ACTA DIDACTICA NAPOCENSIA

Volume 10, Number 2, 2017

CONCEPT MAPPING AS A TOOL TO DEVELOP AND MEASURE STUDENTS' UNDERSTANDING IN SCIENCE

Sema Tan, Omer Erdimez, Robert Zimmerman

Abstract: Concept maps measured a student's understanding of the complexity of concepts, and interrelationships. Novak and Gowin (1984) claimed that the continuous use of concept maps increased the complexity and interconnectedness of students' understanding of relationships between concepts in a particular science domain. This study has two purposes; the first one was to test this claim and examine how the repeated use of concept maps affected the complexity and interconnectedness of concepts independent of science subjects in elementary school, the second one was to compare the sensitivity of the Ruiz-Primo et al. (1997), and the Novak and Gowin (1984) grading systems for concept maps. The sample group consisted of 23 students including 14 male and 9 female students. We employed paired sample t-tests to answer the research questions, and found that the scores obtained for the fifth science unit was significantly different from the first one. Also, Novak and Gowin's (1984) scoring system was better than Ruiz-Primo et al. (1997) to evaluate complexity in students' thinking except for one of the units. We conclude that concept maps have the potential to measure change in complexity and interconnectivity of concept maps. Furthermore, repeated use of concept maps has the potential to increase the complexity and interconnectedness of student concept maps, and therefore improve their understanding of science independent of science content.

Key words: concept maps, scoring, science, assessment

1. The Theoretical Framework

Concept Mapping as a Tool to Develop, and Measure Student's Understanding in Science

Concept maps measured students' understanding of the complexity of concepts, and interrelationships. They consist of concepts enclosed in circles or boxes and connected by a line to show the relationship between the two concepts and are visual images of the concepts and relationships (Novak & Cañas, 2006). The main goal of using concept maps has been to symbolize valid relationships between concepts in the form of propositions which are two concepts linked with a word(s), to develop a meaningful statement (Novak & Gowin, 1984). Several researchers found that concept mapping helped students internalize new crucial concepts, as well as integrate those concepts with previous knowledge, while revealing the students' level of knowledge and misconceptions (Bhattacharya & Han, 2001). Concept maps are constructive tools to help students (a) consider the connections between the science terms being learned, (b) organize their thoughts (c) visualize the relationships between key concepts in a logical way, and (d) reflect on their understanding (Ruiz-Primo, Shavelson, & Schultz,1997). Also, concept maps demonstrated students' understanding of interconnectedness and relationships between new concepts, along with the concepts to be learned (Novak & Gowin, 1984; Watson, Pelkey, Noyes, & Rodgers, 2016).

Researchers found that concept mapping helped students improve the performance on high cognitive level questions (BouJade & Attieh, 2008), increase the accuracy and complexity of the students' knowledge (Zimmerman, Maker, Gomez-Arizaga, & Pease, 2011), and have a more positive attitude toward learning science (Karakuyu, 2010). Some researchers found no significant differences in achievement between the students using concept maps and those using the traditional method

Received February 2017.

(Karakuyu, 2010), and in one study the mean of teacher candidates' concept map scores was considerably lower than scores of the achievement test (Ingec, 2009).

Scoring Concept Maps

Three types of scoring methods: traditional (Novak & Gowin, 1984), holistic (Besterfield-Sacre, Gerchak, Lyons, Shuman, & Wolfe, 2004), and categorical (Segalas, Ferrer-Balas, & Mulder, 2008) were examined and researchers found that traditional scoring was convenient for quick scoring, and holistic scoring was better at detecting the changes in knowledge structure. However, categorical scoring was the most reliable scoring system when the aim was to capture insight into content and students' knowledge structure (Watson, Pelkey, Noyes, & Rodgers, 2016). Austin and Shore (1993) recommended that teachers and researchers use one grading system rather than using multiple grading systems to score students' concept maps.

Research in which different scoring systems were studied has been conducted; however, none of the researchers compared the scoring systems quantitatively. For example, in their research, Stoddart, Abrams, Gasper, and Canaday (2000) included a table that contained the comparison of eight different scoring systems, but the comparison emphasized elaborateness and map components rather than the effectiveness of measuring science understanding. Ruiz-Primo and her colleagues focused mainly on using concept maps as assessment tools, and her research pertained to validity, reliability, and directedness of concept maps, and different scoring systems for concept maps (Ruiz-Primo & Shavelson, 1996; Ruiz-Primo, Shavelson, Li, & Schultz, 2001; Ruiz Primo, Schultz, Li, & Shavelson, 2001). The studies conducted to determine the validity of concept maps yielded correlational results ranging from 0.37 to 0.67 depending on the style of the concept maps (Liu & Hinchey, 1996, Ruiz Primo et al., 2001a).

Expert and Novice Problem Solving

Gifted learners have unique characteristics, and the main rule to nurture and enhance their learning is to use differentiated curriculum that meet the needs and unique characteristics of gifted students (Maker & Nielson, 1996). Differentiated curricula should include the use of higher level thinking skills and complex thinking processes to solve problems, and should help gifted students realize their potential to become experts. Experts' knowledge capacity and thinking styles are similar to those of gifted learners (Chichekian & Shore, 2014), therefore studying how experts differed from novices would help educators design curricula to promote a higher and more complex level of thinking.

