Third Misconceptions Seminar Proceedings (1993)

Paper Title: **The Validity and Reliability of Concept Mapping as an Alternative Science Assessment**

Author: Liu, Xiufeng & Hinchey, Michael

Abstract: Concept mapping as an evaluation tool has been primarily limited to assessing students' conceptual change. To explore the possibility of using concept mapping as an alternative science assessment, questions regarding validity and reliability have to be answered first. This on-going study is being done in two grade 7 science classrooms. After finishing a unit, students are given both conventional end-of-unit tests consisting mainly of multiple choice items and concept mapping tests. Although the results reported in this paper are inconclusive and incomplete, preliminary results have shown that concept mapping yields moderately high construct validity, consequential validity and internal consistency reliability.

Keywords: Concept Mapping,Construct Validity,Test Reliability,Misconceptions,Individual Testing,,,,
General School Subject: Chemistry
Specific School Subject: Inorganic Chemistry
Students: Junior High

Macintosh File Name: Liu - Concept Mapping
Release Date: 5-4-1994 G, 11-9-1994 I

Publisher: Misconceptions Trust
Publisher Location: Ithaca, NY
Volume Name: The Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics
Publication Year: 1993
Conference Date: August 1-4, 1993
Contact Information (correct as of 12-23-2010):
Web: www.mlrg.org
Email: info@mlrg.org

A Correct Reference Format: Author, Paper Title in The Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics, Misconceptions Trust: Ithaca, NY (1993).

great deal of effort has been invested in bringing you this proceedings, on the part of the many authors and conference organizers. The original publication of this proceedings was supported by a grant from the National Science Foundation, and the transformation of this collection into a modern format was supported by the Novak-Golton Fund, which is administered by the Department of Education at Cornell University. If you have found this collection to be of value in your work, consider supporting our ability to support you by purchasing a subscription to the collection or joining the Meaningful Learning Research Group.
----- -----

The Validity and Reliability of Concept Mapping as an Alternative Science Assessment

Xiufeng Liu, St. Francis Xavier University, Canada
Michael Hinchey, St. Andrew's Junior High School, Canada

The positive effect of concept mapping as an instructional tool for learning has been reported in many school subjects, such as earth science (Ault, 1985); biology (Lehman, Carter, and Kahle, 1985; Okebukola, 1990; Stewart ,Van Kirk and Rowell, 1985); chemistry (Stensvold and Wilson, 1990); physics (Hegarty-Hazel, 1991); and for various levels from kindergartens to colleges (Mason, 1992; Novak and Musonda, 1991). A meta-analysis on the efficacy of using concept mapping as an instructional tool based on 19 studies showed that concept mapping had medium positive effects on students' achievement and attitude (Horton, McConney, Gallo, Woods, Senn and Hamelin, 1993). Concept mapping as an evaluation tool was proposed by Novak and Gowin (1984).They argued that conventional achievement assessment mainly consisting of multiple-choice items did not measure what students learned. They discussed examples of poor prediction validity, construct validity and content validity of conventional tests. More recently, empirical results of students' inconsistent performances in standardized tests were reported and the inherent problem with standardized tests from the point view of constructivist approach to science teaching was discussed by Liu (in press). Therefore, a need for an alternative assessment was evident. According to Novak and Gowin (1984), concept mapping as an assessment tool integrated learning and assessment, which were traditionally considered to be separate. It was argued that valid assessment should be conducted in terms of how students learn. Concept mapping is consistent with students' general learning process and, therefore, is promising for assessing students' learning (Novak & Gowin, 1984). Several studies have been reported using concept mapping as a diagnostic tool to assess students' conceptions (Moreira, 1985; Ross and Mundy, 1991; Wallace and Mintzes, 1990). However, the empirical findings regarding the validity and reliability of using concept mapping as an alternative achievement assessment have not been reported.

The current emerging consensus on validity inquiry agree on two aspects of validity: construct validity (including content validity and prediction validity), and consequential validity (Moss, 1992; Messick, 1989). The common approaches to reliability study include three examinations: test-retest reliability, split-half reliability, and internal consistency (Crocker and Algina, 1986). The present study is to provide preliminary evidence on the validity (construct validity and consequential validity), and reliability (internal consistency) of concept mapping as an alternative science assessment. Test-retest reliability was implied indirectly by Novak and Gowin (1984). They found that "...the first concept map a person makes is almost certain to have flaws. ...a second map usually shows key relationships more explicitly. Most students will not have patience or interest to try a third or fourth version of their maps for a topic,...." (p. 35). The split-half reliability is considered to be inappropriate to the situation of concept mapping as an alternative assessment, because a concept map does not have many items which can be split.

