Journal for the Education of Gifted Young Scientists, 5(1) 1-20, March 2017
e-ISSN: 2149- 360X
http://jegys.org

Research Article

The Use of Concept Maps as a Tool to Measure Higher Level Thinking Skills in Elementary School Science Classes

Omer ERDIMEZ[1], Sema TAN[2], Robert ZIMMERMAN[3]

Received: 21 December 2016 **Accepted:** 16 February 2017

Abstract
The need for assessing higher level thinking skills and using appropriate evaluation methods in programs for the gifted is necessary to better evaluate the effectiveness of these programs. The purpose of this study was to test the claim of Novak and Gowin (1984) that concept maps required higher level thinking skills. A related purpose of the study was to assess which type of measure, a multiple choice test or concept maps, could yield a more accurate or detailed picture of the gains in content understanding of students performing at the highest level on the instruments. A mixed method research design was used to answer the research questions. We concluded that concept maps and multiple choice tests did not measure or require the same thinking skills because of non-significant correlations between the two instruments. Three judges' qualitative analysis also indicated that the number of items requiring higher level thinking skills on multiple choice tests was limited. Concept mapping as a whole process and the crosslinks component of concept mapping required analysis or higher level thinking skills. Also we concluded that concept mapping as a whole process has the potential to show greater gains in scores of the students than the multiple choice items, and crosslinks component of concept mapping that required analysis or higher level thinking skills. To have an alignment between the curricula of programs for gifted students and assessment methods used in these programs, the search for assessment methods requiring higher thinking skills is necessary and needs more investigation.

Keywords
concept maps, multiple choice tests, higher level thinking skills, gifted students, Bloom's Taxonomy

To cite this article:

Erdimez, O., Tan, S., & Zimmerman, R. (2017). The Use of Concept Maps as a Tool to Measure Higher Level Thinking Skills in Elementary School Science Classes. *Journal for the Education of Gifted Young Scientists, 5*(2), 1-20. http://dx.doi.org/10.17478/JEGYS.2017.60

1 PhD, Department of Educational Psychology, University of Arizona, Tucson, AZ 85721. E-mail: oerdimez@email.arizona.edu
2 Assist, Prof, Sinop University, Faculty of Education, Gifted Education Department, Sinop, Turkey
3 University of Arizona, Tucson, AZ

Introduction

The Use of Concept Maps as a Tool to Measure Complex Understanding in Elementary School Science Classes

Gifted students differ from other students because of their high abilities and unique needs for teaching and learning environments. When providing learning experiences for gifted students, three components should be considered to meet their needs and to help them perform at their ability levels: context of the classroom or the school (learning environment), appropriateness of curriculum modifications, and evaluation of the result of the first two steps to determine whether one specific teaching-learning model or combination is better to provide a more appropriate program (Maker, 1982; Maker & Nielson, 1995; Maker & Schiever, 2005, 2010; Tortop, 2015).

Learning environment modifications are the essential and prior modifications before making any modifications to the curriculum. Maker and Schiever (2010) listed eight dimensions of the learning environment that should be modified for optimum growth and development of both gifted and non-gifted students and provide comfort, autonomy, and opportunities to those students. According to Maker and Schiever (2010), the learning environment should be learner centered, facilitative of independence, open, accepting, complex, facilitative of varied groupings, flexible, and facilitative of high mobility.

The next step for a better learning experience for gifted students is modification of the curriculum. This modification can be provided by creating qualitatively different curricula that takes into account the qualities that are special and unique about specific students (Maker & Schiever, 2010). Although modifications to the curriculum might provide a better learning experience for gifted students, the needs and qualities of all gifted students are not similar; therefore, different curriculum models have been developed to provide a better learning experience. Because of different opinions of scholars in the field and different needs of gifted students, three main curriculum models are used by educators: the content model, the process-product model, and the concept model (Van Tassel-Baska, 1994).

The main difference among these models has stemmed from the philosophy behind the models and components of the curriculum that has been differentiated. For instance, in the content model, the main modification was acceleration and advanced placement; in the process-product model, the main modifications were in research and information processing skills; and in the concept model, the main modifications were made to the themes and ideas (Van Tassel-Baska, 1994).

Different curriculum models for gifted students have brought the attention of scholars to develop better differentiated curricula that might meet diverse needs of gifted students and help students perform at their ability levels. As a result of a review of research and literature, Maker and Schiever (2010) developed a comprehensive approach to differentiate curricula for gifted students based on curriculum model principles by taking into account their characteristics. With this comprehensive approach, any component of a curriculum-- content, process, and product-- can be differentiated by taking into account the characteristics of gifted students (Maker, 1982; Maker & Nielson, 1995; Maker & Schiever, 2005, 2010).

In content modification, the principles are abstractness, complexity, variety, organization, study of people, and study of methods. In process modification, the principles are higher level thought, open-endedness, discovery, evidence/reasoning, group interaction, pacing, and variety. The principles to differentiate the product component of the curriculum are real problems, real audiences, transformations, variety; self selected format, and appropriate evaluation (Maker, 1982; Maker & Nielson, 1995; Maker & Schiever, 2005, 2010). These principles can be used by educators to differentiate curricula for gifted learners or can be used to challenge the students in regular education classes and improve their learning experiences. Based on these curriculum principles, Maker and her colleagues developed the Real Engagement in Active Problem Solving (REAPS) model that can be used both in programs serving the gifted and in general education settings as a complement to the regular curriculum (Maker, Zimmerman, Gomez-Arizaga, Pease & Burke, 2010; Zimmerman, Maker, Gomez-Arizaga, & Pease, 2011; Maker, Zimmerman, Alhusaini & Pease, 2015).

