Concept Mapping: Connecting Educators
Proc. of the Third Int. Conference on Concept Mapping
A. J. Cañas, P. Reiska, M. Åhlberg & J. D. Novak, Eds.
Tallinn, Estonia & Helsinki, Finland 2008

CONCEPT MAPS AND SHORT-ANSWER TESTS: PROBING PUPILS' LEARNING AND COGNITIVE STRUCTURE

Mauri Åhlberg & Vuokko Ahoranta, University of Helsinki. Finland
Email: mauri.ahlberg@helsinki.fi , www.helsinki.fi/people/mauri.ahlberg

Abstract. The purpose of our research was to study what concept maps reveal about pupils' learning compared to teacher-designed school achievement tests (short-answer questions). Twenty pupils (10–13 years), with ample of experience constructing concept maps, had projects in which they constructed a concept map at the beginning and end of a learning project. Achievement tests were used in the balanced statistical design of four replications. Twice the achievement test was presented after the concept maps and twice the achievement test was given between the concept maps. Nine of the student answers were transformed proposition by proposition into concept maps. The numbers of relevant concepts and propositions were used as measures of meaningful learning. The reliability varied 0.73–0.95. The mode of knowledge representation accounted for variation of measures of meaningful learning 45–89 per cent. (Eta squared coefficients varied from 0.45–0.89.) Individually, pupils had more relevant concepts and propositions in teacher-made tests than they were able to reveal by their own concept maps. By their self-constructed concept maps pupils reveal what they think are the most relevant items in their metacognition. When a teacher helps pupils by asking short questions, the same pupils always write statistically significantly more, and it is evident that they have learnt much more than revealed by their concept maps. They express partly different and complementary knowledge in concept maps and in teacher-made achievement tests. The result suggests that it would be wise for teachers to use both concept maps and short answer tests while monitoring and promoting their pupils' learning.

1 Introduction

Concept maps have been argued to be a measure of cognitive structure (Novak & Gowin 1984, 138; Novak 1998, 52; Stoddart & al. 2000, 1223; Thompson & Mintzes 2002). An important problem is how much of pupils' knowledge and its structure we are able to represent by concept maps. Edwards and Fraser (1983, 24) provide empirical evidence for their claim "… concept maps were as accurate as interviews for revealing student comprehension of concepts." Novak & Cañas (2008) argue in similar fashion referring to Edwards and Fraser (1983). As far as we know no one has tried to use short-answer tests to probe pupils' understanding about the same content area as they made concept maps. In this paper we present a quasi-experimental design and results of statistical tests concerning this issue.

According to Åhlberg (1990) all concepts become accurate only in relation to other concepts, as a part of a developing conceptual structure. That is why it is important to monitor and promote development of pupils' conceptual structures.

2 Methods

From all we knew about concept maps, we created the research problem and planned the methods by which these problems would be answered. Research problem: Are there statistically significant differences between pupils' own concept maps and their answers to the short answer questions in a teacher-made achievement test as indicators of pupils' quality of learning and understanding? The following research methods were used to answer research problems: (1) Counting personal sums of concepts and propositions from the last student-made concept maps; (2) Counting sums of concepts and propositions from pupils' answers to a short-answer test. (3) Using a teacher-made concept map as a tool; (4) Paired-samples t-test and (5) Eta-squared –coefficients used as measures of effect-size.

2.1 Research design

The research and teaching are designed to test practicality of concept mapping and teacher-made short-answer achievement tests. Design experiments (Brown 1992, Brown and Campione 1996, Collins 1996) were planned and executed in a school class in Eastern Finland. There were two series of design experiments: the first one from 1997 to 2000 with one class (group of pupils) and from 2000 to 2002 with another class. In the first series of design experiments (1997–2000) there were 19 pupils, 11 girls and 8 boys in the class. In the second series of design experiments (2000–2003) there were 20 pupils, 10 girls and 10 boys in the class. When we examined the learning in these long series of design experiments, we have used sums of relevant concepts and sums of relevant propositions as the best available indicators of the quality of learning. The more relevant concepts and relevant propositions a pupil has, the more likely s/he is able to learn more and more meaningfully.