Researchers found that experts'extensive knowledge affected what they noticed and how they organized, represented, and interpreted information in their environment, which in turn affected their abilities to remember, reason, and solve problems (Bransford, Brown, & Cocking, 2000, Dogusoy-Taylan & Cagiltay, 2014). Novices experienced learning through concept formation while experts learned through concept integration (Daley, 1999). Furthermore, expertise in a domain helped learners understand the patterns of meaningful information that were not available to novices, and experts' knowledge was organized around core concepts, which helped them to establish meaningful relationships between concepts (Bransford et al., 2000).

Clear differences existed between expert and novice thinkers in the ways of problem solving. Both expert and novice thinkers' schemata contained procedural knowledge; however, experts also thought about the applicability of procedural knowledge while novice thinkers' procedural knowledge lacked abstracted solution methods (Chi, Feltovich, & Glaser, 1981). Several studies found evidence that expert students presented higher level of understanding, quality, and complexity (Austin & Shore, 1993), obtained higher scores for their categorization and representation of information (Pinto, Doucet, & Ramos, 2010), and discovered significant high correlations between the multistep problem solving performance and linkage, score, and good links (Austin & Shore, 1993) compared to novices. Experts also followed a qualitative procedure by using key variables linked together (Heyworth, 1999), used their prior conceptual knowledge and experience during the problem-solving process (Hmelo-Silver, Nagarajan, & Day, 2002), while novice students applied any available formula into which given data were substituted (Heyworth, 1999). In summary, the expert characteristics required to create a concept map include, (1) applicability of procedural knowledge, (2) categorization of information, (3) the ability of using key variables linked together, and (4) the application of prior knowledge to the current problem solving process. Organizing meaningful information has been the basis of both expertise and concept maps. Expertise includes not only the knowledge of facts and formulas specific to the domain, but also the organization of these facts and formulas around core concepts or "big ideas" that guides experts' thinking about their domains more connectively (Bransford et al., 2000). Creating concept maps by organizing ideas around core concepts helps experts to retrieve information by spending less effort than novices do.

In one of the earliest studies about concept maps, Novak and Gowin (1984) claimed that the continuous use of concept maps increased the complexity and interconnectedness of students' understanding of relationships between concepts in a particular science domain, which was another characteristic of the knowledge base of experts. In our study, we tested Novak and Gowin (1984) claim by examining how using concept maps actually affected the complexity and interconnectedness of science concepts over time.

Studies about concept maps conducted in an elementary school setting are rare. This study would make a contribution to the understanding of how concept maps could be used in settings other than high school and college, and help educators understand how concept maps affected the complexity of students' thinking processes, they would also.

Purpose

The first purpose of the study was to examine how the repeated use of concept maps affected the complexity and interconnectedness of concepts independent of science subjects in elementary school. In addition to Novak and Gowin's (1984) claims, researchers (Austin & Shore, 1993; Chi et al., 1981; & Heyworth, 1999) found that experts' thinking skills were different from novices, and experts had a more complex understanding of concepts in their disciplines than did novices. Thus we aimed to test whether concept maps could be used as tools to develop complex thinking skills, and therefore develop expertise.

The second purpose of this study was to compare the sensitivity of the Ruiz-Primo et al. (1997), and the Novak and Gowin (1984) grading systems for concept maps. In these grading systems, different components have been considered. In the Ruiz Primo et al. (1997) grading system, propositions were given points from 1 to 4 based on the accuracy level, and cross-links had no value. In Novak and Gowin's (1984) grading system each proposition was worth 1 point regardless of its quality, but the crosslinks were graded based on the level of their quality. Therefore, choosing the more effective method would help educators save time and make more accurate decisions about achievement or placements based on students' understanding of the content they are teaching.

The following questions guided the study:

1. How did continuous use of concept maps affect the complexity and interconnectedness of students' concepts independent of science content?

2. Which of the two grading systems, Ruiz Primo et al. or Novak and Gowin, showed greater change in complexity of students' thinking from novice to expert?

2. Methodology

Method

A quantitative research design was used to investigate the increase in complexity of concept maps over time and science units, and to compare the effectiveness of the two grading systems, Ruiz Primo et al. (1997) and Novak and Gowin (1984), in measuring increases in complexity and number of connections among concepts. Students were asked to create concept maps before and after each Full Option Science System Unit.

Setting

The setting was an elementary school in a Southwestern city in the United States. The school was a small public neighborhood school near a university. At the time of this study, 343 students of varied ethnicities, languages, and nationalities were enrolled in classes from kindergarten to fifth grade.

Participants

The participants were selected from students who were involved in the study from the beginning of third grade until the end of fourth grade. The classroom teacher had more than 20 years of teaching experience. The sample group consisted of 23 students including 14 male and 9 female students. The students' ethnic backgrounds varied as follows: 10 White-American, 10 Hispanic, 2 Asian-American, and 1 student in two or more racial categories. Fourteen students were identified as gifted based on the Developing Cognitive Abilities Test (DCAT), a test that was designed to assess reasoning abilities in verbal, quantitative, and spatial areas for children in K-12, and Raven's Progressive Matrices, a nonverbal ability test in a spatial format to assess reasoning abilities. Students' socio-economic status was generally high, but 30% of the students in the sample were below poverty level.