Theoretical Framework

The theoretical framework of this study was based on the two principles of Maker and Schiever's (2010) curriculum principles, higher level thought and appropriate evaluation. Curricula for gifted students should require higher level thought (skills) and these skills should be assessed by using appropriate instruments. The REAPS model forced students to use higher thinking skills when they were exposed to real life problems. Furthermore, expecting students to find solutions to real life problems by creating models and concept maps shows how the principles of higher level thought and appropriate evaluation are integrated in the REAPS model.

The REAPS model was a combination of three components: Thinking Actively in a Social Context (TASC), Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER), and Problem- Based Learning (PBL) (Maker, Zimmerman, Gomez-Arizaga, Pease & Burke, 2010; Zimmerman, Maker, Gomez-Arizaga, & Pease, 2011; Maker, Zimmerman, Alhusaini & Pease, 2015). The primary goal of this new model was to develop students` proficiency in solving real-life problems—ill-structured problems that have no pre-determined solutions and can be solved by using a variety of methods.

Reviewing the REAPS model carefully, one can see that most of the principles developed for curriculum modifications are present. Because students are exposed to real life problems and are expected to solve ill defined problems, the curriculum modification principles such as higher levels of thinking, open-endedness, real problems, discovery, group interaction, freedom of choice, complexity, and appropriate evaluation can be observed during the implementation of this model. Although all principles for curriculum modification are equally important to differentiate the curriculum, our focus in this study was on the two principles, higher levels of thinking and appropriate evaluation, that were important to evaluate the effectiveness of differentiation and growth of the students academically.

Need for Assessing Higher Level Thinking Skills

Although in most of the curriculum modification models, the content and process component of the curriculum were differentiated by activities that required higher level thinking skills and complex understanding, the product component of these models was not evaluated by appropriate instruments. In short, alignment between the process and product component of the curriculum is necessary to better evaluate the effectiveness of the modifications and to observe students' academic growth. If curricula is differentiated with activities that require higher thinking skills and complex understanding, products also should be evaluated by instruments to assess those skills.

Unfortunately in programs serving the gifted, alignment between skills required in the process component and the product component is missing. In most of these programs, multiple choice tests, which do not require higher thinking skills such as judgment, analysis, and reflection (Clark & Zimmerman, 2004, p. 128), have been used as an assessment instrument for differentiated curriculum. Using multiple choice tests for assessing the growth of gifted students has been criticized because of the limitations of these tests to measure skills of students at the high and low ends of the normal distribution and because of the ceiling effect of these tests (Clark, 2005; McCoach, Rambo & Welsh, 2012). Using inappropriate instruments such as multiple choice tests for the product component of differentiated curriculum not only prevents educators to measure the effectiveness of programs but also brings criticism about the ineffectiveness.

Concept Maps

As an alternative assessment tool, concept maps might be used to measure higher level thinking skills based on Novak and Gowin's (1984) claim that concept maps require higher levels of thinking skills such as analyzing, synthesizing, and evaluating. Concept maps are the visual images of the concepts that students have in their minds. Concepts are represented in the form of a proposition that is a combination of two concepts labeled with a linking word that explains the relationship between two concepts (Novak & Gowin, 1984; Novak & Canas, 2006). According to Novak and Gowin (1984), the main goal of concept maps has been to symbolize valid relationships between concepts in the form of propositions. They stated that concept maps might help students to internalize new crucial concepts and to integrate those concepts with previous knowledge while stu dents explored their level of knowledge and misconceptions.

Concept maps have had a long history of use in education (Ausubel, Novak, & Hanesian, 1978; Boujaoude & Attieh, 2008; Novak & Gowin, 1984; Novak, & Musonda, 1991; Novak & Canas, 2006; Markow Lonning, 1998; Ruiz Primo & Shavelson, 1996; Zimmerman, Maker, Gomez-Arizaga & Pease, 2011). Although concept maps have been used in different subject areas, they are used especially in science and science education (Asan, 2007; Boujaoude & Attieh, 2008; Markow Lonning, 1998; Novak, & Musonda, 1991; Qarareh, 2010; Rice, Ryan & Samson, 1998; Natividad, 2008; Zimmerman, Maker, Gomez-Arizaga & Pease, 2011).

Although some studies about the use of concept maps in science classes have been conducted, most of these studies were completed with high school and university students. Concept map studies conducted with elementary school students have been rare (Asan, 2007; Qarareh, 2010; Iraizoz Sanzol & Gonzalez Garcia, 2008). Asan (2007) used Cmap Tools, a program used to make and share knowledge as represented in concept maps, to assess 23 fifth grade students` knowledge of science. He then compared the multiple choice test results of students with concept map scores; the correlation between test scores and concept map scores varied from 0.4 to 0.7. Iraizoz Sanzol and Gonzalez Garcia (2008) used Cmap Tools with 24 fifth graders to evaluate the effect of Cmap Tools in a cooperative learning environment. Researchers found that students developed better concept maps when they used Cmap Tools in a cooperative learning environment than when they did concept maps by themselves without the tools. Qarareh (2010) studied the effects of concept maps on student achievement in a fifth-grade science class. He compared students ($n = 80$) studying concepts related to water and matter using concept maps with those using traditional teaching methods. To examine the difference in achievement, the researcher developed an achievement test that consisted of 25 fill-in-the-blank questions to measure the different knowledge levels of Bloom`s Taxonomy. The achievement test was reviewed by curriculum specialists and science teachers; then modified as necessary. The difference in achievement between the group using concept maps and the control group was significant ($p < 0.05$). The author concluded that concept mapping had a greater effect on the academic achievement of the fifth graders than did the traditional teaching method.

