

1-1-2011

# An Action Research Study to Determine the Feasibility of Using Concept Maps as Alternative Assessments by a Novice Teacher

Nancy Smith Mitchell  
*Portland State University*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/open\\_access\\_etds](https://pdxscholar.library.pdx.edu/open_access_etds)

**Let us know how access to this document benefits you.**

---

## Recommended Citation

Mitchell, Nancy Smith, "An Action Research Study to Determine the Feasibility of Using Concept Maps as Alternative Assessments by a Novice Teacher" (2011). *Dissertations and Theses*. Paper 603.  
<https://doi.org/10.15760/etd.603>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

An Action Research Study to Determine the Feasibility of Using Concept Maps as  
Alternative Assessments by a Novice Teacher

by

Nancy Smith Mitchell

A thesis submitted in partial fulfillment of the  
requirements for the degree of

Master of Science in Teaching  
in  
General Science

Thesis Committee:  
Michael Flower, Chair  
Liza Finkel  
Melissa Potter

Portland State University  
© 2011

## **Abstract**

This action research study investigated the feasibility of a novice teacher using concept maps as assessments in secondary science classes. The subjects in this study were the researcher, a novice pre-service science teacher, and students (n=35) in two classes of mixed-grade Foundations in Physics and Chemistry. This study tracked student and teacher experiences over an instructional unit in astronomy. All students received a 50-minute lesson on concept mapping, one class received three additional concept mapping lessons during the unit, then all students took a unit test with a paired short answer and concept map question. Student surveys were conducted to gather student feedback, and teacher reflective journaling was used to track teacher data. Data were analyzed using descriptive statistics. The results indicated that the extra concept mapping lessons did not result in higher scores on the concept maps or the paired short answer responses. The teacher journaling revealed that using concept mapping as an assessment tool was possible for a novice teacher. Advantages and barriers were identified.

## Table of Contents

Abstract	i
List of Tables	iv
List of Figures	v
Introduction	1
Literature Review	5
Overview of Literature Review	5
Brief Background on Assessment in the United States	5
Assessments	7
Alternative Assessments	12
Concept Maps	18
Methods	26
Overview	26
Participants	28
Treatment	32
Instruments	34
<i>Concept map lessons.</i>	34
<i>Tests.</i>	35
<i>Post-test survey.</i>	38
<i>Teacher journaling.</i>	39
Procedure	39
Data Analysis	41
Findings	43
Research Question 1	43
<i>Student response case studies.</i>	43
<i>Student scores.</i>	49
<i>Journal review.</i>	53
Research Question 2	58
<i>Student surveys.</i>	58
Discussion	62
Research Question 1	62
<i>Case studies.</i>	62
<i>Student scores.</i>	63
<i>Advantages and barriers.</i>	64
Research Question 2	67
<i>Student surveys.</i>	67
Limitations and Recommendations	69
References	73
Appendix A: Lesson Materials	76
Concept Map Lesson	76

Big Bang Theory and Origins of the Universe, Unit Day 2	83
Life Cycle of Stars, Unit Day 3	87
Origin of the Solar System, Unit Day 4	89
The Moon, Unit Day 5	91
Mt. Nose, Unit Day 6	93
Eclipses, Unit Day 7	94
Concept Map Lessons, Class E	95
Concept Map Test Questions	99
Concept Map Survey	102
Appendix B: Case Study Student Samples	103
Class E Case Study Samples	103
Class E Case Study Samples	109
Appendix C: Research Approval Documents	114

## List of Tables

<i>Table 1. Ethnicity of student populations at AHS and the high school district</i>	29
<i>Table 2. Ethnicity of city and state where AHS is located (United States Census Data, 2009).</i>	29
<i>Table 3. Public high school characteristics of AHS and the high school district</i>	29
<i>Table 4. AHS Bell Schedule.</i>	30
<i>Table 5: Overview of Astronomy Unit Lessons.</i>	34
<i>Table 6. Novak and Gowin Concept Map Scoring System (1984)</i>	37

## List of Figures

<i>Figure 1: Example of a concept map for water (Novak &amp; Gowin, 1984)</i>	19
<i>Figure 2: Examples of Fill-in-the-node and Fill-in-the-lines concept maps (Shavelson, et al. 2005)</i>	20
<i>Figure 3: Study Flow Chart Showing E and C Conditions</i>	27
<i>Figure 4. Flow Chart of Concept Map Scoring System.</i>	38
<i>Figure 5. Overall Student Test Score by Class Condition</i>	50
<i>Figure 6. Average student test score on paired question by class condition</i>	50
<i>Figure 7. Student Test Score on Star Life Cycle Test Question by Question Format.</i>	51
<i>Figure 8. Student Test Score on Galaxy Test Question by Question Format.</i>	52
<i>Figure 9. Student Test Score on Eclipse Test Question by Question Format.</i>	53
<i>Figure 10. Histogram of minutes spent scoring concept maps sorted by class.</i>	57
<i>Figure 11. Student Assessment of Time Spent in Class on Concept Mapping.</i>	59
<i>Figure 12. Student Self-Assessment of Preparedness for Concept Map Question.</i>	59
<i>Figure 13. Student opinion on amount of class time spent concept mapping and average concept map scores by class.</i>	60
<i>Figure 14. Student self-assessment of preparation for concept map and average concept map score by class.</i>	60
<i>Figure 15. Student preference for question format.</i>	61

## Introduction

The amount of testing students are subjected to in the United States now surpasses most other countries and is seen as excessive by many educators (Kohn, 2000). Many of the critics of our nation's educational system find fault with traditional tests, which rely primarily on selected-response items, and especially high-stakes standardized testing. These tests are fraught with problems such as socioeconomic and racial bias, and diversion of resources from instructional purposes for test preparation purposes (Gonzalez-Espada, 2009; Kohn, 2000; Lawrenz, Huffman, & Welch, 2000; Zurcher, 1998). Instead of using selected-response tests, proponents of educational reform recommend "alternative assessments" that provide wider opportunities for students to demonstrate their knowledge. For example, Grant Wiggins (1993) recommends "authentic" tests that measure how students apply their conceptual understanding to situations they are likely to encounter in daily life.