Full Option Science System (FOSS)

The Full Option Science System (FOSS) was a research-based science curriculum developed for grades K-8 at the Lawrence Hall of Science by a team of researchers from the University of California, Berkeley. The FOSS project started over 20 years ago to meet the need of providing meaningful science education to the students in the USA. In this study students participated in five FOSS modules for third and fourth grade: water, earth materials, ecosystems, changing earth, and structures of life.

Real Engagement in Active Problem Solving (REAPS)

REAPS was a learning model developed by Dr. June Maker and her colleagues (Maker, Zimmerman, Alhusaini, & Pease, 2015). Research on the model is ongoing (Wu, Pease, & Maker, 2015; Gomez-Arizaga, Bahar, Maker, Zimmerman, & Pease, 2016). In this model, the researchers combined three different models, the Thinking Actively in a Social Context (TASC) model, the Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER) model, and Problem Based Learning (PBL), to create a more comprehensive, inclusive, and cohesive model. The REAPS model was not a replacement of any type of school curricula; instead it was a framework to guide teachers and students in their learning environment, and the use of the model involved questioning and using problem-solving approach. In this study, REAPS was used specifically to help students understand science concepts and develop group projects related to the science units. For example, in the water unit, students were divided into groups of 4 or 5 students who built water parks that would be sustainable in desert environments with very little water (Maker, Zimmerman, Gomez-Arizaga, Pease, & Burke, 2010; Zimmerman et al. 2011).

Instruments

Concept Map Assessment

The researchers chose two methods of concept map assessment: Ruiz Primo et al. (1997) and Novak & Gowin (1984). Although several grading systems for concept maps existed in the field, most lacked clarity in methods, which was the main reason we chose these two scoring systems. Both methods had three main components in common: (a) the concepts in the domain of study, (b) the label given to the line connecting the two concepts, and (c) the proposition, which was the combination of the pair of concepts (nodes) and the label.

Novak & Gowin (1984)

The main components of the Novak and Gowin (1984) scoring system were propositions, hierarchy levels, crosslinks, and examples (Table 1).

Table 1. *Scoring Criteria for Concept Maps (Modified from Novak & Gowin, 1984)*

1. Proposition	Are meaningful relationships between two concepts indicated by the connecting lines and linking words? Valid propositions are scored 1 pt.
2. Hierarchy	Does the map have a hierarchy with the more general concept above the specific concepts on the map? Each subordinate valid hierarchy level is awarded 5 pts.
3. Crosslinks	Does the map show meaningful connections between different segments of the hierarchy?
4. Examples	Which specific events or objects are valid, such as "Quartz is a type of rock"? A valid relationship is awarded 1 pt.

Table 2. *Scoring Criteria for Crosslinks*

Score of the crosslink levels	Description of the quality of the crosslink
Invalid-0 pts	The crosslink is incorrect
Below Average - 2 pts	Altough the crosslink is valid, it does not represent a purposeful connection between the two segments of the map
Average - 4 pts	Although the crosslink shows a meaningful connection between two segments, the meaning needs to be clarified more.
Good - 7 pts	The crosslink is valid, correct, and represents a purposeful connection between two segments of the map
Excellent - 10 pts	The crosslink is valid, correct, and shows deep understanding of the relationship between two segments of the map

Propositions were the concepts connected by a linking line, preferably with an arrow if the relationship was directional, and a label. Hierarchy was measured by scoring the number of levels of specificity in information. Levels were rank ordered from the most general concept to the most specific one with the more specific subordinate concepts covered by the concepts above them. Crosslinks were the links that connected one concept segment to another one. They were the most important parts of concept maps in the Novak and Gowin (1984) scoring system because of their potential to represent meaningful connections among concept map sections and to indicate creative ability.

The maps were scored based on each component, and the total scores were calculated. Novak and Gowin’s scoring criteria for crosslinks ranged from 2 points to 10 points based on the crosslinks’ validity and synthesis. If the student made unique or creative cross-links, additional points could be awarded. The researchers modified Novak and Gowin’s scoring criteria for crosslinks, and ranked the scores by degree of validity and quality of synthesis (Table 2).

Ruiz Primo et al. (1997)

The biggest difference between the Novak and Gowin (1984) and the Ruiz Primo et al. (1997) grading systems was a criterion map. In Ruiz Primo et al. (1997) the construction of a criterion map was mandatory while in Novak and Gowin's (1984) it was optional. The criterion map was constructed as a combination of an expert's, a teacher's, and a researcher's concept maps to determine the substantial links between concept pairs. A squared matrix based on the key concepts was made to define all possible links between concept pairs. To determine the substantial links, in this study the teacher, the educator of the gifted, and the scientist constructed their own concept maps.

Table 3. *Accuracy of Propositions (Modified from Ruiz-Primo et al. 1997)*

Accuracy of Proposition	Definition
Excellent	Outstanding proposition. Complete and correct. It shows a deep understanding of the relationship between the two concepts. • metamorphic rocks can be made from igneous rocks
Good	Complete and correct proposition. It shows a good understanding of the relationship between the two concepts. • climate exposes minerals
Average	Incomplete but correct proposition. It shows partial understanding of the relationship between the two concepts. • minerals help geologists
Below Average	Although valid, the proposition does not show understanding of the relationship between the two concepts. • water is earth materials
Invalid/inaccurate	The proposition is incorrect. • limestone is quartz

Note. Adapted from "The Use of Concept Maps in Facilitating Problem Solving in Earth Science"by R. Zimmerman, C. J. Maker, M. P. Gomez- Arizaga, and R. Pease, 2011

The teacher's map served as a point of reference for the substantial links students were expected to have after studying the module. The educator's map was used to be a reflection of the substantial links of the earth materials as a unit of the curriculum in an educational system, and the scientist's map was used to provide the substantial links based on the structure of the science unit.