Use of Concept Maps in Education of Gifted Students

Novak and Gowin (1984) claimed that concept maps required higher levels of thinking such as analyzing, synthesizing, and evaluating. However, a search for concept map studies conducted in programs serving the gifted or in programs that included a sample containing gifted students resulted in only a few studies (Austin & Shore, 1993; Diket & Abel, 2001; Zimmerman et. al, 2011).

In this study our purpose was to test the claim of Novak and Gowin (1984) that concept maps required higher levels of thinking such as analyzing, synthesizing, and evaluating. Investigating this claim will also help us to understand how appropriate the concept maps are for assessing higher thinking skills for a differentiated curriculum. A related purpose of the study was to assess which type of measure could yield a more accurate or detailed picture of the gains in content understanding of students performing at the highest level on the multiple choice tests and concept maps. The following questions guided the study:

- What was the relationship between students` scores on concept maps and the multiple choice test scores on the same science content?
- Which of the assessment methods was the best to measure higher level thinking skills?
- How did the scores of the three highest students change on multiple choice tests and concept maps between the pre and post assessment?
- How did the scores of students change on multiple choice items and concept maps' components that measured higher level thinking skills for the Ecosystems Module?

Method

Research Design

A mixed method research design was used in this study. Concept maps and multiple choice test scores of these students were compared to determine the similarities and differences in both assessment tools. Also, scores of the three highest-scoring students on both assessment tools were tracked to observe changes over time. Students` scores on five consecutive science modules, three in the third grade science curriculum and the last two in the fourth grade science curriculum, on both assessment tools, were analyzed in this study. Furthermore, three doctoral students were asked to rate qualitatively the multiple choice items in the Ecosystems Module and components of concept mapping that required level thinking skills based on Bloom's revised Taxonomy (Krathwohl, 2002).

Participants

The participants in this study were 23 students who attended third and fourth grade classes for two consecutive years: 2009 and 2010. Students were involved because their classroom teacher agreed to be a part of this research and parents' consent forms were gathered. Institutional research review had been submitted and approved. The sample consisted of 23 students, 14 male and 9 female and their ethnic backgrounds were varied: 10 White American, 10 Hispanic, 2 Asian American, and 1 student from two or more ethnic backgrounds. In the sample, 11 students were identified as gifted based on either the Developing Cognitive Abilities Test (DCAT) or the Raven Progressive Matrices. Although the school was located close to one of the largest universities in the Southwest and socio-economic status of the most students was generally high, 30% of the participants were receiving free or reduced-price lunch.

Setting

This study took place in an elementary school science class for two consecutive years in a Southwestern city of the United States. As a complement to regular science modules from the Full Option Science System (FOSS), the Real Engagement in Active Problem Solving (REAPS) teaching model was implemented; the students solved ill-structured real life problems and participated actively in hands-on projects while creating three-dimensional models for their solutions.

Full Option Science System (FOSS). This research-based curriculum for grades K-8 was developed at the University of California, Berkeley, through forty years of research, to provide a meaningful science education for the diverse school population in the United States (FOSS, 2013). Five of the FOSS science modules, water, earth materials, ecosystems, changing earth, and structures of life, were taught in third and fourth grade.

Real Engagement in Active Problem Solving (REAPS). This model was a combination of three parts: Thinking Actively in a Social Context (TASC), Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses (DISCOVER), and Problem- Based Learning (PBL) (Maker, Zimmerman, Gomez-Arizaga, Pease & Burke, 2010; Zimmerman, Maker, Gomez-Arizaga, & Pease, 2011). The primary goal was to develop students` proficiency in solving real-life problems—ill-structured problems that have no pre-determined solutions and can be solved by using a variety of methods. The REAPS model was not the main curriculum; it was used when students created a group project at the beginning or end of each science module (Zimmerman, Maker, Gomez-Arizaga, & Pease, 2011).

The DISCOVER model is student-centered and has an emphasis on developing multiple intelligences and problem solving skills (Maker, Zimmerman, Gomez-Arizaga, Pease & Burke, 2010). Students solve real life problems with varying degrees of structure, ranging from closed to open-ended (Zimmerman, Maker, Gomez-Arizaga, & Pease, 2011). The problem continuum, adapted from the works of Getzels and Csikszenmihayli (1976, 1967), which had six types of problems, was an important aspect of the DISCOVER curriculum model (Maker, 2001). The type of problem was determined by how well three aspects of the problem, definition, methods for solving it, and the solution, were known by the presenter and the solver (Maker, 1993a). At one end of the continuum, the problem types were structured or well-defined, Type I; at the other end of the continuum the problem types were the least structured or ill-defined, Type VI (Maker, 2005). Based on these three components of the problem, the type of the problem was determined (Table 1).

Every project was started with an open-ended question, statement, or a challenge to expose students to ill-defined real life problems. For instance, in one of the projects, students were asked to construct one of the ecosystems that was taught in the regular curriculum, with its all complex elements (Zimmerman, Maker, Gomez-Arizaga, & Pease, 2011).