There is an abundance of research and theory supporting the use of alternative assessment (Klassen, 2006; Shavelson, Ruiz-Primo, & Wiley, 2005; Wiggins, 1993), but this pre-service teacher's recent observations in a classroom were eye-opening. A veteran secondary high school biology teacher gave her class an alternative assessment as a summative measure of knowledge gain. The result was a chaotic class period that lacked meaningful demonstration of students' knowledge. Clearly there is more to alternative assessment than a teacher's selection of a nontraditional test format. In order for any assessment to have meaning, it must be carefully designed to allow students to demonstrate understanding of the learning objectives.



It is essential for pre-service teachers preparing to teach science at the secondary level to develop the ability to assess effectiveness of instruction. In light of the problems with selected response tests, novice teachers need to learn how to use various alternative assessment methods. Concept mapping is an alternative assessment that some claim allows for a better evaluation of changes to students' conceptual knowledge in students over time compared to selected-response testing (Ruiz-Primo, Schultz, Li, & Shavelson, 2001a; Ruiz-Primo, Schultz, Li, & Shavelson, 2001b). However, as informative as the literature is regarding validity of concept mapping as assessments as demonstrated by university based researchers in secondary classrooms (Ruiz-Primo, et al., 2001a; Ruiz-Primo, et al., 2001b, Shavelson, Ruiz-Primo, & Wiley, 2005; Yin, Vanides, Ruiz-Primo, Ayala, Shavelson, 2005), there is a lack of action research detailing the reality of using concept maps as assessments by secondary educators in their own classrooms. Additionally, no published literature could be found regarding the use of concept map assessments by inexperienced or pre-service science teachers for student assessment purposes. It is unclear whether the implementation of concept maps as alternative assessments can be successfully carried out by a novice teacher at the secondary level, and it is unclear whether a single student training lesson on concept map construction, as used in the research cited as part of this research project, is ideal for a novice teacher and students who will use a concept map to assess learning gains (Ruiz-Primo, et al., 2001a; Ruiz-Primo, et al., 2001b; Yin, et al., 2005).

This purpose of this study is to investigate the feasibility of the use of concept maps with high school students by a novice teacher. The study will address the following

research questions: 1) How feasible is it for a pre-service teacher to use concept maps to assess student understanding and what are the advantages or possible barriers of using this method? 2) Is a single lesson on concept mapping sufficient for students to prepare them for a unit test based on concept maps? If not, are there any benefits are observed when 3 additional concept map experiences are part of classroom instruction? The hypotheses tested were: novice teachers can use concept mapping as an effective assessment method and that students are likely to need multiple experiences creating concept maps in order to demonstrate their full understanding on a unit test.

The study used action research as a means of answering these questions. Action research is “teachers researching their own practice of teaching” (Feldman & Minstrell, 2000). According to Feldman and Minstrell, there are two purposes of action research: to improve teaching practices with the goal of improving student learning, and to become part of the “knowledge base of teaching and learning” (Feldman & Minstrell, 2000).

Subjects of the research included the researcher, who is a Masters’ degree student at a public urban university in the Pacific Northwest, and who was completing practicum field observations at a local high school in the 2010-2011 school year, and her students in two classes of Foundations of Physics and Chemistry. All students received the concept map training lesson described by Ruiz-Primo, Schultz, Li, and Shavelson (2001a) prior to starting a unit in astronomy, and all students completed a construct-a-map assessment question paired to a traditional format short answer question as part of the unit assessment test. During the unit instruction, half the classes received three additional concept mapping exposures in class, in order to evaluate whether the additional time

devoted to concept mapping was a beneficial use of time, as determined by scores and by student feedback. Throughout the study the teacher maintained notes on the value of concept mapping as a means of determining what students have learned at the end of a unit, as well as on the implementation challenges (such as scoring the tests), and modifications for next steps.

The independent variable was the extent of training for students on the concept map assessment tool and other modifications made by the novice teacher during the study. The dependent variable was the quality of student performances on concept map tests as measured by a rubric for scoring the tests.

## **Literature Review**

### **Overview of Literature Review**

The literature review will first briefly cover the history of testing in schools, from the early adoption of testing in the U.S. to current reform efforts based on No Child Left Behind. The next section focuses on the different types and functions of tests used in the modern classroom in the U.S. This is followed by a literature review on the use of alternative assessments in the classroom, including science notebooks, open-ended tests, hands-on tests, student portfolios, concept mapping, and student/teacher developed assessments. The final section of the literature review covers concept maps in detail, from the initial development of concept maps to the study of the use of concept maps as assessment tools, along with the features of the possible types of concept map assessments, as presented in the literature.