After the classroom teacher and the expert's discussion about the science unit and vocabulary of the unit, mandatory concepts were listed in written instructions that were given to the students to use when constucting concept maps. Thirty-five mandatory propositions were identified and put on the criterion map for ecosystems. Students were given a list of concepts in their instruction sheets.

A Proposition Inventory was developed to examine the quality and the variation of the propositions (Ruiz Primo et al.,1997), which included the propositions (nodes and links) provided on the three experts' maps. Based on the degree of accuracy, each proposition was classified into one of five categories, and was scored accordingly (Table 3). In addition to the propositions in the criterion map, the ones that were not included in the criterion map but were in the students' maps also were graded for accuracy.

Although in the original scoring system, Ruiz Primo et al. (1997) calculated three forms of concept map scores, the proposition accuracy score, the convergence score, and the salience score, in this study the researchers evaluated only two of the scores: (a) a total proposition accuracy score, the sum of the scores obtained on all propositions; and (b) convergence score, the proportion of valid propositions in the student's map to all mandatory propositions in the criterion map (i.e., the degree to which the student's map and the criterion map converged).

Procedure

Students in third grade were taught water, earth materials, and ecosystem science modules while the students in fourth grade participated in the structures of life and changing earth modules as part of the FOSS science units. Units were taught using the REAPS model.

Students participated in a one-hour training session before creating their concept maps. They created two types of maps: maps from the beginning and fill-in-the blank maps. The teacher then lead a discussion for thirty minutes and answered any questions about making concept maps. To help students develop the pre and post science unit concept maps, they were given written instructions, which included the following elements: (a) a broad question that encompassed the main idea of the science unit, (b) the required concepts from the criterion map developed from the assessment procedure of Ruiz Primo et al. (1997, see above), (c) optional linking words for connecting concepts, and (d) examples of concept maps related to other science topics. Students were asked to draw their maps from the most general concepts to more specific ones, but were not required to make a hierarchical map because any concept on the map could be raised to form a hierarchy (Novak & Gowin, 1984). Students were given 45 to 50 minutes to construct their maps. The teacher, the specialist in education of the gifted, and the scientist were available if the students had any questions and to keep them on task. Before and after each FOSS unit the students created concept maps, which were graded by two special education doctoral students and one scientist.

Data Analysis

Inter-rater Reliability

Each researcher graded the same 30 concept maps out of 230 to obtain data for interrater reliability (13%). A Pearson correlation coefficient (r) was used to determine interrater reliability (STATISTICA software program, 2004) by comparing the raters' the total accuracy scores in the Ruiz-Primo et al. (1997) scoring system and total scores in the Novak and Gowin (1984) scoring system. The correlation between total accuracy scores using the Ruiz-Primo et al. scoring system among the three raters was significant ($p < .05$), and ranged from 0.85 to 0.92. For the total scores in the Novak and Gowin (1984) scoring system the correlation also was significant ($p < .05$), and varied from 0.70 to 0.87. Because of the agreement, the concept maps were divided equally among the raters, and each rater scored one-third of the concept maps.

To determine how the continuous use of concept maps affected the complexity and interconnectedness of concepts, we employed several procedures. First, we examined crosslinks and hierarchy scores of Novak and Gowin`s (1984) scoring system and variation in the number of the two highest level accuracy scores (3s and 4s) of the Ruiz-Primo scoring system to analyze complexity. Second, we calculated total crosslink scores and the number of relationships from Novak and Gowin's scoring system (1984), and the total accuracy scores from the Ruiz-Primo scoring system to determine the increase in interconnectedness of science concepts. The students' pretest and posttest concept map scores were analyzed separately. A t-test for paired samples was used to evaluate the differences in means between groups.

To determine which grading system, (i.e. Ruiz-Primo, Novak and Gowin), showed greater change in complexity, we used a criterion map. According to Novak and Gowin (1984) "a criterion map may be constructed, and scored, for the material to be mapped, and the student scores divided by the criterion map score to give a percentage for comparison" (p. 36). Therefore we graded the criterion maps for each science unit based on the two scoring systems, and divided the students' scores by the criterion map scores to obtain the percentages. Finally, they calculated the changes in percentages between students' pre and post scores, and compared these percentages using t-test paired samples analysis.

3. Results and Discussion

Results

How did repeated use of concept maps affect the complexity and interconnectedness of students' concepts independent of science content?

A paired sample t-test was used to compare the means of post concept maps for six scoring criteria over five sequential science units: Water, Earth Materials, Ecosystems, Changing Earth, and Structures of Life. The increase in complexity of concept maps over time and science units was examined using crosslinks ($M=3.04$, $SD=2.72$) and hierarchy scores ($M=3.00$, $SD=1.29$) of Novak and Gowin`s (1998) scoring system and variation in the number of the two highest level accuracy scores (3s and 4s) of the Ruiz-Primo et al. (1997) scoring system (M=12.13, SD=8.87). The effect sizes have been reported in Table 4.