Thinking Actively in a Social Context (TASC) was another component of the REAPS model. It was used to teach students how to solve a problem step by step and to monitor their progress through different phases. With using the TASC model, students completed a group project and created their models following the steps provided in this model. TASC model provided a roadmap for the students to follow while they were solving ill-structured real life problems.

The third part of the REAPS model was problem-based learning (PBL). Because students were asked to solve a problem that had a series of challenges, throughout all steps of the process of model creation, they used PBL to develop creative and critical thinking skills. All the members of the research team and the classroom teacher were involved as facilitators to help students think about and create solutions to challenging problems.

Instruments

Multiple choice tests about the content of the FOSS Modules and concept maps were used as assessment tools. Both were administered to the students before and after each science module.

Multiple choice tests. Multiple choice tests were provided for each FOSS module used in the school district. They were developed by a group of K-12 science teachers from the school district after a review of each module. The number of questions on these tests ranged from 24 to 28 for different modules. Because students' scores on Ecosystems module was the only module compared to scores of students on concept maps to measure higher level thinking skills, we only conducted reliability analyses for this module. The Split-half reliability method was used to calculate the scores for Ecosystems module. Based on scores of 23 students, the Spearman-Brown coefficient for equal length was .71, $M= 17.60$, $SD= 2.79$, $N=22$ and was in acceptable range (Haertel, 2006). Because items 11 and 23 had zero variance these items were automatically excluded by SPSS and analyses were made based on 22 items of the Ecosystems module.

Concept maps. Concept maps are graphical tools used to represent the knowledge of students (Novak & Canas, 2006). They consist of propositions, the smallest meaning unit that is a combination of two concepts (nodes) in circles or boxes; and a linking word describing the relationship between two concepts (Novak & Gowin, 1984). Since the development of concept maps, many scoring systems have been developed (Austin & Shore, 1993; Novak & Gowin, 1984; Ruiz Primo et. al, 1997) based on the purpose for their use. Although many scoring systems have been used, we chose to use Novak and Gowin`s (1984) because it was more appropriate to the purpose. Because of the participants' exposure to open-ended real life problems, a scoring system that would allow the students to show their higher levels of thinking and problem solving skills was appropriate.

Propositions, hierarchy levels, crosslinks, and examples were the four elements of the Novak and Gowin concept map scoring system (Table 2). Propositions were the smallest pieces of students' knowledge, and consisted of two nodes connected by a linking word. Hierarchy was measured by scoring the number of levels of hierarchy: the most general concept would be the highest level; thereafter, the levels were ranked from the next most general concept to the most specific concept. The more specific subordinate concepts were covered by the concepts above them. Another scoring criterion was example, specific objects or events that were valid. Crosslinks connected one segment of the concept hierarchy to another segment. In Novak and Gowin`s (1984) scoring system, the crosslink was considered the most important part of the concept map because it showed a meaningful connection between concept map sections, indicated creat ive ability, and had the potential to measure higher levels of thinking skills (Novak & Gowin, 1984). Although Novak and Gowin (1984) assigned different scores for each component of concept maps in their scoring system, they assigned the highest score to crosslinks, 2 to 10 times the points assigned for a proposition, based on the quality of the crosslink. We created a specific rubric for the crosslink component of the concept maps based on the suggestion of Novak and Gowin (see Table 2). The total score for Novak and Gowin's (1984) scoring system was obtained by adding the scores from the four components.

Procedures

The FOSS science modules were taught over the academic year. In third grade, water, earth materials, and ecosystems and in fourth grade, structures of life and changing earth modules were taught. The REAPS model was not the main curriculum, but was used as a complement at the end of each module. When students were exposed to real life problems and were asked to create group projects to solve real life problems related to the specific science module, the REAPS model was the framework for this process.

At the beginning of third grade, students participated in a concept development discussion before concept map creation, and students practiced making maps. In this discussion, a main topic or concept from the curriculum was chosen to show students how to organize ideas related to a topic and how concepts can be connected and related to that main topic or concept. For the concept development session, the Hilda Taba Teaching Strategies were used to teach students how to list, label, group, subsume, and to recycle all the previous concepts related to the main topic (Maker & Zimmerman, 2008). After the concept development discussion, at the beginning of third grade, students participated in a one-hour concept map training session. They practiced making concept maps and discussed the concept maps they created. After the first session, the students practiced with fill-in-the-blank concept maps, and discussed the results before they constructed their own maps for the science unit. The second session was thirty minutes long.

At the beginning of each science module, pre concept maps and multiple choice tests were administered to students. After each science module, including the group project and post concept maps, post multiple choice tests were administered. In consecutive concept mapping sessions, pre and post, written instructions were given to help them develop better concept maps. The instruction sheet included a broad question that encompassed the main concept of the science unit, a list of concepts created from the three expert's agreement about the most important concepts taught in the module and examples of words to use to connect the concept on the map. Although sample connecting words were given, students did not have to use connecting words. However, they were instructed to use as many of the concepts on the list as possible. Students were asked to start from a more general concept and move to more specific concept when they created their maps. This also was not mandatory and students had freedom not to follow the suggestion to create hierarchical concept maps. We believed that providing written instructions might help the students to develop better concept maps. Each concept mapping session, pre and post, took 45 to 50 minutes. The classroom teacher, the specialist in education of the gifted, and the scientist were available if the students had any questions and to keep students on task.