### **Brief Background on Assessment in the United States**

Klassen briefly described the history of testing in the US in Part 1 of a two-part review (2006). Given that examinations were in use in Europe as a means to quantify grades since about 1792, and the use of examinations for other purposes goes to 210 B.C. in China, the use of assessments in the United States began relatively recently in 1845 in Boston (Klassen, 2006). Horace Mann, a significant proponent for public education, administered written examinations to Boston teenagers, and from 1845 to 1900, essay tests became the norm for testing (Klassen, 2006). Mann argued that the testing results were reflections of the quality of teachers, and used poor test scores to drive reform policies regarding public education standards (Klassen, 2006). While exams in their

earliest forms served as public policy tools for accountability of teachers and schools, this did not continue to be the case from about 1920 to 1960, when assessment was primarily confined to use by classroom teachers to judge student performance and confer grades (Klassen, 2006).

The shift in 1960 back to using exams as accountability instruments to inform decisions about public policy was due to the growing concern that the U.S. was falling behind the Soviet Union in science and technological innovations (Bruner, 1977; Klassen, 2006). The success of the Soviet Sputnik satellite program inspired the 1959 “Woods Hole Conference” in Woods Hole, Massachusetts; the conference was held out of concern for the condition of American science education with the goal to improve the “quality of curriculum” in science education (Bruner, 1977, pg. 14). The building concern over the quality of science education due to the perceived advances in the Soviet Union was one important influence in science education reform in the US, resulting in increasing the number of exams administered to students. The affects of this policy shift continue as seen in the current emphasis on the use of testing for accountability measures of schools and teachers (Klassen, 2006).

In 2001, President Bush made accountability the focus of his education agenda (The White House, 2001), further promoting the view that testing is a necessary tool for gauging teachers’ effectiveness. This education reform to improve teachers and increase academic achievement resulted in the No Child Left Behind Act of 2001 (NCLB), which was signed into law in January 2002 (Public Law 107-110). NCLB specifies accountability measurements by schools, districts, and states, and mandated the use of

statewide learning objectives as the benchmark for measurement. Schools, and subgroups of students within the school (race-ethnicity, and economic background) are measured for Adequate Yearly Progress (AYP) towards meeting these learning objectives. These requirements have forced many states to revise testing practices by adding extensive statewide tests that were not previously required with ambitious performance goals that many states are finding difficult to meet (Linn, Baker, & Betebenner, 2002). Teachers are increasingly pressured by student performance on assessments, and spend increasingly more time “teaching to the test” and on test preparation (Kohn, 2000).

### **Assessments**

Classroom assessments serve multiple functions in today’s schools; they are used for diagnostic purposes, assigning grades, monitoring student progress, informing instructional and curricular decisions, and gauging the teacher’s own effectiveness of instruction (Dietel, 2004). The following literature review briefly describes a study of the testing effect, then provides information on selected response testing, which is the format most commonly used today, and concludes with a discussion of concept maps as an alternative and more authentic test format.

Kuo and Hirshman investigated a new method of experimental testing to better characterize the testing effect (Kuo & Hirshman, 1996). The testing effect is the idea that when recalling an item from memory in a testing environment, there is a higher probability that it will be recalled than in a non-testing environment. The purpose of the research was to determine whether a new method of studying testing effects could isolate the learning effects of taking a test, using two experimental protocols. The first

experiment used 64 undergraduate students with 16 testing conditions, using word recall as the subject. Each condition had a unique set of five experimental trials followed by a “test”; the experimental conditions differed in the number and order of different “study” (S) and “test” (T) trials. For example, condition 5 was “SSTSST” and condition 16 was “STTTTT.” During the S trials students were asked to study the words and during the T they were told to study the words then recall them. Each trial had a “retention interval” during which the subjects were asked to complete unrelated tasks. The results from Experiment 1 showed that test trials enhance memory as compared to study trials and that testing effects are seen even when students are not given an opportunity to restudy material. To rule out “interitem association,” which is a retrieval benefit in a series of unrelated words, the researchers designed Experiment 2 using single items for recall. Experiment 2 had 18 subjects and used single words during each trial, but followed the same procedure as Experiment 1, with small changes. The results demonstrated significant testing effects seen for the single-item recall, suggesting that inter-item association was not a significant factor in the observed testing effect, and it was the test condition itself that produces the increase in recall observed.

The preferred test format in the U.S. continues to be the traditional selected-response test (Neill & Medina, 1989). The reasons teachers favor this format vary, but the greatest reason for the popularity for selected-response testing has to do with ease of scoring (Klassen, 2006; Roediger & Marsh, 2005). Selected-response tests include the ubiquitous multiple-choice format, but also include matching, fill-in-the-blanks, and true-false formats (Klassen, 2006). It is not a coincidence that the invention of digital scoring

methods was accompanied by an increase in the use of selected-response tests (Klassen, 2006). Teacher time required for scoring remains an important consideration when selecting or constructing an assessment (Roediger & Marsh, 2005; Yin, et al., 2005).

The positive and negative effects seen in multiple-choice testing are presented in a study by Roediger and Marsh (2005). This study of 24 undergraduate college students consisted of a multiple-choice test followed by a cued-recall test, with test questions covering 36 different passages. The purpose of the study was to evaluate the relationship between the testing effect (the idea that taking a test will improve later performance on a test) and interference effects of multiple-choice questions (the misinformation lure effect), as well as how the number of multiple-choice options related to the outcome. To duplicate a testing situation where a student has not prepared or read the material before a test, half of the students were given the first 18 passages and the rest of the students read the second 18 passages, though all students completed questions for all passages. In the first test, students were advised to provide an answer to every question, but in the cued-recall test, participants were advised to not guess.