METHOD

The study reported in this paper is a part of an ongoing project conducted at a rural junior high school in a Canadian Atlantic province. Two grade 7 general science classes taught by the second author of this paper participated in the study. The two classes (38 students altogether) were initially formed randomly after the top 60 high achievement and bottom 30 low achievement students were assigned to enriched and adjusted classes. Therefore, the two classes of students can be considered to be equivalent, and the results for both the classes were pooled together for analysis. Before the first author introduced the idea of concept mapping to the teacher (the second author), the teacher was not quite aware of the potential of concept mapping for instruction and assessment, but enthusiastic to apply it. After having finished a unit, the teacher gave the classes a conventional test as before, and also a concept mapping as a test. The conventional tests took one class period (45 minutes) and concept mapping took one class period as well. The conventional test was given before the concept mapping because of the instructional effect of concept mapping as reviewed above. The data analyzed in this study were mainly from the tests on the first unit (solution) of the research project. After having finished the unit on solution, students were given a conventional test. The conventional test consisted of 6 parts. Part A included 20 multiple choice items; part B was a matching subtest; part C contained 10 fill-in-blank items; part D contained 10 true-false items; part E contained 10 judgmental items; and part F contained 4 short-answer items. The concept mapping test was given two days later than the conventional test. Because students had not been exposed to concept mapping before, the teacher spent the first 15 minutes of the class period explaining what concept mapping was using the example from Novak and Gowin (1984, p. 16). Students then were asked to draw concept maps to represent what they had learnt in the unit on solution. All the students returned their concept maps at the end of the period. The conventional test was scored by the teacher independently, and the concept maps were scored by the first author independently. The weight scheme for scoring concept maps was that proposed by Novak and Gowin (1984). Different weight schemes for scoring concept maps have been proposed (Cleare, 1983; Wallace and Mintzes, 1990), but they were basically variations of Novak's scoring scheme (Novak and Gowin, 1984). Students' scores on conventional and concept map tests from the first unit were used to investigate the construct validity in this study based on the consideration that concept mapping itself had great positive potential to improve learning as reviewed above; especially the effect of concept mapping to improve learning was confounded with students' characteristics, such as verbal ability (Stensvold and Wilson, 1990), internal/external locus of control, individual differences in learning process, spatial ability, and academic performance measures (Zeitz, Anderson-Inman, 1993).

RESULT

CONSTRUCT VALIDITY

Students' concept maps were initially pooled into a big concept map on a Bristol board. The frequency of each link (valid and invalid) demonstrated by the students were counted and the percentage of students demonstrating the link was marked on the linking line. Because the pooled concept map was very big, a simplified concept map was made by deleting all the single occurrence links demonstrated in only one student's map. The final integrated concept map is displayed in Figure 1. From Figure 1, we see that 97% of students indicated that solutions had two parts: solute and solvent, and 61% of students indicated that solutions had strengths. A certain percentage of students also mentioned saturation, the separation of a solution, solubility, ways to speed up dissolving, etc.

Insert Figure 1 about here

According to the textbook the students were using (Atlantic Science Curriculum Project, 1986), students were expected to understand and explain:

. How is a solution different from a non-solution?
. What is a solute? a solvent?
. What does the word "hard" in "hard water" mean?
. What makes water hard?
. What things are very soluble? not very soluble? insoluble?
. How can you separate a solute from a solvent?
. How does a solute affect the boiling and freezing points of a solvent?
. What is a "dilute" solution? a "concentrated" solution?
. How can you test a solution to see whether it is dilute or concentrated?
. What is meant by "saturation" and "solubility"
. Does temperature affect the amount that can be dissolved? (p.68)

It can be seen that the concepts demonstrated in students' concept maps were basically in agreement with what were expected by the textbook. A few concepts, such as hard water and the effect of solutes on the boiling and freezing points of a solvent, were indicated only by individuals, therefore were not shown in the pooled concept map. Those concepts mentioned by individuals reflected a difference of students in understanding those concepts. It has to be pointed