Three experts were involved: one scientist, one researcher who was an expert in the field of education of the gifted, and the classroom teacher. Two special education doctoral students and the scientist scored the concept maps. Two special education doctoral students participated a two hour training for scoring concept maps with the scientist in two sessions.

Data Analysis

To answer the first research question, students' pre-concept map scores were correlated with pre multiple choice test scores and post-concept map scores were correlated with post multiple choice test scores by using Pearson Product Moment Correlations. To answer the second question, a qualitative process was employed. To identify the multiple choice test items and concept map components that required higher level thinking skills, a judges' worksheet, in which Bloom's revised Taxonomy (Krathwohl, 2002) and concept map components were defined, was given to three judges, special education doctoral students in their last year of study. The three judges had teaching experience in different levels of education and were knowledgeable about Bloom's Taxonomy. In addition, multiple choice test questions from one of the modules, Ecosystems, and three students' post concept maps of the same module were attached to the worksheet for evaluation by the judges. The judges first were asked to read information about Bloom's Taxonomy and then to classify each question on the multiple choice test of the Ecosystem module based on this taxonomy. Second, the judges were asked to read information about concept maps (Novak & Gowin, 1984) and examine three sample concept maps of the students to become more familiar with the mapping process. Third, the judges were asked to classify concept mapping as a whole process and the components of the concept mapping process—proposition, hierarchy, crosslinks, and examples—based on Bloom's Taxonomy.

To answer the third research question, the changes in scores was calculated for the three students who scored the highest on both instruments for each science module between pre and post assessment. A one-way ANOVA was used to compare the change in the scores for the five science modules.

To answer the fourth question, a two step process was used. First, based on the three judges' ratings, the multiple choice items and concept map components that required analysis or higher levels of thinking skills was determined for the Ecosystems module. Then, all students' score changes on multiple choice items and concept maps components that required these skills were calculated for the ecosystems module. Finally, the change in students' scores on multiple choice items and concept maps components that required analysis or higher level thinking skills between pre and post assessment was compared by using one-way ANOVA.

Inter-rater reliability. To determine the inter-rater reliability for scoring of concept maps, 30 (13%) of 230 concept maps were scored separately by three raters according to Novak and Gowin`s (1984) scoring criteria. The correlation among the three raters for Novak and Gowin's (1984) total scores was significant ($p<.05$) and correlations varied from 0.70 to 0.87. Because of the significant correlations among scores of the three raters, we were able to divide all the concept maps equally among the raters, and each rater scored one-third of the total concept maps.

Results

Question 1: What was the relationship between students' scores on concept maps and the multiple choice test scores of the same science content?

To answer this question, a Pearson Product Moment Correlation was used. Students' multiple choice pretest scores were correlated with concept map pretest scores and multiple choice posttest scores were correlated with concept map posttest scores of the same science module. Out of 10 comparisons, students' scores on the two instruments were correlated at a significant level on only two units of study, pre Water and post Changing Earth (Table 3). The correlation between multiple choice pretest scores and concept map pretest scores for the Water module was significant, $r= .43$, $p= .037$, and the correlation between multiple choice posttest scores and concept map posttest scores for the Changing Earth module was significant, r= .62, p= .001.

Question 2: Which of the assessment methods was the best to measure higher level thinking skills?

Each judge's classification of multiple choice items in the Ecosystem module based on Bloom's Taxonomy have been shown in Table 4. The judges differed in the numbers of items they classified as requiring thinking at first three levels: Judge A, 11; Judge B, 15; and Judge C, 18. They also differed in the numbers they classified as requiring thinking at the next two levels: Judge A, 13; Judge B, 9; and Judge C, 6. None of the judges classified any of the questions as requiring thinking skills at the highest level of the taxonomy, creation. The three judges agreed that items 9, 10, 15, and 16 on the multiple choice tests of the Ecosystems module required analysis or higher level thinking skills. Inter-rater agreements among the three judges for multiple choice items were calculated by using Spearman's rank correlation coefficient. Agreements among all three judges were significant and ranged from r= .51(p= .011) to r= .86, (p= .000).

Each judge's classification of concept map components and concept mapping as a whole process based on Bloom's taxonomy have been shown in Table 4. Judges A, B, and C classified the concept mapping process and its two components as requiring thinking at the three highest levels of the taxonomy. Although the three judges classified concept map components differently, all of them agreed that concept mapping as a whole process and crosslinks required analysis or higher level thinking skills. Inter-rater agreements among the three judges were calculated by using Spearman's rank correlation coefficient. The only significant correlation was between Judge B and Judge C, r= .89, p= .042.

Question 3: How did the scores of the three highest-scoring students change on multiple choice tests and concept maps between pre and post assessment on five science modules?

The changes in scores of the three highest scoring students on pre and post assessment, on both instruments were calculated for five science modules. The mean raw score change between pre and post assessments for concept maps was 11.66 with a standard deviation of 14.09. The mean raw score change between pre and post assessments for multiple choice tests was 12.78 and the standard deviation was 11.68. To assess the raw score changes on both instruments, all were converted to z-scores and compared by using one-way ANOVA. The raw score change for the three highest students between pre and post assessment for five science modules, on both instruments, multiple choice tests and concept maps, was not significantly different, $F(1,28)= 3.50$, $p= 0.72$.

Question 4: How did the scores of students change on multiple choice items and concept map components that measured higher thinking skills on the Ecosystems module?