In this paper we present only the results of four design experiments in which teacher-made short-answer achievement tests were used as a control to pupils' concept maps. In two of the design experiments teacher-made short-answer achievement test was presented after the last pupil-made concept maps, in two design experiment teacher-made short-answer achievement test was presented before the last pupil-made concept maps. The total research design of four design experiments was balanced as regards to timing of concept mapping and short answer testing. The design experiments were as follows:

1) The Atmosphere-design experiment: The first concept maps were created on March 26, 2001, the second concept maps were created on April 19, 2001. The teacher-made short-answer test was on May 3, 2001.
2) The Human Biology-design experiment: The first concept maps were constructed on October 19, 2001, the second concept maps on January 17, 2002. The teacher-made short-answer test was on January 17, 2002.
3) The Australia-design experiment: The first concept maps were constructed on January 17, 2003. The teacher-made, short-answer test was on February 19, 2003. The second concept maps were made on February 21, 2003.
4) The Sun and Planets-design experiment: The first concept maps were made on April 4, 2003. The teacher-made, short-answer test was on April 24, 2003. The second concept maps were done on April 28, 2003.

The answers to the teacher made short-answer test of the nine intensively studied pupils were transformed into concept maps. Sums of relevant concepts and sums of relevant propositions were calculated.

2.2 Generalizations in quantitative and qualitative research: Importance of replications

Our research data are not random samples, but purposeful samples, which are information rich (applying Patton 1990, 181–185 and Patton 2002, 242). Statistical tests are used, but not for statistically generalizing to any accurate population. In educational sciences there is the problem that populations are constantly changing. This means that there is no real way to take a random sample of pupils or teachers in the strict statistical sense. Strictly speaking, there is no statistically sound way of statistical generalization to any larger population. But theoretical generalization is possible, because the researched pupils and their teacher are cases of real pupils and real teachers respectively (Cook, Leviton & Shadish 1985, 763 – 764; Yin 1998; de Vaus 2002, 148). Purposeful samples of real pupils allow us to conclude that under the similar conditions, similar phenomena are likely to happen.

In this article statistical significance tests are applied to test how much the figures from the samples differ from what would be expected by chance. This is a test of whether variation in the sample statistically significantly differs from random variation. Testing four times, allow us further check whether the statistical analysis results of purposeful samples the tentative regularities stand up to the repeated statistical testing. In behavioural and educational sciences there are far too few replications (Cook & Campbell 1979, 79–80; Rosenthal & Rosnow 1984, 181–191; Robinson & Levin 1997; Thompson 1997).

2.3 Data Analyses and Findings

Principal Vuokko Ahoranta, MSc collected the data. She has used concept maps and Vee heuristics regularly with her pupils since 1997, about four to five times per school year. She had started to work with these pupils in the same school year when the data was collected. All pupils were involved and very attentive during their part in the design experiments. They knew and were pleased that they collaborated in educational research as a part of their teacher's university studies. The validity of concept maps was checked using pupil interviews from time to time. In the four design-experiments, pupil's answers to teacher-made short-answer test provided an extra validity check of pupils' concept maps.

Data was analyzed both qualitatively and quantitatively (Miles and Huberman 1994; Miles & Huberman 2002; Cresswell 2002, 559-601). Following Åhlberg & Ahoranta (2004), we assume that meaningful learning has at least two practical indicators: personal sums of relevant concepts and of relevant propositions. Variation of meaningful learning is indicated by variation of sums of relevant concepts and sums of relevant propositions, both of which are calculated from concept maps. They are the best available indicators of meaningful learning and understanding. These figures were statistically analyzed.