The results showed that as the number of possible answers increased, student performance decreased depending on whether students had read the passages (Roediger & Marsh, 2005). Questions with 4 and 6 alternative answers had significant lower testing effects observed compared to questions with 2 alternatives. Despite warnings against guessing, researchers found that on the cued-recall test, participants chose wrong answers in proportion to the number of alternatives offered, without regard to whether they read the passage or not. For wrong choices selected on the cued-recall test, 75% had been



selected on the initial test, showing that subjects were choosing incorrect lures as correct answers. However, subjects answered more questions correctly on the recall test when they had been tested previously on a multiple-choice test relative to when they had not, an example of the positive effect of testing. The researchers concluded that taking a multiple-choice test causes subjects to answer questions on a cued-recall test incorrectly, with a greater effect observed in the questions corresponding to passages the participants did not read (designed to mimic a poorly prepared student taking an exam) and when the number of alternatives was higher on the initial test. Though there was a positive testing effect seen when the participants answered the questions correctly in the initial test, and the researchers note that the mechanism of reading multiple incorrect answers may help students retrieve the correct answer and recall it later, this mechanism can create misinformation, especially when many possible answers are presented as alternatives. The researchers concluded that educators need to consider both the positive and negative effects of multiple-choice testing, and that further research is needed to investigate multiple-choice testing in situations that closely imitate educational settings.

In science education, the reliance on selected-response tests is especially problematic, as the cognitive goals for teaching and learning science should include different types of knowledge that cannot be measured using traditional tests (Shavelson, Ruiz-Primo, & Wiley, 2005). Shavelson, Ruiz-Primo, and Wiley (2005) describe four discrete types of knowledge in science: declarative, procedural, schematic, and structural. Declarative knowledge involves “knowing that,” and is associated with factual and conceptual knowledge. Procedural knowledge is “knowing how,” and involves being

able to carry out steps to answer some kind of question. “Knowing why” is schematic knowledge, and is a result of being able to access long term memory and working memory in order to address a question. Strategic knowledge is “knowing when,” and is characterized by the ability to discern when, where, and how to apply the appropriate knowledge. Selecting an assessment tool for use in the science classroom requires consideration of what knowledge type is being assessed (Shavelson, et al., 2005).

Part of the assessment selection process should include an awareness of the difference in measuring the extent of knowledge and the structure of knowledge, and there are various forms of assessments that have been shown to measure the different types of knowledge effectively for extent and structure (Shavelson, et al., 2005). For example, hands-on performance assessments and lab notebooks have been shown to be accurate in measuring the extent of procedural knowledge, and procedural maps are effective measures of the structure of procedural knowledge (Ruiz-Primo, 2005; Shavelson, et al., 2005). There are also examples of alternative assessments that are not measuring any knowledge gains (Waters, Smeaton, & Burns, 2004). Research has demonstrated that selected-response tests do measure the extent of students’ declarative knowledge, but they do not measure the structure of the knowledge (Shavelson, et al. 2005). Concept maps have been shown to be an effective and valid assessment technique for measuring the structure of conceptual knowledge, depending on the mapping technique selected (Ruiz-Primo et al. 2001a; Ruiz-Primo et al. 2001b; Shavelson, et al., 2005; Yin, et al. 2005). Because of the already overwhelming number of tests students are subjected to, it is crucial that science teachers select assessments with care, to provide

meaningful information about learning gains (Kohn, 2000; Shavelson, et al., 2005; Wiggins, 1993).

In summary, test taking can increase memory retention as a result of the “testing effect” (Kuo & Hirshman, 1996). However, the testing effect is largely annulled by the high negative correlation between the number of alternatives provided on a selected-response test and the benefit to the student, as well as an increased probability that the misinformation will serve as “lures” to create false knowledge on follow-up recall tests (Roediger & Marsh, 2005). Finally, the use of selected-response tests in science is problematic because these tests only measure the extent of declarative knowledge and do not consider the four different types of knowledge in science; they also do not distinguish between assessing extent and structure of knowledge (Shavelson, et al., 2005). Some alternative assessments can be used to measure different types of knowledge in science (Shavelson, et al., 2005). The next section of the literature review will review some different forms of alternative assessments.

### **Alternative Assessments**

Alternative assessments have become more popular in the last 30 years, as educators began to apply theories of learning to assessment techniques and more research on alternative assessment validated their use in the classroom (Klassen, 2006). There are many types of alternative assessments possible that have been researched in science classrooms (Klassen, 2006). One form of alternative assessment that would potentially show both conceptual and procedural understanding is evaluation of students’ science notebooks (Shavelson, Ruiz-Primo, & Wiley, 2005). Ruiz-Primo presents the results of a

study that aimed to create a framework for using science notebooks as an assessment as well as a tool for scoring those notebooks (Ruiz-Primo, 2005). This study involved six classrooms out of 20 fifth grade classrooms that were participating in a larger study involving science curriculum reform. A total of 72 notebooks were used in this study, and were collected after completion of a unit taught as a part of the main study. Pre and post assessment scores were determined as part of the main study and the scores were added to the data for the science notebook study. Science notebooks were scored using a criteria table and notebook scoring form, which were created based on the specific activities of the unit that was taught, with five types of scores gathered for each notebook: unit implementation, quality of communication, conceptual understanding, procedural understanding, and teacher feedback.