The three judges agreed that items 9, 10, 15, and 16 on the multiple choice test of the Ecosystems module required analysis or higher level thinking skills. The crosslinks component and concept mapping as a whole process were the two items related to concept maps that all three judges agreed require analysis or higher level thinking skills. A one-way ANOVA was used. The mean change scores between pre and post assessment for the multiple choice items (M= 1.56, SD= 4.57), crosslinks scores (M= 7.88, SD= 11.8), and total concept map scores (M= 17.92, SD= 20.91) that required analysis or higher level thinking skills were calculated.

A significant difference was found between mean change scores of pre and post assessment for the multiple choice items, crosslink scores, and total concept map scores that rated as required analysis or higher level thinking skills by the judges, $F(2,69)= 8.18$, $p=.001$. By using Tukey post-hoc test, we found that the change in total concept map scores was significantly higher than the change in the multiple choice items (MD= 16.35, SE= 4.08, p= .000) and the change in crosslinks scores (MD= 10.04, SE= 4.08, p= .043). The change in crosslinks scores was not significantly different from the change in the multiple choice items (MD= 6.31, SE= 4.08, p= .275).

Discussion

To answer first question, the relationship between concept maps and multiple choice tests was answered by using the Pearson Product Moment correlation to determine the relationship between students' scores on pre multiple choice tests and pre concept map scores and post multiple choice test scores with post concept map scores for the five science modules. Out of 10 comparisons, students' scores on the two instruments were correlated significantly only on two units of study. In the other eight comparisons, no significant correlations were found between students' multiple choice test scores and concept map scores. In previous studies, some researchers found significant correlations between multiple choice test scores and concept maps scores. Ruiz Primo et al., (1997) found correlations that ranged from .36 to .40, Rice et al., (1998) found correlations that ranged from .41 to .70, and Asan (2007) found correlations that ranged from .40 to .70 between multiple choice test scores and concept map scores. Although we found significant correlations between multiple choice test scores and concept maps scores on two units of the study, our results are inconsistent with previous research because of the non-significant correlations on the other eight comparisons.

The non-significant correlations between multiple choice test scores and concept map scores might be explained by reviewing the thinking skills that each instrument requires or measures. Clark and Zimmerman (2004, p.128) claimed that multiple choice tests do not require higher thinking skills such as judgment, analysis, and reflection. Novak and Gowin (1984) claimed that concept maps require higher thinking skills such as analysis, evaluation, and creation. Based on our results and previous research about the skills that these instruments measure (Clark and Zimmerman, 2004; Novak and Gowin, 1984), we might conclude that concept maps and multiple choice tests do not measure the same science knowledge or do not require the same thinking skills because of the non-significant correlations between multiple choice test scores and concept map scores.

To answer the second question of this study, we further analyzed the claim of Novak and Gowin (1984) that concept maps require or measure higher thinking skills by asking three doctoral students who had teaching experiences in different educational settings and who were familiar with Bloom's Taxonomy to classify multiple choice items of one module (Ecosystems) and the components of concept maps based on Bloom's revised taxonomy (Krathwohl, 2002). Although the agreement among the three judges for multiple choice items was significant, agreement between only two judges (Judge B and C) was significant for the concept map components. The reason for non-significant correlations for classification of concept map components might have been from the limited number of items that the judges had to classify for concept maps based on the taxonomy. The three judges classified 24 multiple choice items and only 5 components of the concept maps. Having the limited number of items to classify for concept map components might have caused the non-significant correlations among the judges. All three judges agreed that only 4 out of 24 items on the Ecosystems module's multiple choice test required analysis or higher level thinking skills. Although a significant correlation was found only between two judges for classification of concept map components, all three judges agreed that concept mapping as a whole process and the crosslinks component of concept maps require analysis or higher thinking skills. These qualitative findings support the claim of the Novak and Gowin (1984) that concept maps require higher level thinking skills.

In the third question of this study, we investigated the relationship between performance of the three highest students on both instruments for the five science modules. The change in scores of the three highest students on both instruments between pre and post assessment for each science module were calculated and turned into z-scores to be able to compare these changes. No significant differences were found between the three highest students' score change between pre and post assessment for five science modules, on both instruments.

One of the reasons for this non-significant result might be the small sample size. Because we had five science modules and we selected the three highest students for each science module and for each instrument, we had only 15 students' score change for each instrument. Having 15 students in each group might have decreased the variability of the scores and resulted in non-significant results. Further investigation is needed by increasing the number of students in the sample or increasing the percentages of students whose scores are analyzed to be able to make a better comparison.

To answer our last question, we tested the accuracy of the judges' classification of multiple choice items and concept map components based on the taxonomy. The three judges agreed that four items on the multiple choice tests of the Ecosystems module, concept mapping as a whole process, and the crosslinks component of the concept mapping process require analysis or higher level thinking skills. To be able to see students' score change on the multiple choice items, concept mapping process, and crosslinks that require analysis or higher level thinking skills, we calculated the change in scores of the students between pre and post assessment for the Ecosystems module. By comparing these change scores we could observe the similarities and differences between those components of the concept maps and the multiple choice items that require higher level thinking skills. The changes in scores between pre and post assessment for the multiple choice items, crosslink scores, and total concept map scores that required higher level thinking skills were significantly different from each other. The Tukey post-hoc test revealed that the change in total concept map scores was significantly higher than the change in multiple choice items and the change in crosslink scores. However, the change in crosslink scores was not significantly different from the change in the multiple choice items.