We present two examples of a pupil's concept maps and an example of a teacher-constructed concept map based on a pupil's answers in the teacher-made, short-answer test.

Atmosphere-design experiment

Pupils #208 constructed himself the following concept maps: Figures 1 and 2. We have added short interpretations of the maps.

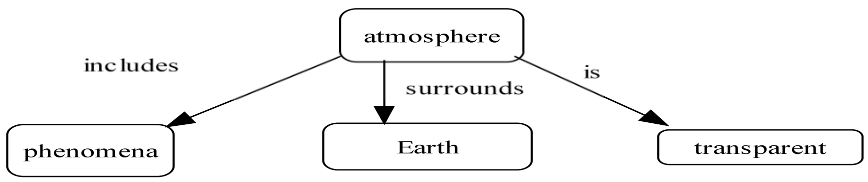

Figure 1. The first concept map of the pupil #208 at the beginning of the design experiment on Atmosphere.

Interpretation of Figure 1: There is an appropriate hierarchy in the first concept map of this design experiment. There are four relevant concepts and three relevant propositions. There are no clear misconceptions.

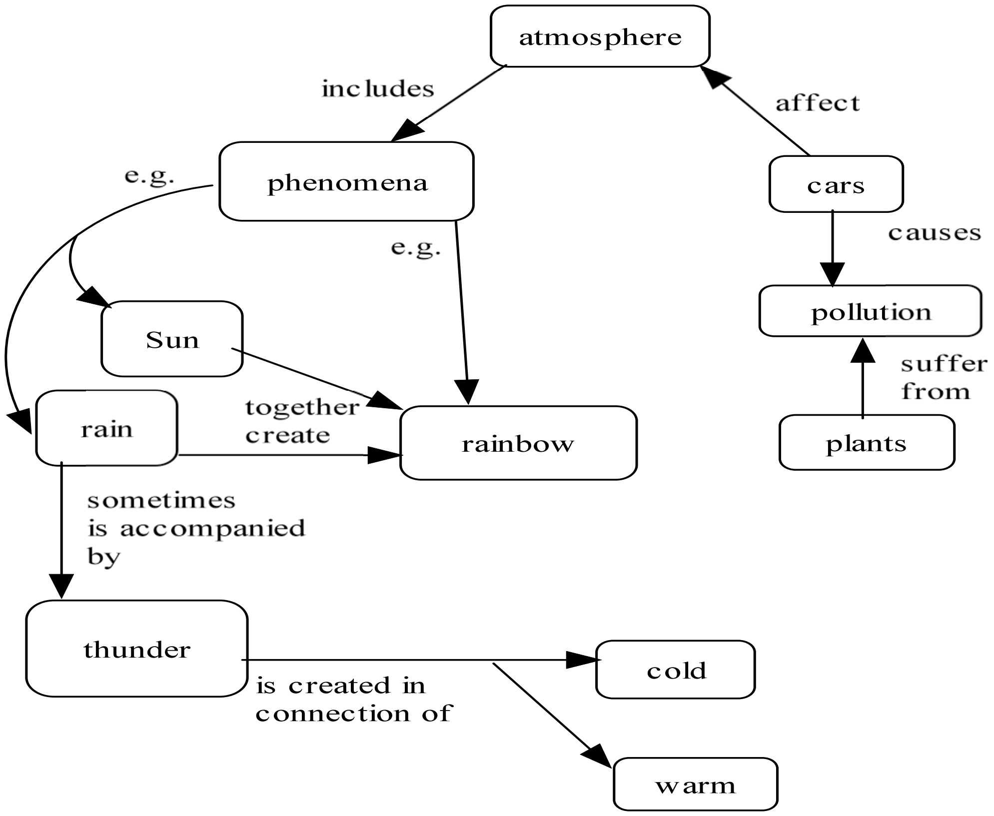

Figure 2. The last concept map of the pupil #208 at the end of the design experiment on Atmosphere.