Ruiz-Primo (2005) found that student notebooks could be scored reliably, and there was a high degree of correlation between the notebook scores and the other assessment scores. She also found that student communication and understanding did not improve over the course of the unit, and teacher feedback was not present in notebooks. Additionally, the student writing in notebooks was described as “mechanical,” lacked explanations and conclusions, and demonstrated that teachers were not using notebooks in an efficient way to improve student learning and understanding and lacked alignment to unit goals. The researcher concluded that if notebooks are going to be used as an assessment, teachers need to ensure that the entries have purpose, contain teacher feedback, and coordinate the goals of improving understanding and inquiry.

Another study evaluating various alternative assessment formats was a large side-by-side comparison research study performed by Lawrenz, Huffman, and Welch (2000); this study looked at student achievement using various alternative assessments compared to traditional multiple-choice tests for 3500 ninth grade students at 13 high schools across the United States. The schools were selected as a result of using curriculum linked to the NRC Science Standards and teachers were all part of a teacher enhancement project associated with the development of the curriculum. The researchers used a multiple-choice test, a written open-ended test, a hands-on laboratory content test, and a hands-on full investigation test in their evaluation of assessments. All students completed the multiple-choice test and students were randomly selected at each school for completion of the other assessments. Additional data were collected on ethnicity and gender, and students self-reported typical course grades. The researchers found that students performed well on all assessments, but least well on the open-ended test and best on the full investigation. Significant differences were seen among the various assessment formats when data were analyzed by ethnicity but not by gender. Student reported typical grades were predictive of outcomes on all assessments for all groups, and correlations were explored between the various assessments and typical grades.

Lawrenz, Huffman, and Welch found that the strongest correlations were between performance on the multiple-choice test and the open-ended test, with correlations weaker as typical grade decreased (2000). The authors note that these findings are in agreement and disagreement with prior research. They suggest that based on differing performance by ethnic group and by achievement group on the various assessments, these

different tests are measuring different things and future research is needed. The better performance by African American students on the full investigation test is of particular interest to the researchers due to the traditionally large gap between Caucasian and African American students on tests, which was seen in all test formats here other than the full investigation. Also of interest to the researchers was the diminished gap in low and moderate performing students observed on the hands-on test, suggestive of some increased motivation to do well compared to a traditional test. The authors raise questions about teacher responsibility to offer assessments that allow all students to demonstrate knowledge, and since it seems that hands-on testing does that best here, conclude that perhaps more hands-on testing should be used in the classroom.

In an extensive historical review paper on assessment in science education by Klassen, three formats of alternative assessments are reviewed (2006). His review for each assessment is detailed and includes a description of the assessment, the basis for it, research using it, followed by author opinion regarding the assessment. The first assessment discussed is the concept map assessment. Research by Ziilik et al. are cited as an example of the concept map assessment; they used a concept map and a multiple-choice assessment as a pre- and post-assessment tool to evaluate conceptual understanding and found that the concept map and multiple-choice test had high reliability and moderate correlation, and the concept map was able to show a high degree of improvement in conceptual understanding. The author supports the use of concept mapping as an assessment. Performance assessments are described and research by Erikson and Meyer is discussed, in which students were asked to draw conclusions about

pitch and tension in guitar strings. The research findings regarding performance assessment are unfavorable according to Klassen: research on performance assessment has shown that performance assessments do not address the shortcomings of deconceptualized assessments. The final version of assessment discussed is the portfolio assessment. Slater et al. used the portfolio assessment at a community college to show that portfolio assessment can be used instead of conventional assessments. Research on portfolio assessments is positive, but incomplete, and there are concerns about the use of the portfolio as assessment given the opportunity for fraud and plagiarism. Klassen briefly discusses the need for validity and reliability in assessment, and specifically issues for these three assessment types reviewed. Problems include a lack of standardization in concept maps, reliability in portfolios, and a lack of validity in performance assessments. Klassen concludes his paper with recommendations regarding science assessments, which include use of validated conceptual assessments, replace numerical grading with letter only grading, and use of the school as the level for achievement and not the student.

In one of the only action research studies found investigating alternative assessment, Waters, Smeaton, and Burns (2004) present the findings of one teacher with 17 years of teaching experience and her efforts to implement differentiated, alternative assessments in place of selected-response tests in her secondary science classroom. This research used survey tools to examine student reaction to the changes as part of an action research study. The students in this study are described as a “diverse mix” of students attending a rural school of 3250 students, with some recently relocated from urban areas. The study involved 79 ninth graders enrolled in the teacher’s earth science class. The

teacher created a rubric for the new assessments, shared this with the students, promoted discussion of the new assessments, and solicited student responses. Students were then allowed to choose an assessment activity, whether they wanted to work alone or in a group, and decide on the product they would create for the assessment. A variety of options were given, including board game creation, performance, models, newspapers, creative or formal writing, computer presentations, and web pages, or combinations of these.

Students were required to complete assessments in 3 different formats throughout the semester and were given 180 hours of class time per assessment for completion (Waters, Smeaton, & Burns, 2004). Students met periodically for group time and peer comments, and presented their projects formally to the class and teacher, then submitted materials to the teacher for evaluation. At the end of the semester, the students completed a 24-question survey with a Likert scale response for students to provide opinions on the new assessments. There was also an open ended 2-question questionnaire. The results showed that this group of students preferred the alternative assessments to multiple-choice tests, and most enjoyed the freedom allowed in the alternative format. The teacher noted that enthusiasm for the alternative assessments seemed higher in the classroom, and the classroom tone was friendly and collegial. While students self-reported learning more with the alternative assessments, the teacher noted that this was not measured in the study.