Table 4. *The t and p Values of the Paired-Samples T-test for the Means of the Post Concept Maps Using the Six Scoring Criteria and Five Science Units*

Pairs	t	p	d
WaterCrossHier-WaterThreeFour	2.188	0.039*	0.446
EarthMaterialsCrossHier-Earth MaterialsThreeFour	0.729	0.473	0.148
EcosystemCrossHier-EcosystemThreeFour	4.842	0.000***	0.988
ChangingEarthCrossHier-ChangingEarthThreeFour	2.494	0.020*	0.509
StructureofLifeCrossHier-StructureofLifeThreeFour	-3.161	0.006**	0.766

Note. Cross=Crosslinks Hier=Hierarchy *=p<0.05, **=p<0.01 ***=p<0.00

To determine the increase in interconnectedness of science concepts, the total crosslink scores (M=12.58, SD=11.57) and the number of relationships (M=24.83, SD=10.59) from the Novak and Gowin scoring system (1984), and the total accuracy scores (M=64.83, SD=34.68) from the Ruiz-Primo et al. (1997) scoring system were used. Descriptive statistics for all units have been presented in Table 5. In general, the students` concept map scores increased for all six criteria from the water unit to the ecosystem unit: the number of cross links, total cross link scores, number of hierarchies, number of 3s and 4s (the two highest proposition accuracy scores), and total accuracy scores. The one exception was that the number of relationships for the second science unit, earth materials, was less than for the first science unit, water. The scores obtained from the selected criteria for the changing earth unit, which was the first unit that the students studied in fourth grade, were less than the ecosystem unit, the last science unit in third grade. The means for the last science unit, structures of life, were more variable than the means for the other units.

Because the ecosystem unit was the 3rd unit, using that unit as a criterion enabled us to make a comparison of students' developments before and after that unit. The ecosystem (3rd) unit scores were significantly higher than those for the water (1st) unit (crosslinks t= 4.92 p=.000, total crosslinks t=4.89 p=.000, number of hierarchy levels t=3.04 p=.000, and accuracy scores t=2.44 p=.02) and earth materials (2nd) units for four of the six scoring criteria (crosslinks t=3.53 p=.000, total crosslinks t=2.97 p=.010, number of hierarchy levels t=2.64 p=.010, and accuracy scores t=2.55 p=.010), and significantly higher than the scores for the number of crosslinks, t=3.23 p=.000, and the total crosslink scores, t=3.21 p=.000 for the changing earth (4th) unit and the structure of life (5th) unit (crosslink t=2.63 p=.010, total crosslink t=3.24 p=.000. Five out of the six scoring criteria for the complexity and interconnectivity of concept maps (e.g., number of crosslinks, total crosslink scores, number of hierarchies, number of relationships, and total accuracy scores) for the structures of life (5th) unit were significantly different from the number of crosslinks, total crosslink scores, number of hierarchies, number of relationships, and total accuracy scores for the water (1st) unit (Table 6).

Table 5. *Means and Standard Deviations of Six Scoring Criteria for the Post Concept Maps for the Five Sequential Units, (t-test, <p.65)*

Science Unit	Scoring Criteria: Novak						Ruiz Primo					
	No. of Crosslinks		Total Crosslinks Score		No. of Hiearchic Levels		No. of Relationship		No. of 3s+4s		Total Accuracy Scores	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Water	0.25	0.53	0.88	1.98	2.04	0.86	21.7	11.98[b]	7.71	7.15	42.17	27.85
Earth Materials	0.98	0.95	4.79	5.63[a]	2.13	0.99	14.38	6.93	9.13	4.29	44.35	16.6
Ecosystems	3.04	2.72[abcd]	12.58	11.57[abcd]	3	1.29[ab]	24.83	10.59[b]	12.13	8.87	64.83	34.68[ab]
Changing Earth	1.04	1.33[a]	4.33	4.98[a]	2.67	0.64	24.13	9.76[b]	11.17	6.65	56.39	25.48
Structure of Life	1.18	1.24[a]	3.24	3.07[a]	3.06	0.75	29.18	10.25[ab]	10.06	6.56	66.35	26.57

Note.[a] Significantly different from the water unit (p< .05); b Significantly different from the earth materials unit (p< .05); c Significantly different from the changing earth unit (p< .05); d Significantly different from structures of life (p< .05)

Table 6. *T-values and P-values for the Comparison of the Ecosystems Unit with the other Four Science Units on Four Scoring Criteria*

Science Unit	Crosslinks		Total Crosslinks		Hierarchic Levels		Accuracy Scores	
	t	p	t	p	t	p	t	p
Water	4.92	0.000***	4.89	0.000***	3.04	0.000***	2.44	0.020*
Earth Materials	3.53	0.000***	2.97	0.010**	2.64	0.010**	2.55	0.010**
Changing Earth	3.23	0.000***	3.21	0.000***	1.14	0.260	0.94	0.350
Structures of Life	2.63	0.010**	3.24	0.000***	0.17	0.870	0.15	0.880

Note. The degrees of freedom for the structure of life unit was df = 38; all others were df = 46. Total scores were calculated using Novak and Gowin's (1984) scoring system and accuracy scores were calculated using Ruiz-Primo`s (1997) scoring system. ***= p< .000, **=p< .01, *= p< .05

Which of the two grading systems, Ruiz Primo et al. or Novak and Gowin, showed greater change in complexity of students' thinking from novice to expert?