Interpretation of Figure 2: There is an appropriate hierarchy in the concept map. There are 11 relevant concepts and 12 relevant propositions. There are no clear misconceptions. Based on the number on links to and from each concept, we find three equally central concepts in this concept map: ‘phenomena’, ‘rainbow’ and ‘thunder’. Each of them is connected to other concepts by three links. Probably these concepts were the most prominent in the pupil’s thinking when he constructed his concept map.

Atmosphere: Transforming a pupil’s writing (short answers) to a concept map

The teacher-made short-answer questions/tasks were as follows: 1) What gases the air is made of? 2) Explain how the wind is formed; 3) What do you know about air pressure? 4) In which ways the atmosphere is important for humankind? 5) How is a rainbow created? 6) How are the Northern lights created? 7) What are the uses of air for humans? 8) What causes air pollution and how it can be prevented? 9) How is a thunderstorm created? 10) How is burning linked to the earlier questions and their answers?

The content validity of the test is high, because these short questions correspond with the local curriculum, its main themes and concepts. We transformed the pupils' answers to a concept map (Figure 3).

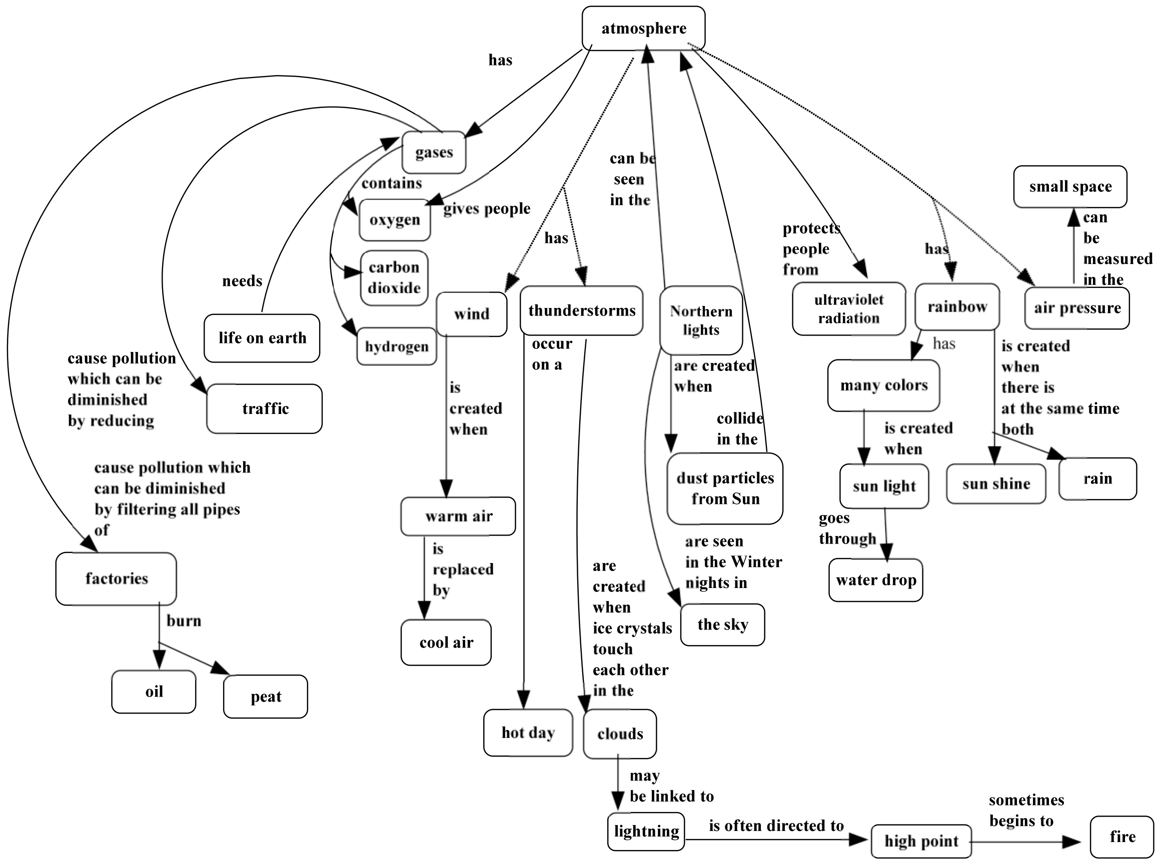

Figure 3. An example of a concept map which the teacher constructed transforming proposition by proposition pupil #208's answers in the teacher-made short-answer test on 'Atmosphere'.

Interpretation of Figure 3: When a pupil herself/himself constructs a concept map s/he can only take into the concept map what is in her/his metacognition. However, all pupils were able to provide much more information when explicitly asked in the teacher-made short-answer test. These two methods to gather knowledge about what and how pupils learn and think are complementary. Both provide useful knowledge for both pupil and teachers. It is interesting to learn that the most central concept of this pupil's thinking was 'gases'. It has seven links with other concepts. 'Atmosphere' has nine links with other concepts, but four of them are inferred implicit links.

2.4 Effect sizes