We conclude that concept mapping has the potential to measure higher thinking skills of the students and has the potential to show a greater increase in scores of the students between pre and post assessment. Based on the results of the third question, we confirm the claim of Novak and Gowin (1984) that concept maps require or measure higher thinking skills. Because total concept map scores had a greater increase from pre to post assessment when comparing crosslink scores and the four multiple choice items that measure higher level thinking skills, we suggest the use of concept maps as a better way to measure higher level thinking skills of students.

Limitations

One of the limitations of this study was the small sample size. Only 24 students participated: students from one classroom and one school in a southwestern city of the U.S. For this reason, results of this study should be generalized with caution. The other limitation of this study was the number of judges who classified multiple choice items and concept map components. In some situations low or non-significant correlations occur because of limited numbers of items or judges.

Theoretical Implications

Assessment methods that will measure higher level thinking skills of gifted students and general education students has been gaining more attention. In this study, we compared multiple choice tests and concept maps to assess their abilities to measure higher level thinking skills of the students. Based on our results, we conclude that concept maps have the potential to measure higher thinking skills and we provided evidence for the claim of Novak and Gowin(1984) that concept maps require higher level thinking skills in assessing gifted or high performing students and in assessing skills all students need for the future.

Austin and Shore (1993) concluded that concept maps constructed by high performing students differed from those of average performing students, and closely resembled the concept maps of physics experts. Although we did not compare the performance of high and average students on the concept mapping process, this finding should be retested in different school settings and grade levels to examine the effectiveness of this assessment method with high performing or gifted students in future research. We suggest that researchers study more about the use of concept mapping with gifted students to better understand its effectiveness in measuring higher level thinking skills.

Practical Implications

We suggest the use of concept maps as an assessment method in general education programs and in programs serving gifted students. In contrast to multiple choice tests, concept maps have the potential to measure higher level thinking skills and to show greater increase in students' scores between pre and post assessments than multiple choice tests. Because concept mapping does not have a predetermined maximum score, the concept mapping process will eliminate the ceiling effect in the assessment process and students will have the chance to show a greater increase in knowledge and skills.

Conclusion

The results of this study provided some evidence for the claim of Novak and Gowin (1984) that concept maps measure higher level thinking skills. The three judges agreed that concept maps measure higher thinking skills such as analysis, evaluation, and creation. Also, the results of our last question showed that the change in scores of students on the concept maps is greater than the change on multiple choice items and crosslink scores that measure higher thinking skills. Based on these results, we suggest the use of concept maps as an assessment method both with general education students and with students in programs serving the gifted.

References

Asan, A. (2007). Concept mapping in science class: A case study of fifth grade students. *Educational Technology & Society, 10*(1), 186-195. Retrieved from http://ifets.info/journals/10_1/17.pdf

Austin, L. B., & Shore, B. M. (1993). Concept mapping of high and average achieving students and experts. *European journal of ability, 4*(2), 180-195. doi: 10.1080/0937445930040207

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

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.

Clark, G., & Zimmerman, E. (2004). *Teaching Talented Art Students: Principles and Practices*. New York: Teachers College Press

Clark, L. (2005). Gifted and growing. *Educational Leadership, 63*(3), 56-60

Diket, M. R., & Abel, H. T. (2001). Metacognitive instrument for tracking graduate student learning in gifted education. *Gifted Child Quarterly, 45*(1), 24-34. doi: 10.1177/001698620104500105

Getzels, J. W., & Csikszentmihalyi, M. (1967). Scientific creativity. *Science Journal, 3*(9), 80-84, retrieved from http://www.apa.org/

Getzels, J. & Csikszentmihalyi, M. (1976). *The creative vision: A longitudinal study of problem finding in art.* New York: Wiley.

Frasier, M. M., & Passow. A. H. (1994). *Toward a new paradigm for identifying talent potential (Research Monograph 94112).* Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut. Retrieved from http://www.gifted.uconn.edu/nrcgt/reports/rm94112/rm94112.pdf

Haertel, E. H. (2006) Reliability in Brennan, R L. (Ed.) *Educational Measurement*, Westport, CT: American Council on Education/Prager Publishers, 4th ed., 65-110.

Irazioz Sanzol, N. & Gonzalez Garcia, M. F (2008) *The concept map as an aid to cooperative learning in primary education: A practical Experiment*. In 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.

Krathwohl, R. D. (2002). A revision of Bloom's Taxonomy: An overview. *Theory into Practice, 41*(4), 212-218.

Maker, C. J. (1982). *Teaching Models in Education of the Gifted*. Rockville, MD: Aspen Systems Corp.

Maker, C. J. (1993a). Creativity, intelligence, and problem-solving: A definition and design for cross-cultural research and measurement related to giftedness. *Gifted Education International, 9*, 68-77. doi: 10.1177/026142949300900020

Maker, C. J. (2001). DISCOVER: Assessing and developing problem solving. *Gifted Education International, 15,* 232-251.

Maker, C. J. (2005). *The DISCOVER project: Improving assessment and curriculum for diverse gifted learners.* The National Research Center on the Gifted and Talented (NRC/GT). retrieved from http://www.gifted.uconn.edu/nrcgt/reports/rm05206/rm05206.pdf

Maker, C. J, Schiever, S.W. (2005). *Teaching Models in Education of the Gifted.* (3rd ed.) Austin, TX: Pro-ed Inc.