A paired-samples t-test was conducted on students' change scores for each science unit to answer this research question. First we calculated the total of the students' crosslink and hierarchy scores for complexity in the Novak and Gowin (1984) scoring system, added these two scores for each student and normalized these scores by calculating the same scores for the criterion map, and changed students' scores into percentages by using criterion map scores. Second, we calculated the total of students' accuracy scores on which they received 3s and 4s for complexity in the Ruiz Primo et al. (1997) scoring system. Next, we calculated the same score for the criterion map and changed the students' scores into percentages to normalize them. We, then, calculated the change in percentages for pre and post scores of each unit for each student. Thus, we calculated the changes in students' percentages within each science unit for both scoring criteria, the total of the students' crosslink and hierarchy scores, and the total of the accuracy scores on which they received 3s and 4s. We then paired five units according to the criteria to measure complexity as follows: Water Crosslink Hierarchy to Water Three Four (M=8.86, SD=19.85), Earth Materials Crosslink Hierarchy to Earth Materials Three Four (M=3.10, SD=20.82), Ecosystem Crosslink Hierarchy to Ecosystem Three Four (M=21.87, SD=22.12), Changing Earth Crosslink Hierarchy to Changing Earth Three Four (M=6.20, SD=12.18), and Structures of Life Crosslink Hierarchy to Structures of Life Three Four (M=-13.14, SD=17.14). Four were statistically significant: water, t= 2.188, p= 0.039, ecosystem, t = 4.842, p=0.000, changing earth, t= 2.494, p= 0.020, and structures of life, t=-3.161, p= 0.006 (Table 4).

Discussion

Novak and Gowin (1984), and Novak and Musonda (1991) suggested that students who use concept maps have the potential to increase knowledge and to improve understanding in science. In this study, we were specifically interested in knowing if sequential use of concept maps improved the complexity and interconnectedness of the concepts students used in their concept maps. We examined the change in complexity and interconnectedness over five science units from third to fourth grade.

We found that in fourth grade, the scores for crosslinks were significantly less than those obtained for the ecosystems (last unit in 3rd grade), but were higher than scores for the first science unit, water. This could be explained by the reduced practice that occurred in three-month vacation that students had between third and fourth grade. Because students did not practice concept mapping for three months, and they started a higher level science unit right after this vacation, regression could occur. On the other hand, the hierarchical levels of the concept maps from the second science unit to the fifth science unit were significantly higher than those for the first science unit, water. This finding supports Novak and Gowin's (1984) and Novak and Musonda's (1991) claim, that the continuous use of concept maps will affect students' understanding of relationships between concepts by increasing complexity and interconnectedness. The third and fifth science units' concept maps' hierarchical levels also were significantly higher than the second science unit's concept maps. Although the number of higher level accuracy scores (3s and 4s) showed a general increase, they were not significantly different from the first science unit, water. The five out of six scoring criteria (the number of crosslinks, total crosslink scores, the number of hierarchical levels, number of relationships, and the total accuracy scores) created for the structures of life (5th) science unit were significantly different from the water (1st) science unit which referred to the change in complexity and interconnectedness of concept maps.

Novak and Gowin's (1984) scoring system was better than Ruiz-Primo et al. (1997) to evaluate complexity in students' thinking except for the earth materials unit. This result might have been due to the unit itself. The Earth Materials unit had vocabulary that was more complex and more unlikely to be encountered in daily life than the other four units. Thus, the students were not as familiar with the vocabulary in this unit as they were in the other four units. This might have resulted in concept maps with lower quality. Another explanation for this non-significant result could have been the time spent on this unit. Because of the school district's established schedule for each school to use the materials, the teacher had a limited amount of time to teach the unit. In other words, because of the complex nature of this unit, the teacher was not able to finish the unit by the time he had to return the materials.

These findings are consistent with several studies. Zimmerman et al. (2011) found that students' scores for accuracy and the complexity level of their maps increased from pre to post test. BouJade and Attieh (2008) also found significant differences between two groups of chemistry students, using and not using concept maps, favoring the concept map group on the knowledge level questions. Austin and Shore (1995) found significant correlations between multi-step problem solving performance, and linkage, score, and good links in students' concept maps.

However, our results were both consistent and inconsistent with the results of the Karakuyu (2010). In his research, he found no significant difference between attitudes and achievement of students using concept maps and students using the traditional method, but found that concept mapping instruction was more effective than traditional instruction in improving student's physics achievement.

The results of this study were also consistent with those of Austin and Shore (1993), who found that high performing physics students' concept maps were very similar to physics experts' maps, and clearly differed from the novices' concept maps. Also, the results were consistent with the study of Chi et al. (1981) who found that the problem schemata of experts clearly differed from the novices' schemata and expert students used more structured problem solving methods using key variables linked together, while novice students applied any available formula (Heyworth, 1999). Experts used their prior and conceptual knowledge and experience during the problem solving process (Hmelo-Silver et al., 2002), and experts spent more time in analyzing the problem, planning, and organizing the data. We believe our results confirm those of the other studies.

We conclude that the number of crosslinks, total crosslink scores, number of hierarchies, number of relationships, and total accuracy scores have the potential to measure change in complexity and interconnectivity of concept maps. Furthermore, repeated use of concept maps has the potential to increase the complexity and interconnectedness of student concept maps, and therefore improve their understanding of science independent of science content.