The fifth edition of APA (2001, 25-26) Publication manual emphasizes the importance of reporting effect sizes, e.g., product moment r-squared, and eta-squared (or effect size r-squared). Kier (1999, 95–96), Kramer & Rosenthal (1999, 63), Rosenthal and DiMatteo (2001, 72) and Field and Hole (2003, 166, 170, 180) present how eta-coefficient or "the effect size r" can be calculated from t-values and sums of squares resulting from Student's t-test and F-values and sums of squares of ANOVA.

2.5 Evaluating the quality of concept maps

There are many methods for evaluating the reliability and validity of concept maps (e.g., Ruiz-Primo and Shavelson 1996; Rice, Ryan, Samson 1998; McLure, Donak and Suen 1999; Ruiz-Primo, Schultz, Li and Shavelson 2001). These tests are based on Ausubelian learning theory as Novak and Gowin (1984) applied it to concept maps. Åhlberg and Ahoranta (2004) have presented a more general approach based on the idea that concepts and propositions are basic units of thinking, and in evaluation the main task is to check how many concepts and propositions are relevant.

Cronbach's alphas based on the total sum of number of concepts and propositions in the beginning, middle and in the end of the design experiments in the four intensively studied design experiments described in this report are:

- Atmosphere, based on 1st cmap, 2nd cmap, cmap based on teacher-made achievement test: alpha = 0.95
- Human biology, based on 1st cmap, 2nd cmap, cmap based on teacher-made achievement test: alpha = 0.92
- Australia, based on 1st cmap, cmap based on teacher-made achievement test, 2nd concept map: alpha = 0.88
- Sun and planets: based on 1st cmap, cmap based on teacher-made achievement test, 2nd cmap: alpha = 0.73.

3 Results and discussion

Are there statistically significant differences between pupils' own concept maps and their answers to the short answer questions in a teacher-made achievement test as indicators of pupils' quality of learning and understanding?

A paired-samples t-test was used to find out whether there is a statistically significant difference between pupils' own last concept maps and teacher-constructed concept maps based on pupils' answers in the short-answer test. The results of these tests make it possible to calculate, how much variation of meaningful learning (as indicated by the number of relevant concepts and relevant propositions) does a form of knowledge representation (self-made concept maps vs. teacher-made concept map from pupils' answers to short-answer test) statistically explain.

Atmosphere–design experiment

Data are presented in Table 1.