In summary, there are many different types of alternative assessments that have been researched in science classrooms. Student notebooks can be scored reliably, but to



use them as an assessment tool would require extensive effort by the teacher to improve the quality of entries by the students by offering constructive feedback (Ruiz-Primo, 2005). Hands-on investigation assessments show promise for reducing traditional gaps in scores seen by ethnicity, may increase student motivation, but also need more research in the classroom setting (Lawrenz, Huffman, & Welch, 2000). Concept mapping was determined to be a viable alternative assessment format when reviewed alongside portfolio assessments and performance assessments (Klassen, 2006). Finally, an action research study demonstrated that the successful introduction of alternative assessment in the high school classroom is possible, though it also demonstrates the importance of remaining aware of the purpose of classroom assessment (Waters, Smeaton, & Burns, 2004). Based on the viability of concept mapping as a form of alternative assessment presented in the initial literature review and the perceived simplicity of implementation compared to some alternative assessment formats discussed previously, concept maps became the focus of this research project. The next section of the literature review examines the use of concept maps in the classroom as assessment tools.

### **Concept Maps**

Concept maps are two-dimensional visual diagrams that represent a set of concepts (Novak & Gowin, 1984). The “map” is formed with nodes, each containing one concept, usually organized in a hierarchy. The nodes are linked together with labeled lines to form propositions between the ideas. Dotted lines can be used to bridge branches of the map to suggest new relationships (Mintzes, Wandersee, & Novak, 1998). See Figure 1 for an example concept map of water.

Concept mapping was introduced in 1984 by Novak and Gowin as a way to increase meaningful learning in science by guiding instructional planning, pre-assessing student knowledge and identifying misconceptions. They also suggested it could be used as an evaluation tool. Novak and Gowin recognized that traditional testing in science was not a valid measure of student understanding and also noted that it allowed misconceptions to go unnoticed. They proposed that alternative assessment tools should be considered for use in the classroom and that concept maps would be practical

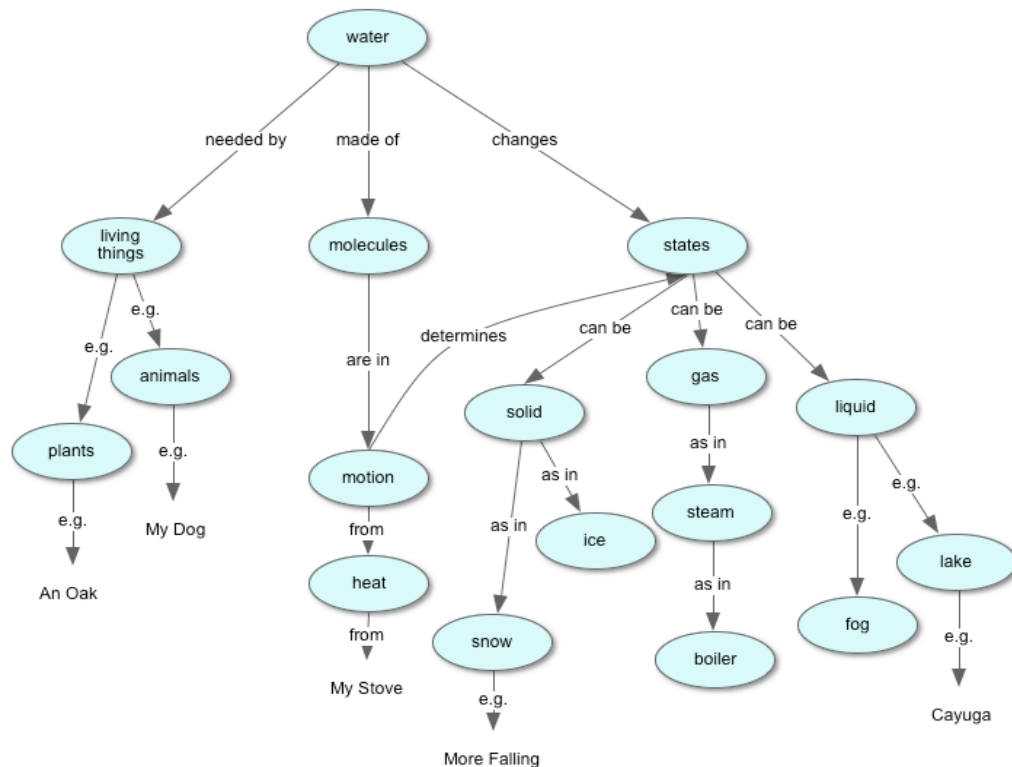


Figure 1: Example of a concept map for water (Novak & Gowin, 1984)

alternatives because completion of a concept map demonstrates an understanding of interrelatedness by requiring an ability to organize and relate ideas in a subordinate manner (Novak & Gowin, 1984).