Limitations

Only one school, two grade levels, and one teacher were involved in the study. The results may be different for another school setting, for another teacher, for students from different grade levels, or students from different ethnic backgrounds, thus may not be generalized to other elementary school populations. Another limitation is the time that was spent for vacation between third and fourth grade. In this vacation, the students did not practice concept mapping skills, so they needed more time at the beginning of the next school year to recall these skills and create complex concept maps.

4. Future Implications and Conclusion

Future Practical Implications

We graded all the pre and post concept map scores for all five science units using both scoring systems, analyzed them, and concluded that Novak and Gowin's (1984) scoring system is better to measure complexity, and therefore is recommended for use in concept map grading rather than the Ruiz Primo et al. (1997) method. Novak and Gowin's (1984) scoring system is definitely less subjective and less time consuming. We found that the complexity and interconnectivity of concepts used in the concept maps increased over time. Because complexity and interconnectivity of the concepts are the main characteristics of experts' maps, students (novices) who consistently use concept mapping skills will be more likely to become like experts in their thinking. We recommend using concept mapping in teaching and assessment.

Future Research Implications

Although we found that Novak and Gowin's (1984) scoring system is better to measure complexity in students' thinking, we agree that both scoring systems have strengths in assessing science knowledge. For example, Novak and Gowin (1984) suggested that concept maps should be structured hierarchically, and the propositions from different segments should be linked to each other a much as possible, while Ruiz-Primo et al. (1997) suggested that the propositions should be graded according to their quality. We believe that, grading concept maps with a scoring system based on the strengths of the two systems, Novak and Gowin (1984) and Ruiz-Primo et al. (1997), and then examining the effectiveness of this system to show change in complexity of students' thinking would make a great contribution to developing students' expertise.

We observed a decrease in the concept map scores for the first science unit of the fourth grade, changing earth. Although, we think this result stems from the fact that students were on vacation between third and fourth grade and did not participate in any activities related to concept maps, a qualitative study can be designed to investigate the reason for this decrease. Students can be interviewed about the changing earth unit, the amount of time they spent to draw concept maps during their vacation, and the possible change in school setting when they become fourth graders.

Theoretical Implications

This study was built on the theoretical framework of expert-novice research. We confirmed that using concept maps in teaching and assessment in certain domains helps students to think more like experts in these domains, and therefore using concept maps is appropriate for differentiated curricula to provoke higher level thinking skills.

Conclusion

Based on the analyses of data, we concluded that the repeated use of concept maps increased the complexity and interconnectivity of concepts independent of science content, and therefore is more appropriate than the traditional assessment methods to use in science. When we compared the two scoring systems, Ruiz-Primo et al. (1997) and Novak and Gowin (1984), we found that the Novak and Gowin scoring system showed greater change in complexity of students' thinking as they progressed from novice to expert. Thus, when assessing students' understanding of science using concept maps, we recommend teachers and educators to use the Novak and Gowin scoring system.

References

[1] Ausubel, D.P., Novak, D.J., & Hanesian, H. (2 ed.). (1978). *Educational psychology: A cognitive view*. New Yok, NY: Holt, Rinehart and Winston.

[2] Besterfield-Sacre, M., Gerchak, J., Lyons, M. R., Shuman, L. J., & Wolfe, H. (2004). Scoring concept maps: An integrated rubric for assessing engineering education. *Journal of Assessing Conceptual Knowledge Using Concept Map Scoring 141 Engineering Education*, 93(2), 105–115. http://dx.doi.org//10.1002/j.2168-9830.2004.tb00795.x

[3] Bhattacharya, K. & Han, S. (2001) Piaget and Cognitive Development. In M. Orey (Ed.) *Emerging perspectives on learning, teaching, and technology*. Athens, GA: University of Georgia Press.

[4] Bransford, J., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. (Expanded ed.). Washington, D.C.: National Academy Press.

[5] BouJaoude, S., Attieh, M. (2008). The Effect of Using Concept Maps as Study Tools on Achievement in Chemistry. *Eurasia Journal of Mathematics, Science & Technology Education, 4*(3), 233-246.

[6] Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive sciences*, *5*(2), 180-195.

[7] Chickean, T., & Shore, B. M. (2014). Cognitive Characteristics of the Gifted. In J. A. Plucker & C. M. Callahan (Eds.), *Critical Issues and Practices in Gifted Education* (pp. 119-133). Waco, TX: Prufrock Press.

[8] Daley, B. J. (1999). Novice to expert: An exploration of how professionals learn. *Adult Education Quarterly*, 49, 133–147. doi: 10.1177/074171369904900401

[9] Dogusoy-Taylan, B., & Cagiltay, K. (2014). Cognitive analysis of experts' and novices' concept mapping processes: An Eye tracking study. Computers in Human Behavior, 36, 82–93. doi:10.1016/j.chb.2014.03.036

[10] Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), The nature of expertise (pp. xv–xxviii). Hillsdale, NJ: Erlbaum.

[11] Gomez-Arizaga, Bahar, Maker, Zimmerman, & Pease. (2016). How Does Science Learning Occur in the Classroom? Students' Perceptions of Science Instruction During the Implementation of REAPS Model. *Eurasia Journal of Mathematics, Science & Technology Education*, *12*(3), 431-455.