			Sum of relevant concepts			Sum of relevant propositions		
subjects	School achievement level	Sex	The 1st concept map	The 2nd concept map	Short-answer-test	The 1st concept map	The 2nd concept map	Short-answer-test
201	Advanced	male	6	17	36	5	17	40
202	advanced	female	8	21	48	7	20	53
203	advanced	female	11	24	50	10	23	61
204	average	female	5	24	29	4	23	31
205	average	male	2	11	21	1	10	22
206	average	female	9	15	35	8	14	39
207	low	male	7	8	32	6	7	33
208	low	male	4	11	31	3	10	33
209	low	female	3	5	27	2	4	28

Table 1: The data of the design experiment of 'Atmosphere'.

The statistical results are presented in Table 2. There are statistically significant differences both in the number of relevant concepts and propositions between the last pupil-constructed concept map and the teacher-constructed concept map of pupils' answers to the short-answer test. The mode of pupils' knowledge presentation explains statistically almost 90 per cent of the variation of personal sums of both concepts and propositions.

Variables, which are explained statistically by the form of knowledge representation	Arithmetic means (M), standard deviations (s) and correlation between measurements (r_{12})	t = t-test value, df = degrees of freedom, p = statistical probability (2 – tailed)	Eta squared = effect size correlation between the second concept map and teacher-made short-answer test, based on personal sums of relevant concept/propositions .	percentage of variation explained
Personal sums of relevant **concepts** in the second concept map and the concept map based on short-answer test	The sample of pupils M_1 = 15.11, s1 = 6.92 M_2 = 34.33, s2 = 9.43 r_{12} = 0.641, df = 7, p = 0.063	t = -7.907, df = 8, p = 0.000	0.89	89%
Personal sums of relevant **propositions** in the second concept map and the concept map based on short-answer test	The sample of nine pupils M_1 = 14.22, s1 = 6.96 M_2 = 37.78, s2 = 12.32 r_{12} = 0.670, df = 7, p = 0.048	t = -7.65, df = 8, p = 0.000	0.88	88%

Table 2: The statistical results of the design experiment of 'Atmosphere'.

Human Biology–design experiment

There are statistically significant differences both in the number of relevant concepts (t = -6.433, df = 8, p = 0.000, Eta squared = 0.84) and propositions (t = -6.228, df = 8, p = 0.000, Eta squared = 0.83) between the last pupil-constructed concept map and the teacher-constructed concept map of pupils' answers to the short-answer test. The mode of pupils' knowledge presentation explains statistically over 80 per cent of the variation of personal sums of both concepts and propositions.

Australia–design experiment

There are statistically significant differences both in the number of relevant concepts (t = -2.573, df = 8, p = 0.033, Eta squared = 0.45) and propositions (t = -2.705 df = 8, p = 0.027, Eta squared = 0.48) between the last pupil-constructed concept map and the teacher-constructed concept map of pupils' answers to the short-answer test. The mode of pupils' knowledge presentation explains statistically over 45 per cent of the variation of personal sums of both concepts and propositions.

Sun and Planets–design experiment

There are statistically significant differences both in the number of relevant concepts (t = -5.244, df = 8, p = 0.001, Eta squared = 0.77) and propositions (t = -5.146, df = 8, p = 0.001, Eta squared = 0.77) between the last pupil-constructed concept map and the teacher-constructed concept map of pupils' answers to the short-answer test. The mode of pupils' knowledge presentation explains statistically 77 per cent of the variation of personal sums of both concepts and propositions.

4 General Discussion

All pupils were involved and very attentive during their part in the design experiments. They knew and were pleased that they were used in educational research as a part of their teacher's university studies. This may have resulted in the Pygmalion effect or the Rosenthal effect that motivates the pupils to be more effective than usually. When anything new is introduced into a classroom, many kinds of unintended side effects may result (Ball 1988, 490). The Pygmalion side effect was used in this study in a positive way. We expected that all pupils would learn more and better while using concept maps and Vee heuristics, because both concept maps and Vee heuristics are designed and tested metacognitive tools to promote meaningful learning. Pupils' concept maps reveal to pupils themselves and to their teacher much more detailed knowledge than they earlier imagined possible to know about their learning.