The validity of concept mapping as an assessment technique was investigated by Ruiz-Primo, Schultz, Li, and Shavelson (2001a). The purpose of the research was to examine whether the mapping technique employed would result in differences in validity of interpreting student understanding and inferred cognitive activities. The study consisted of 152 high school students, and they were asked to complete a series of concept maps on the topic “Elements, Atoms, and Compounds.” The first task was to create a concept map from scratch (“construct-a-map”) given only a set of concepts to use, the second task they were asked to fill in a blank node map (“fill-in-the-node” map), then the third task was to fill in the blank linking lines between nodes (“fill-in-the-lines”

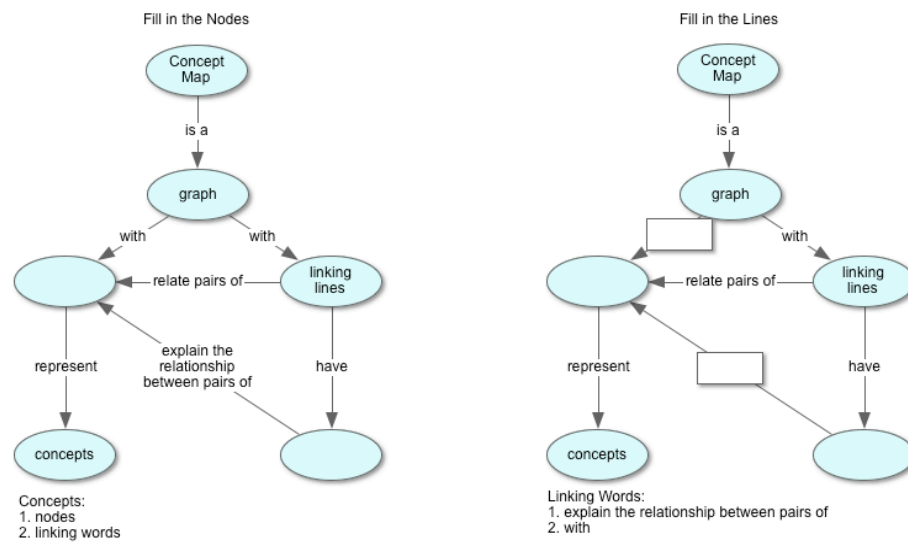


Figure 2: Examples of Fill-in-the-node and Fill-in-the-lines concept maps (Shavelson, et al. 2005) map). See Figure 2 for examples. Students were provided with either of two incomplete lists of concepts for completion of the latter two maps, which revealed that providing students with half the concepts facilitated the student generating the other half for

completion of the map. The researchers found that the completion of “skeleton maps,” either fill-in-the-node or fill-in-the-line, were equivalent, though the fill-in-the-node completion was easier for students, resulting in the researchers compiling the data into one unit for “fill-in-the-map.” The fill-in-the-map scores were overall very high, suggesting a high level of understanding by the students. However, the construct-a-map scores showed a much different performance outcome, with lower scores, indicating only partial knowledge. The researchers concluded that because there was not a high correlation between the scores using these techniques, there were different aspects of understanding being demonstrated, with the construct-a-map technique better reflecting students’ understanding.

To determine whether the concept mapping techniques were valid representations of understanding, the researchers followed up the previous study with a small cognitive validity study (Ruiz-Primo, Schultz, Li, & Shavelson, 2001b). In this study, the researchers enrolled 6 students and two teachers from the previous study; based on concept mapping scores from the previous study the researchers selected 3 students designated as “high” and 3 that were “low.” Students and teachers were asked to complete a concept map for a topic using three mapping techniques: fill-in-the-node, fill-in-the-line, and construct a full map. They also were asked to answer multiple-choice questions. To evaluate inferred cognitive activities, the subjects were instructed to “think-outloud” and researchers audio-recorded the verbalizations of the students and teachers as they completed the tasks. They were then asked to describe the strategies used and decisions made in the completion of the task. The maps were graded and the

verbal transcripts were coded for types of explanations and monitoring strategies used in order to quantitatively analyze the results.

Ruiz-Primo, Schultz, Li, and Shavelson found that using a fill-in-the-node concept map resulted in a skewed result, appearing that all students have high knowledge, while the construct a map technique revealed that the students only had partial knowledge, and the fill-in-the-lines result had a high correlation to the construct a map result, which was unexpected (2001b). The researchers hypothesized that this was an effect of a small sample size and required further research. The analysis of the verbalizations for cognitive activities showed that the construct a map technique was accompanied by significantly more “explanatory” comments than the other formats. The researchers concluded that it was because the technique required more knowledge, so it allowed the subjects more opportunities to demonstrate knowledge but also required more “self-checking” activities, or monitoring statements. The fill-in-the-node format had the most “monitoring” verbalizations, with the researchers suggesting this was a result of the subjects being more aware of right and wrong answers in this format. The results of the map scores and the verbal codings were compared, with the researchers finding strong correlations between the “explanation” type statements and the construct-a-map technique. The highest degree of conceptual errors correlated strongly to the fill-in-the-node format. The researchers concluded that the construct-a-map technique had the best correlation with the cognitive activities inferred and the scores obtained.

In a follow-up study reported in 2005, Yin et al. further explored the differences between concept mapping techniques. In this study, the two techniques evaluated were

the construct-a-map with created linking phrases (coded “C”) and construct-a-map with selected linking phrases (“S”) with a goal of showing conceptual and empirical evidence of the characteristics of the techniques to provide a valid comparison of the two techniques. The study participants were 92 eighth graders from 6 classes taught by the same teacher. The students were ranked as low, medium, or high performing by their teacher, based on previous scores and teacher observation. The design used was a 4x2 design, with students randomly assigned to one of 4 sequences: CC, SS, SC, and CS. The concept mapping tasks were based on an established curriculum on buoyancy, with the concepts to be included selected by a team of researchers, assessment experts, and a master teacher. For both mapping tasks, students were provided with nine concepts related to buoyancy, and instructed to connect the concepts with one-way arrows. Students then either created their own phrases describing the relationship in the C task, or used a phrase from a list provided in the S task. Think-aloud protocols were video taped for four students for later analysis. To score the maps, the unique propositions from students were collected into a database and points were assigned to the propositions, based on a 4 point scale, with scientific intent favored over grammar or wording, and partial credit for recognizing relationships even if they were not fully correct. Once point values were assigned, 3 experienced raters evaluated the maps and assigned scores.