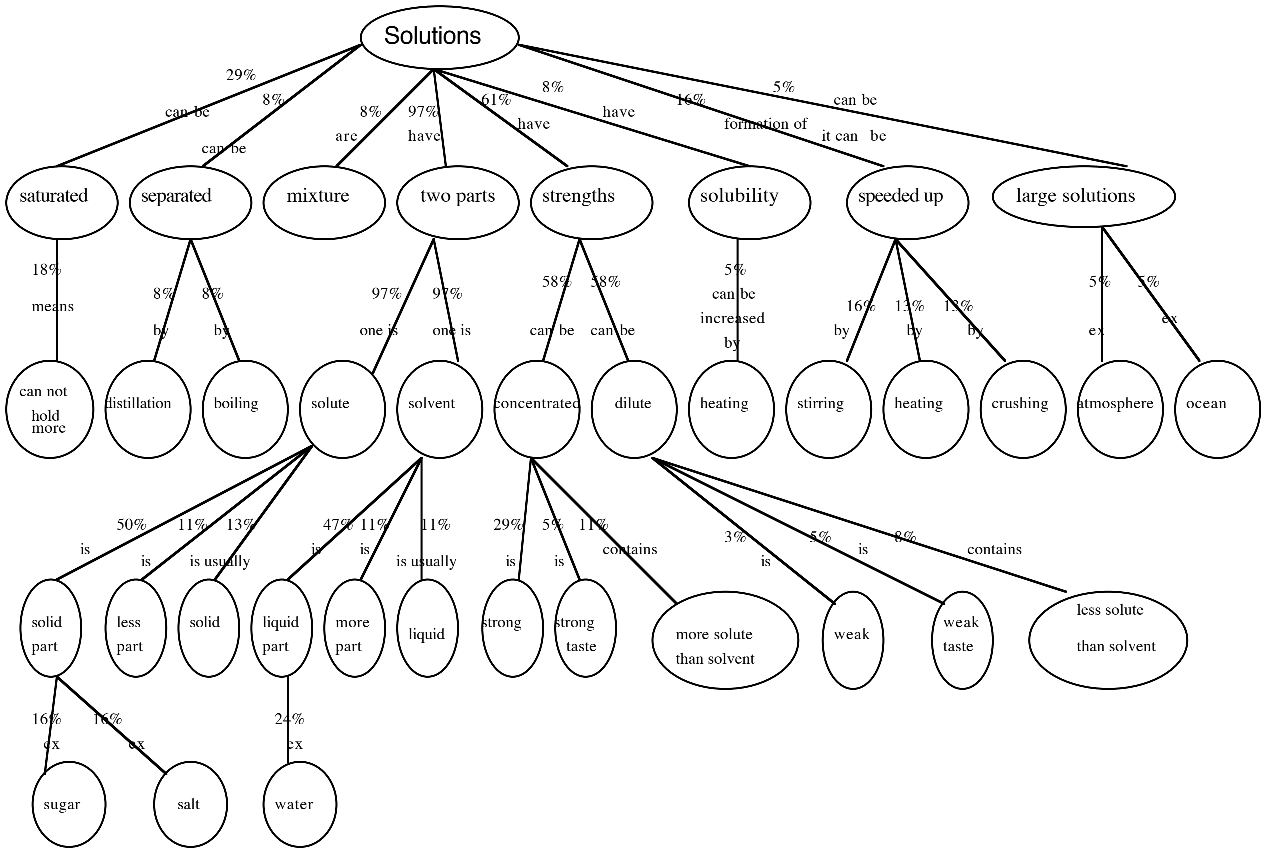

Figure 1. Pooled Concept Map

out that those individual links were enormous. It was found that the majority of students had at least one unique link. In summary, the concept maps produced by students reflected a certain degree of centrality on some concepts such as the components of a solution and variation on some other concepts such as hard water.

Table 1 displays students' scores on both the conventional and concept mapping tests. Students' conventional test scores ranged from 51.0 to 93.0 with a mean of 77.7 and a standard deviation of 10.1. Students' concept mapping scores ranged from 6 to 62 with a mean of 22.8 and a standard deviation of 11.7. The absolute range (the difference between the maximum score and minimum score) for conventional test scores was 42 and the absolute range for concept mapping test scores was 56. The standard deviations for both the conventional and concept mapping test scores were close (with 10.1 and 11.7 respectively). The Pearson's correlation between students' conventional and concept mapping scores was computed; it was .437, which was significant at the .05 significance level ($p<.05$). The Spearman's rank order correlation between the conventional test scores and concept mapping scores was also computed; it was .3115, which was not significant at the .05 significance level ($p=.057$). However, when excluding one extreme student who had a very high conventional test score but a very low concept mapping score (student 21), the Spearman's rank order correlation was .3777, which was significant at the .05 significance level ($p=.021$). Student 21's concept map is displayed in Figure 2. From Figure 2, there were many invalid links and hierarchies in the concept map. Many links were not connected by appropriate linking words. The student clearly demonstrated a logic confusion about the concepts.

Insert Table 1 about here

Insert Figure 2 about here

Figure 3 and Figure 4 show two typical students' concept maps. The student who produced the concept map in Figure 3 had the highest conventional test score (93), and the student who produced the concept map in Figure 4 had the lowest conventional test score (51). Comparing Figure 3 and Figure 4, we can see that the abler student demonstrated much more valid links and hierarchies than the less able student. The less able student showed many concepts in the map, but most of them were invalidly connected by linking words or put in wrong hierarchies.

Insert Figures 3 and 4 about here

Table 1 Students' Scores on the Conventional and Concept Mapping Tests

Student	Conventional Test Scores	Concept Mapping Scores
1	75	18
2	87	33
3	82	26
4	61	6
5	51	6
6	79	23
7	66	7
8	81	23
9	93	30
10	78	38
11	72	17
12	93	16
13	89	43
14	84	13
15	92	62
16	71	32
17	64	24
18	82	13
19	86	34
20	81	15
21	92	9
22	90	11
23	79	30
24	72	21
25	74	14
26	79	37
27	81	12
28	74	26
29	77	19
30	68	20
31	85	34
32	84	21
33	70	6
34	79	16
35	70	28
36	57	16
37	87	34
38	69	28