In four design experiments pupil's answers to a teacher-made short-answer test provided an extra check on the corresponding pupil's concept maps. The authors were surprised to learn from pupils' answers to the short-answer test, that pupils knew much more than they presented in their concept maps. However, what was presented in the pupils' concept maps was often, in the answers to short-answer tests, but not always all. These two ways to probe pupils' learning and understanding are complementary. Principal Vuokko Ahoranta has continually discussed concept maps with her pupils, and her conclusion is that when pupils construct their concept maps they put into them everything relevant that comes into their minds. Improved concept maps then provide pupils and their teacher knowledge about pupils' metacognition. When remembering and knowledge construction is facilitated by short questions in the test situation, much more comes into mind. Our results differ from Edward's and Fraser's (1983, 24) conclusions after using pupil interviews: "… concept maps were as accurate as interviews for revealing student comprehension of concepts." These results need independent checking in other contexts. There is a need for replications in other classrooms, with larger populations of pupils and teachers.

References

Acton, W., Johnson, P. and Goldsmith, T. (1994). Structural knowledge assessment: Comparison of referent structures. Journal of Educational Psychology 86(2), 303–311.

Åhlberg, M. (1990). Käsitekarttatekniikka ja muut vastaavat graafiset tiedonesittämistekniikat opettajan ja oppilaiden työvälineinä. [Concept mapping and other graphical knowledge representation methods as tools for teachers and pupils.] University of Joensuu. Research Reports of the Faculty of Education N:o 30.

Åhlberg, M. & Ahoranta, V. (2004). Six years of design experiments using concept mapping – At the beginning and at the end of each 23 learning projects. In Cañas, A. J., Novak, J. D., Gonzales, F. M. (Eds.) Concept Maps: Theory, Methodology, Technology. Proceedings of the First International Conference on Concept mapping. CMC 2004 (pp. 45–51). Pamplona, Spain, Sept 14 – 17, Vol. 1., http://cmc.ihmc.us/papers/cmc2004-220.pdf

APA. (2001). Publication manual of the American Psychological Society. 5th edition. Washington, D.C.: American Psychological Society.

Ball, S. (1988). Unintended effects in educational research. In Keeves, J. (Ed.) Educational research, methodology, and measurement. An international handbook (pp. 490–493). New York: Pergamon Press.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. The Journal of the Learning Sciences 2(2), 141–178.

Brown, A. & Campione, J. (1996). Psychological theory and the design of innovative learning environments. On procedures, principles and systems. In Schauble, L & Glaser, R. (Eds.) Innovations in learning. New environments for education (pp. 289–325). Mahwaw, N. J:: Lawrence Erlbaum.

Collins, A. (19969. Design issues for learning environments. In Schauble, L & Glaser, R. (Eds.) Innovations in learning. New environments for education (pp. 347-361). Mahwaw, N. J: Lawrence Erlbaum.

Cook, T. & Campbell, D. (1979). Quasi experimentation. Design and analysis for field settings. Boston: Houghton Mifflin.

Cook, T., Leviton, L. & Shadish, W. (1985). Program evaluation. In Lindsey, G. & Aronson, E. (Eds.) Handbook of Social Psychology (pp. 699–777). New York: Random House.

Cresswell, J.(2002). Educational research. Planning, conducting, and evaluating quantitative and qualitative research. Upper Side River, N.J.: Pearson Education.

de Vaus, D. (2002). Analyzing social science data. London: SAGE.

Edwards, J. & Fraser, K. (1983). Concept maps as reflectors of conceptual understanding. Research in Science Education 13, 19 – 26.

Field, A. (2000). Discovering statistics using SPSS for windows: advanced techniques for the beginner. London: SAGE.

Field, A. & Hole, G. (2003). How to design and report experiments. London: SAGE.

Kier, F.(1999). Effect size measures. What they are and how to compute them. In Thompson, B. (Ed.) Advances in social science methodology (vol. 5) (pp. 87–100). Stamford, CT: JAI Press.