Maker, C. J, Schiever, S.W. (2010). *Curriculum development and teaching strategies for gifted learners* (3rd ed.). Austin, TX: Pro-Ed.

Maker, C. J., & Nielson, A. B. (1995). *Curriculum development and teaching strategies for gifted learners.* (2nd ed.). Austin,TX: Pro-Ed Inc.

Maker, C. J. & Zimmerman, B. (2008). *Problem solving in a complex world: Integrating DISCOVER, TASC, and PBL in a teacher education project.* Gifted Education International, 24, 160-178.

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.

Maker, C. J., Zimmerman, B., Alhusaini, A., Pease, R. (2015) Real Engagement in Active Problem Solving (REAPS): An evidence-based 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.

Markow, P., & Lonning, R. (1998). Usefulness of concept maps in college chemistry laboraties: Students' perceptions and effects on achievement. *Journal of Research in Science Teaching, 35,* 1015-1029.

McCoach, B. D., Rambo, E. K., & Welsh, M. (2013). Assessing the growth of gifted students. *Gifted Child Quarterly, 57*(1), 56-67. doi: 10.1177/0016986212463873

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.

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 study of science concept learning. *American Educational Research Journal, 28*(1), 117-153. Sage Publication.

Qarareh. O. A. (2010). The Effect of Using Concept Mapping in Teaching on the Achievement of Fifth Graders in Science. Stud Home Comm Sci, 4(3): 155-160. Retrieved from http://www.krepublishers.com/02-Journals/S-HCS/HCS-04-0-000-10-Web/HCS-04-3-000-2010-Abst-PDF/HCS-04-3-155-10-127-Qarareh-A-O/HCS-4-3-155-10-127-Qarareh-A-O-Tt.pdf

Rice, C. D., Ryan, M. J., & Samson, M. 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.

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

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.

Tortop, H.S. (2015). Üstün Zekalılar Eğitiminde Farklılaştırılmış Öğretim Müfredat Farklılaştırma Modelleri *[Differentiated Instruction and Curriculum Differentiation Models in Gifted Education]*. Düzce: Genç Bilge Yayıncılık.

Van Tassel-Baska, J. (1994). *The national curriculum development projects for high ability learners: Key issues and findings.* In N. Colangelo, S. G. Assouline, & D. L. Ambroson (Eds), Talent Development: Proceedings the 1993 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development, (pp. 19-38). Dayton, OH: Ohio Psychology Press.

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.

Appendix

Table 1

Problem Continuum

Type		Problem		Method		Solution	
		Presenter	**Solver**	**Presenter**	**Solver**	**Presenter**	**Solver**
Closed	**I**	Specified	Known	Known	Known	Known	Unknown
	II	Specified	Known	Known	Unknown	Known	Unknown
	III	Specified	Known	Range	Unknown	Known	Unknown
Open-Ende	**IV**	Specified	Known	Range	Unknown	Range	Unknown
	V	Specified	Known	Unknown	Unknown	Unknown	Unknown
	VI	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown

Note. Adapted from "Problem Continuum" by Maker, J., & Schiever, W. (2010). *Curriculum development and teaching strategies for gifted learners* (3rd Ed.). Austin, TX: Pro-Ed

Table 2

Scoring System for Concept Maps

1. Proposition	The meaningful relationship between two concepts indicated by the connecting lines and linking words. Valid propositions are scored 1 point.
2. Hierarchy	Is there a hierarchy with the more general concept above the specific concepts on the map? Each subordinate valid hierarchy level awarded 5 points.
3. Cross links	Does the map show meaningful connections between different segments of the hierarchy?
(a) Invalid -	Incorrect Crosslink - 0 pt
(b) Lack of effort -	Although valid, the crosslink does not represent a purposeful connection between the two segments- 2 pts
(c) Poor -	Although the crosslink shows a meaningful connection between two segments, the meaning needs to be clarified more- 4 pts
(d) Good -	Valid, correct and the crosslink represents purposeful connection between two segments-7 pts
(e) Excellent -	Valid, correct and shows deep understanding between two segments-10 pts
4. Examples	Specific events or objects that are valid, such as, "Quartz is a type of rock." Valid relationship awarded 1 point.

Note. Adapted from Novak & Gowin, (1984). *Learning how to learn.* New York : Cambridge University Pres

Table 3

Descriptive Statistics for both Assessment Methods and the Correlations between methods

Science Modules	Multiple Choice				Concept Maps				Correlations			
	Pretest		Posttest		Pretest		Posttest		Pre Scores	P	Post Scores	P
	M	SD	M	SD	M	SD	M	SD				
Water	57.9	12.65	80.82	9.24	19.04	9.04	33.75	15.86	0.43*	0.037	0.33	0.119
Earth Materials	54.5	12.36	74.00	11.97	22.16	10.29	33.95	14.03	0.33	0.122	0.14	0.52
Ecosystems	63.23	14.05	82.29	9.69	35.33	15.73	53.25	24.73	0.16	0.444	0.29	0.173
Changing Earth	60.50	15.02	78.00	8.98	32.04	13.03	41.79	13.04	0.10	0.638	0.62**	0.001
Strcs of Life	65.74	9.96	93.2	6.35	45.05	8.59	48.17	12.55	0.27	0.302	0.37	0.143

Note. * significant at $p < .05$, ** significant at $p < .001$, Strcs of Life= Structures of Life