)*

Structural complexity: dominant structural pattern of concepts and links

- *Chain/Linear*: linear sequence of concepts (no branching)
- *Circular*: concepts daisy-chained with the ends joined
- *Hub/Spokes*: radial structure in which all the related aspects of the topic are linked directly to the core concept, but are not directly linked to each other.
- *Tree*: a linear chain with branches attached
- *Net(work)*: a highly integrated, complex, interconnected set of propositions

Kinchin, Hay, & Adams
(2000)†
Yin et al. (2005)†
Hay & Kinchin (2006)†
Hay, Wells, & Kinchin
(2008)†

Integrated Rubric: 3 categories, 3 levels per category

- *Comprehensiveness*: definition of the subject area, level of knowledge in the area, breadth and depth of that knowledge
- *Organization*: systematic arrangement of concepts, hierarchy of concept placement, connections and integration of branches
- *Correctness*: accuracy of the material presented (considering level of sophistication, use of appropriate terms, any misconceptions)

Besterfield-Sacre et al.
(2004)†

* = with expert/master map; † = without expert/master map