Kramer, S. & Rosenthal, R. (1999). Effect sizes and significance levels in small sample research. In Hoyle, R. (Ed.) Statistical strategies for small sample research (pp. 59–79). London: SAGE.

Markham, K., Mintzes, J. & Jones, G. (1994). The concept maps as a research and evaluation tool: Further evidence of validity. Journal of Research in Science Education 31(1) 91–101.

Martin, B., Mintzes, J. & Clavido, I. (2000). Restructuring knowledge in biology: cognitive processes and metacognitive reflections. International Journal of Science Education 22(3), 303–323.

McLure, J., Sonak, B. & Suen, H. (1999). Concept map assessment of classroom learning: Reliability, validity and logistical practicality. Journal of Research in Science Teaching 36(4), 475–492.

Miles, M. & Huberman, M. (1994). Qualitative data analysis. Second edition. London: SAGE.

Miles, M., & Huberman, M. (2002). Reflections and advice. In Huberman, M. & Miles, M. (Eds.) The qualitative researcher's companion (pp. 393–397). London: SAGE.

Mintzes, J., Wandersee, J. & Novak, J. (1997). Meaningful learning in science: The human constructivist perspective. In Phye, G. (Ed.) Handbook of academic learning. Construction of knowledge (pp. 406–447). New York: Academic Press.

Mintzes, J., Wandersee, J. & Novak, J. (2001). Assessing understanding in biology. Journal of Biological Education 35(3), 118-124.

Nicoll, G., Fransisco, J. & Nakhleh, M. (2001). A three-tier system for assessing concept map links: a methodological study. International Journal of Science Education 23(8), 863–875.

Novak, J. (1990). Concept Maps and Vee Diagrams: Two Metacognitive Tools to Facilitate Meaningful Learning. Instructional Science 19(1), 29–52.

Novak, J. (1998). Learning, creating and using knowledge. Concept Maps™ as facilitative tools in schools and in corporations. London: Lawrence Erlbaum.

Novak, J. (2002). Meaningful learning: The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. Science Education 86(4), 548–571.

Novak J. D & Cañas, A. J. (2008) The Theory Underlying Concept Maps and How to Construct and Use Them. Institute for Human and Machine Cognition. Pensacola Fl, 32502, Retrieved January 26, 2008 from http://cmap.ihmc.us/Publications/ResearchPapers/TheoryCmaps/TheoryUnderlyingConceptMaps.htm

Novak, J. & Gowin, B. (1984). Learning how to learn. Cambridge: Cambridge University Press.

Patton, Q. (1990). Qualitative evaluation and research methods. Second edition. London: SAGE.

Patton, Q. (2002). Qualitative evaluation and research methods. Third edition. London: SAGE

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

Robinson, D. & Levin, J. (1997). Reflections on statistical and substantive significance, with a slice of replication. Educational Researcher 26(5), 21–26.

Rosenthal, R. & Di Matteo, M. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. Annual Review of Psychology 52, 59–82.

Rosenthal, R. & Rosnow, R. (1984). Essentials of behavioral research. Methods and data analysis. New York: McGraw-Hill.

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

Stoddart, T., Abrams, R., Gasper, E. & Canaday, D. (2000). Concept maps as assessment in science inquiry learning – a report of methodology. International Journal of Science Education 22(12), 1221–1246.

Thompson, B. (1997). Editorial policies regarding statistical significance tests: Further comments. Educational Researcher 26(5), 29–32.

Thompson, T & Mintzes, J. (2002). Cognitive structure and the affective domain: on knowing and feeling in biology. International Journal of Science Education 24(6), 645–660.

Yin, R. (1998). The abridged version of case study research. Design and method. In Bickman, L. & Rog, D. (Eds.) Handbook of Applied Social Research Methods (pp. 229–259). London: SAGE.