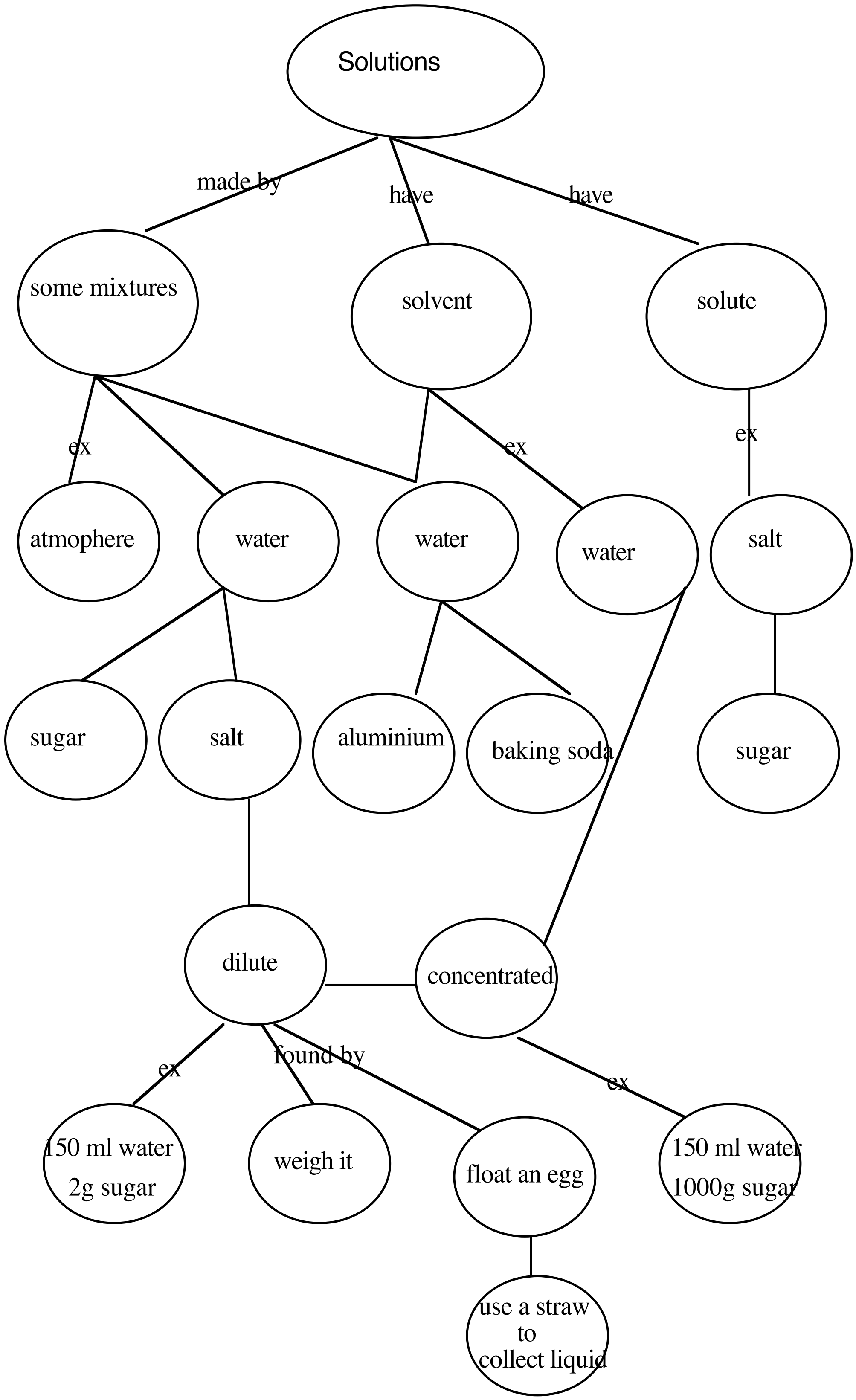

Figure 2. A Concept Map Made by the Student Who Had High Conventional Test Score but Low Concept Mapping Score

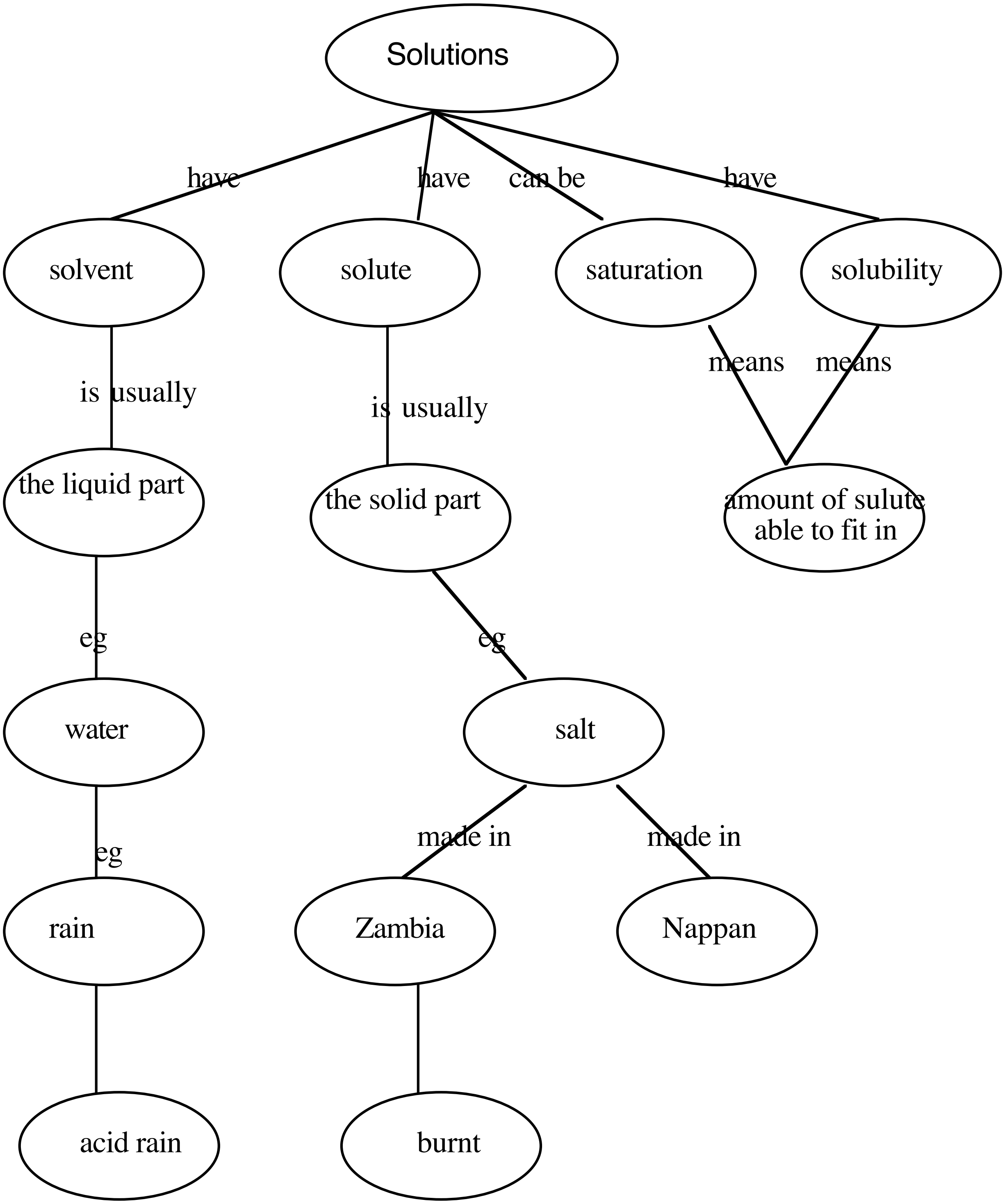
Solutions
have
have
can be
have
solvent
solute
saturation
solubility
is usually
is usually
means
means
the liquid part
the solid part
amount of sulute able to fit in
eg
eg
water
salt
eg
made in
made in
rain
Zambia
Nappan
acid rain
burnt