[12] Heyworth, R. M. (1999). Procedural and conceptual knowledge of expert and novice students for the solving of a basic problem in chemistry. *International Journal of Science Education*,*21*(2), 195-211. doi: 10.1080/095006999290787

[13] Hmelo-Silver, C. E., Nagarajan, A., & Day, R. S. (2002). It's harder than we thought it would be: A comparative case study of expert–novice experimentation strategies. *Science education*, *86*(2), 219-243. doi: 10.1002/sce.10002

[14] Ingec, S.K. (2009). Analyzing concept maps as an assessment tool in teaching physics and comparison with the achievement tests. *International journal of science education*. *31*(14). 1897-1915. doi: 10.1080/09500690802275820

[15] Karakuyu, Y. (2010). The effect of concept mapping on attitude and achievement in a physics course. *International journal of the physical sciences*, *5*(6), 724-737.

[16] Liu, X. & Hinchey, M. (1996). The internal consistency of a concept mapping scoring scheme and its effect on prediction validity. *International journal of science education*, *18*(8).921-937.doi: 10.1080/0950069960180805

[17] Maker, C.J., & Nielson, A.B. (1996). *Curriculum development and teaching strategies for gifted learners.* (2nd ed.). Austin, TX: Pro-Ed Inc.Maker, C. J, Schiever, S.W. (2005). *Teaching Models in Education of the Gifted.* Austin, TX: Pro-Ed Inc.

[18] Maker, J., & Zimmerman, R., Alhusaini, A., & Pease, R. (2015). Real Engagement in Active Problem Solving (REAPS): An evidencebased model that meets content, process, product, and learning environment principles recommended for gifted students. APEX: *The New Zealand Journal of Gifted Education,* 19(1). Retrieved from www.giftedchildren.org.nz/apex.

[19] Maker, C. J., Zimmerman, B., Gomez-Arizaga, P. M., Pease, R., & Burke, M. E. (2010). Developing Real-Life Problem Solving: Integrating the DISCOVER Problem Matrix, Problem Based Learning, and Thinking Actively in a Social Context In Vidergor, H. E. Harris, C. R. & Subhi, T. Y. , *The Practical Handbook for Teaching Gifted and Able Learners.* The International Centre for Innovation in Education

[20] Novak, J. D., Cañas, A. J. (2006). The origins of the concept mapping tool and the continuing evolution of the tool. *Information Visualization*, 5(3), 175-184. Basingstoke, UK: Palgrave Macmillan.

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

[22] Novak, J. D., Musonda, D. (1991). A twelve-year longitudinal study of science concept learning. *American educational research journal*, *28*(1), 117-153. Sage Publication.

[23] Pinto, M., Doucet, A. V., & Fernandez-Ramos, A. (2010). Measuring students' information skills through concept mapping. *Journal of information science*, *36*(4), 464-480. doi:10.1177/0165551510369633

[24] Ruíz-Primo, M. (2000). On the use of concept maps as an assessment tool in science: What we have learned so far. *Revista Electrónica de Investigación Educativa*, *2* (1). Retrieved month day, year from:http://redie.uabc.mx/vol2no1/contents-ruizpri.html

[25] Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of research in science teaching*, *33*(6), 569-600. doi: CCC 0022-4308/96/060569-32

[26] Ruiz-Primo, M.A., Shavelson, R.J., Schultz, S.E. (1997). On the validity of concept maps-base assessment interpretations: An experiment testing the assumption of hierarchical concept maps in science (CSE 455). Los Angeles, CA: National center for research on evaluation, standards, and students testing; Center for the study of evaluation; University of California Graduate School of Education and Information Studies.

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

[28] Ruiz-Primo, M. A., Shavelson, R. J., Li, M., & Schultz, S. E. (2001b). On the validity of cognitive interpretations of scores from alternative concept-mapping techniques. *Educational Assessment*, *7*(2), 99-141.

[29] Segalas, J., Ferrer-Balas, D., & Mulder, K. F. (2008). Conceptual maps: Measuring learning processes of engineering students concerning sustainable development. *European Journal of Engineering Education*, 33(3), 297–306 http://dx.doi.org/10.1080/03043790802088616

[30] Statistica [Computer Software]. (2004). Retrieved from http://en.freedownloadmanager.org/Windows-PC/STATISTICA.html

[31] 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.

[32] Wallace, B. (2008). The early seedbed of the growth of TASC: Thinking actively in a social context. *Gifted education international, 24*(2-3), 139-155.

[33] Watson, M. K., Pelkey, J. G., Noyes, C., and Rodgers, M. O. (2016). Assessing conceptual knowledge using three concept map scoring methods. *Journal of engineering education*, 105(1), 118–146

[34] Wu, I. C., Pease, R., & Maker, C. J. (2015). Students' perceptions of real engagement in active problem solving. Gifted and Talented International, 30(1-2), 106-121.

[35] Zimmerman, R., Maker, C. J., Gomez- Arizaga, M.P., Pease, R. (2011). The use of concept maps in facilitating problem solving in earth science. *Gifted education international 27*(3): 274-287

Authors

Sema Tan, PhD, Sinop University, Sinop, Turkey, e-mail: sematan@sinop.edu.tr

Omer Erdimez, PhD, University of Arizona, Tucson, AZ, USA, e-mail: omererdimez@yahoo.com.tr

Robert Zimmerman, PhD, University of Arizona, Tucson, AZ, USA, e-mail: rhzimmer@hotmail.com