Figure 3. A Concept Map Made by an Able Student

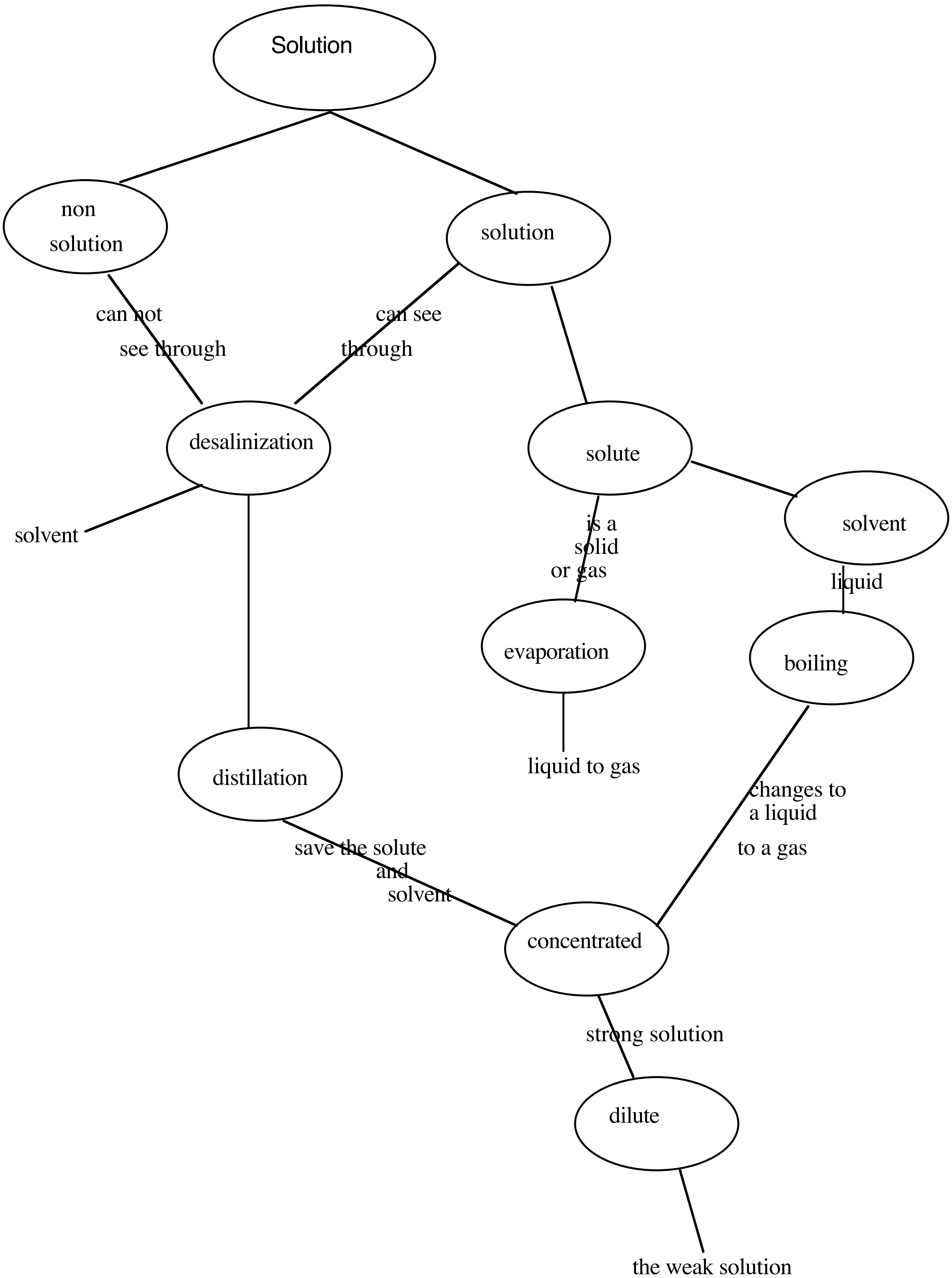

Figure 4. A Concept Map Made by a Less Able Student

CONSEQUENTIAL VALIDITY

The concept maps produced by the students had richer information on students' conceptions than the conventional test. In our study, we differentiated students' misconceptions and incorrect conceptions. Students' misconceptions were those indicated by partially valid links, hierarchies, cross-links and examples, and students' incorrect conceptions were those indicated by invalid links, hierarchies, cross-links, and examples. An example of students' misconceptions can be identified in Figure 1. It can be seen that 50% of students considered solute to be the solid part of a solution and 47% of students considered solvent to be the liquid part of a solution. It can also be seen that 5% of students considered that "dilute" meant "weak taste" and "concentrated" meant "strong taste". Students' incorrect understandings can also be identified in their concept maps. In Figure 1, 11% of students thought incorrectly that "concentrated" meant more solute than solvent. Figure 5 displays a concept map produced by a middle level achievement student indicated by the conventional test score. The student correctly demonstrated that solutions were made of solutes and solvents, but had the misconceptions that solute was the solid part and solvent was the liquid part of a solution. The student also incorrectly demonstrated that when a solution was evaporated, it could turn to gas such as air. Concept maps clearly showed where students had misconceptions and where students held incorrect understandings. Students' misconceptions and incorrect conceptions were also scored using the same weight scheme as scoring valid links, hierarchies, examples, and cross-links. It was found that students' misconception scores ranged from 0 to 31 with a mean of 7.9 and a standard deviation of 7.6. Students' incorrect conception scores ranged from 0 to 19 with a mean of 2.9 and a standard deviation of 4.1. It can be seen that students demonstrated more misconceptions than incorrect conceptions according to the means. Pearson correlations between students conventional test scores and their misconception scores, and between students conventional test scores and their incorrect conception scores were computed. The two correlations were -.301 and .061 respectively; neither of them was significant at the .05 significance level, indicating that the conventional test scores could not predict students' misconceptions nor incorrect conceptions. Students' misconception and incorrect conception scores could provide unique and valuable information for teachers on students' learning and on their instruction as well.

Insert Figure 5 about here

The teacher spent little time in preparation for using concept mapping as a test for the classes, compared to a considerable amount of time spent in designing the conventional test. In this sense, concept mapping would reduce teachers' time on preparing tests significantly. However, the time on scoring maps was found not to be superior to scoring conventional tests, which will be discussed later.

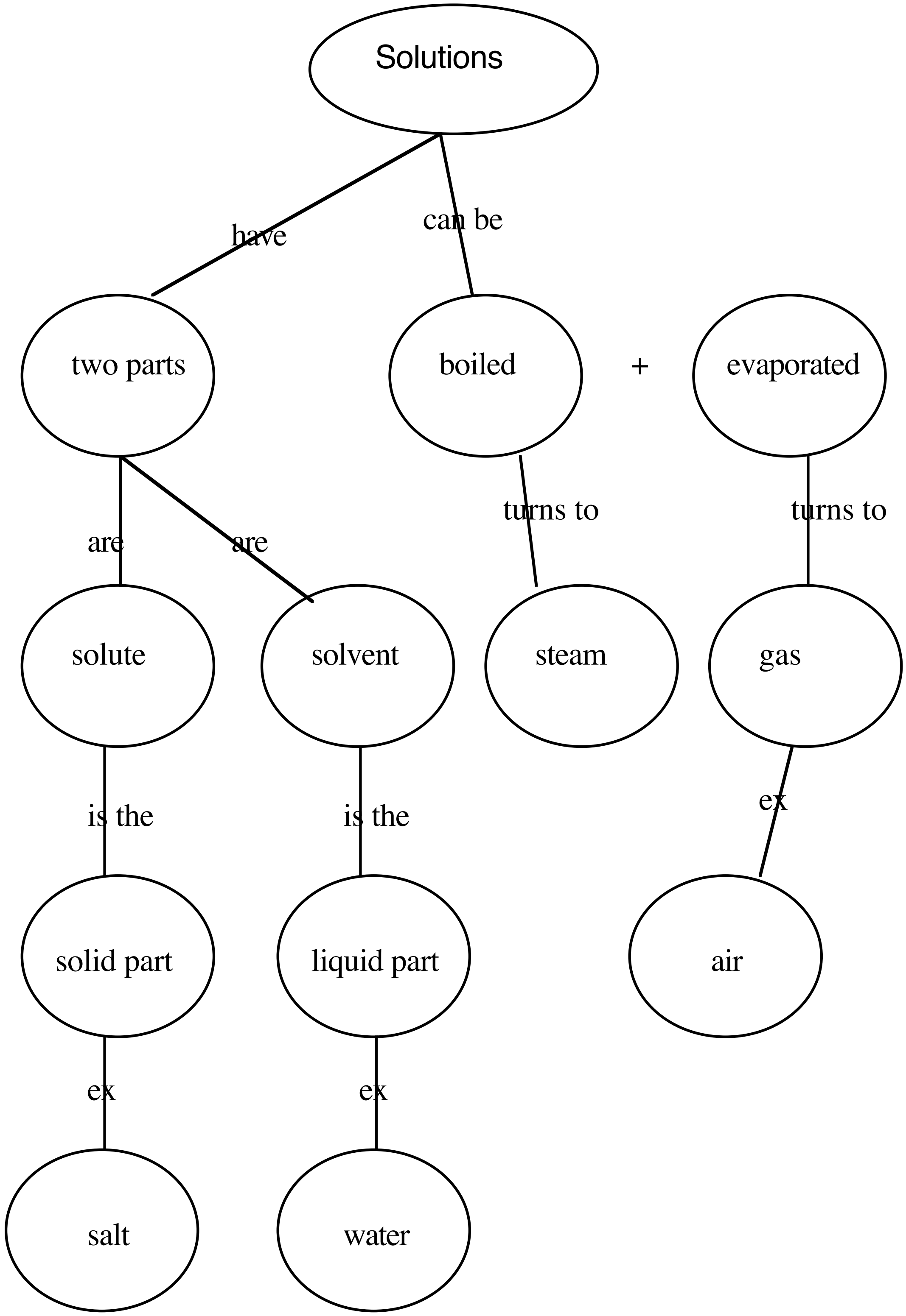

Figure 5. A Concept Map Made by an Ordinary Student

In addition to the free-style concept mapping discussed above, students were also given two other forms of concept mapping tests on other units. One was to ask students to construct concept maps according to a few questions provided and another one was to ask students to construct concept maps by filling in the blanks in a concept map frame with some concepts and linking words left blank. The last form of concept mapping was close to conventional tests in the sense that only certain types of answers were expected. The three forms of concept mapping constituted a continuum from a free-style concept mapping to a highly restricted concept mapping. Several representative (good, middle, and poor achievement) students were interviewed about their preference to the three forms of concept mapping. They all indicated that they preferred the first form of concept mapping (free-style concept mapping). With free-style concept mapping, they could show whatever they knew and were not afraid of making mistakes. With restricted concept mapping, especially with the third form, they felt confined and were afraid of making mistakes because they had to figure out what was expected by the teacher. Therefore, when the first form of concept mapping was used as an assessment, students' confidence was increased and their test anxiety was reduced.

INTERNAL CONSISTENCY RELIABILITY

Students' scores for valid links, hierarchies, cross-links, and examples in concept maps on solution are displayed in Table 2. Many students had zero cross-link scores, i.e. many students did not demonstrate cross-links in their concept maps. The reliability of concept mapping in terms of internal consistency can be defined in different ways. In this study, the internal consistency was defined in terms of Cronbach's alpha (Crocker and Algina, 1986), considering concept map scores are also composite scores consisting of scores on valid links, hierarchies, cross-links, and examples. The variance for total concept map scores was 141.13; the variance for valid link scores was 15.56; the variance for valid hierarchy scores was 40.20; the variance for valid cross-link scores was 12.87; and the variance for example scores was 4.131. Based on these variances, the Cronbach's alpha was computed to be .6495. This moderately high reliability can be found in many conventional tests as well, especially in classroom tests.

Insert Table 2 about here

DISCUSSION

The valid concept map scores were significantly correlated with students' conventional test scores. This result is consistent with previous findings. For example, Bousquet (1982) found that concept map scores could predict students' achievements in a college natural resources class. Fraser and Edwards (1985) found that students who scored at high levels on end-of-unit tests showed high levels of concept mastery as indicated by the concept maps they made. Compared to conventional tests, concept mapping can also provide information on students' misconceptions and incorrect conceptions, which are usually unavailable in conventional tests. In this sense, concept mapping as an alternative assessment has advantages.

The Spearman's rank order correlation was not significant when including the extreme student (who had very high conventional test score but very low concept map score). This indicates that concept mapping as an alternative assessment might be inappropriate to a certain type of students. The appropriateness of concept mapping as an alternative assessment might be confounded with students' characteristics. No studies have been reported on this effect, although a number of studies reported a confounding effect of concept mapping as an instructional tool with students' characteristics (Stensvold, 1990; Zeitz, et al., 1993).

Although the consequential validity is in favor of concept mapping in terms of the information provided, students' attitudes towards the concept mapping and the time the teacher spent on the preparation of concept mapping test, a difficulty in terms of scoring concept maps was found in this study. The difficulty was that scoring concept maps was much more time-consuming than scoring conventional tests. This difficulty can be significant if concept mapping is used in a large-scale achievement assessment. Two improvements may solve this problem. One way is to categorize students concept maps (links, hierarchy, cross-links, and examples) related to a topic (such as solution), and develop a key system for standardizing and speeding scoring. Another improvement is to develop a computer administered concept mapping test system. Typical concept maps (links, hierarchies, cross-links, and examples) are stored and students concept maps are compared to the typical maps for scoring. If we consider other forms of consequential validity as well, such as that concept mapping integrates learning and assessment into one process as argued by Novak and Gowin (1984), we believe that the positive consequential validity outweighs the negative consequential validity when using concept mapping as an alternative test.

Students demonstrated few cross-links in their concept maps. This result was also reported in Stensvold (1990). Although this may be due to students' inexperience with concept mapping, we believe it also reflects students' understanding differences, because a few students did show a number of cross-links and thus had higher concept mapping scores.

In this study, we investigated internal consistency reliability. The inter-rater reliability can become important if concept mapping is used in a large-scale alternative assessment. Inter-rater reliability in large scale performance assessment was discussed in Gipps (1993). Although it might be arbitrary to consider concept mapping to be a form of performance assessment, we tend to consider concept mapping to be in the middle between conventional achievement tests and current performance assessments. In large scale performance assessments, a major difficulty is the manageability (Gipps, 1993). This is not the case if concept mapping is used in a large scale

Table 2 Students' Concept Mapping Scores by Categories

Student	Link	Hierarchy	Cross-link	Example
1	6	10	0	2
2	11	15	0	7
3	7	15	0	4
4	1	10	0	0
5	1	5	0	0
6	8	15	0	0
7	2	5	0	0
8	8	15	0	0
9	6	20	0	4
10	6	20	10	2
11	3	2	0	4
12	2	10	0	4
13	13	25	0	5
14	3	10	0	0
15	11	25	20	6
16	15	15	0	2
17	7	15	0	2
18	3	10	0	0
19	16	15	0	3
20	7	5	0	3
21	4	10	0	2
22	3	5	0	3
23	8	20	0	2
24	2	15	0	4
25	7	5	0	2
26	13	20	0	4
27	8	0	0	4
28	11	15	0	0
29	9	10	0	0
30	10	10	0	0
31	10	20	0	4
32	9	10	0	2
33	5	0	0	1
34	6	10	0	0
35	5	20	0	3
36	6	10	0	0
37	14	20	0	0
38	7	15	0	6

assessment. In order to use concept mapping in large scale assessment, inter-rater reliability becomes crucial. This issue is also closely related with the consequential validity (scoring times). The standardized procedures and scoring systems have to be developed if concept mapping is used in large scale assessments.

REFERENCES

Atlantic Science Curriculum Project (1986). Science Plus (1). Harcourt Brace Jovanovich, Canada.

Ault, C. (1985). Concept mapping as a strategy in earth science. Journal of College Science Teaching, 15, 38-44.

Bousquet, W. S. (1982). An application of Ausubel's learning theory to environmental education: A study of concept mapping in a college natural resources management course. Unpublished doctoral dissertation: Ohio State University.

Cleare, Catherine C. (1983). Using concept mapping to detect intervention effective in improving pre-service elementary education majors' understanding of science topic. In Hugh Helm and Joseph Novak (Eds.) Proceedings of the International Seminar on Misconceptions in Science and Mathematics (vol. 1). Ithaca, NY: Cornell University. 272-81.

Crocker, Linda, Algina, James (1986). Introduction to classical and modern test theory. Holt, Rinehart and Winston, Inc.

Fraser, K., & Edwards, J. (1985). The effects of training in concept mapping on student achievement in traditional tests. Research in Science Education, 15, 158-165.

Gipps, Caroline V. (1993, April). Reliability, validity and manageability in large scale performance assessment. Paper presented at the annual meeting of the American Educational Research Association. Atlanta, GA.

Hegarty-Hazel, Elizabeth (1991). Relationship between students' conceptual knowledge and study strategies-part I: Student learning in physics. International Journal of Science Education, 13(3), 303-312.

Horton, Phillip B., McConney, Andrew A., Gallo, Michael, Woods, Amanda L, Senn, Gary J., & Hamelin, Denis (1993). An investigation of the effectiveness of concept mapping as an instructional tool. Science Education, 77(1). 99-111.

Lehman, J. D., Carter, C., & Kahle, J. B. (1985). Concept mapping, vee mapping, and achievement: Results of a field study with black high school students. Journal of Research in Science Teaching, 22(7), 663-673.

Liu, X. (in press). The inconsistency of students' performance in standardized testing. Journal of Research in Science Teaching.

Mason, C. L. (1992). Concept mapping: A tool to develop reflective science instruction. Science Education, 76(1), 51-63.

Messick, Samuel (1989). Validity. In Robert L. Linn (Ed.). Educational Measurement (3rd. edition). American Council on Education and Macmillan Publishing Company. 13-103.

Moreira, Marco (1985). An alternative strategy for evaluation. Assessment and Evaluation in Higher Education, 10(2), 159-68.

Moss, Pamela A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. Review of Educational Research, 62(3), 229-258.

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

Novak, J. D., Musonda, D. (1991). A twelve-year longitudinal learning. American Educational Research Journal, 28, 117-153.

Okebukola, P. A. (1990). Attaining meaningful learning of concepts in genetics and ecology: An examination of the potency of the concept-mapping technique. Journal of Research in Science Teaching, 27, 493-504.

Ross, Bertram, Mundy, Hugh (1991). Concept mapping and misconceptions: A study of high-school students' understandings of acids and bases. International Journal of Science Education, 13(1), 11-23.

Stensvold, Mark S., Wilson, John T. (1990). The interaction of verbal ability with concept mapping in learning from a chemistry laboratory activity. Science Education, 74(4), 473-480.

Stewart, J., Van Kirk, J., & Rowell, R. (1979). Concept maps: A tool for use in biology teaching. American Biology Teacher, 41(3), 171-175.

Wallace, Josephine D., Mintzes, Joel J. (1990). The concept map as a research tool: Exploring conceptual change in biology. Journal of Research in Science Teaching, 27(10), 1033-1052.

Zeitz, Leigh E., Anderson-Inman, Lynne. (1993, April). Computer-based concept mapping in a high school biology class: The effects of students' characteristics. Paper presented at the annual meeting of the American Educational Research Association. Atlanta, GA.