The data show that there was a testing effect in the SS and CC groups, and a format effect in the SC and CS groups; the formats are not equivalent. To analyze the data without any learning effects, the results were grouped by test format for the first occasion only. The researchers found the C scores were significantly lower than the S

scores, despite similar variances within each data set. The correlation between the test scores and the outside criterion (the teacher's assessment of low, medium, or high) was higher for the S format than the C. Though the difference between the two correlations was not significant, this was also an indication that the two formats were not equivalent. Analysis of the propositions showed that there were higher numbers of partial scores in the C format, indicating that with the S format students were limited by the number linking-phrase options. Another analysis using the propositions showed that there were nine propositions that were "popular" (constructed by over 50% in either condition) in the C group not found in the S group, though there were only five in the S and not the C. This result showed that the S group may have received some benefit with selection of propositions for more difficult concepts, but also were limited by offered phrases. Students created maps that were more complex, of a network structure, significantly more in the C technique than the S technique, indicating that the C technique allows more freedom for students to demonstrate knowledge. The think-aloud video analysis showed that the S condition required an extra "checking" step, when the C condition allowed for more spontaneous proposition building, showing that different cognitive processes were going on during the different techniques. The researchers concluded that though the C and S conditions are not equivalent, and the C condition is the "benchmark" technique for concept mapping, because the scoring was much more difficult for the C condition, the S condition might be better suited to assessments. The researchers surmise that the S condition could be improved by including linking phrase possibilities from other students' C maps for selection and by ensuring that the phrases were understandable.

In summary, research has demonstrated that concept maps can be used as assessments of structural conceptual knowledge in secondary science. Research into the implementation of concept map assessments, or mapping techniques, has demonstrated that not all forms will work equally well for assessing structural conceptual knowledge. The easiest mapping technique to grade and the most basic form of a concept map assessment is the “fill-in-the-nodes” map, when the student is given a blank concept map and a list of possibilities for completion (Ruiz-Primo, Schultz, Li, & Shavelson, 2001a). However, this format generally results in misleadingly high scores. Further evaluation has shown that students with a high score on a fill-in-the-node map can have a low conceptual structural understanding (Ruiz-Primo, Schultz, Li, & Shavelson, 2001b). The technique of concept map assessment that has the highest correlation between score and conceptual knowledge, and is considered the “gold standard,” is the “construct-a-map” format; in this assessment the student must complete a concept map from scratch using a list of required concepts (Yin, et al., 2005). Scoring individual students’ maps will reveal a great deal about their understanding and misconceptions, but the difficulty of scoring these maps renders the construct-a-map format impractical for use in the classroom as an assessment. Research suggests the best type of concept map for assessment is the modified construct-a-map, where students are offered a set of propositions and concepts to use in constructing a concept map (Yin, et al., 2005).



## Methods

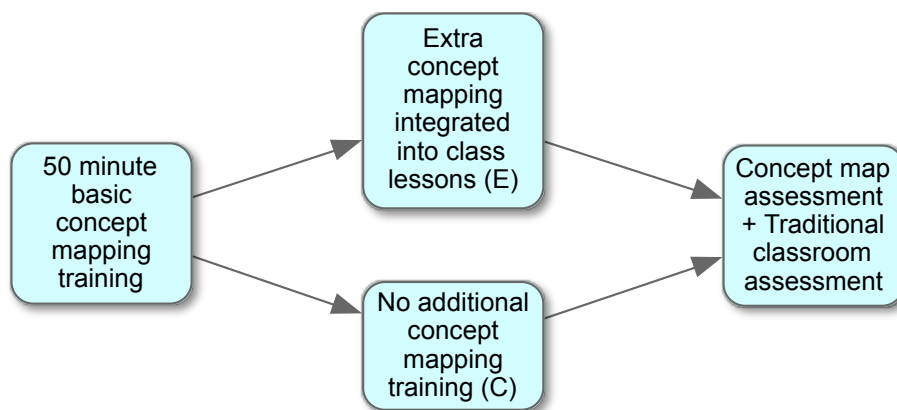
### Overview

This was an action research study to answer the following research questions: 1) How feasible is it for a pre-service teacher to use concept maps to assess student understanding and what are the advantages or possible barriers of using this method? 2) Is a single lesson on concept mapping sufficient for students to prepare them for a unit test based on concept maps? If not, are there any benefits are observed when 3 additional concept map experiences are part of classroom instruction?

Action research, or teacher-based research, is historically not grouped with traditional educational research conducted by university-based researchers (Cochran-Smith & Lytle, 1990). According to Cochran-Smith and Lytle, research topics of interest tend to diverge between teacher research and university research, resulting in a segregation in the content, ownership, and structure of research. There is a lack of structural and financial support for teacher research when compared with university research, and most teacher research is performed outside the university setting. Another cause for the division, as related by Cochran-Smith and Lytle (1990) is the methodology used in action research, as seen in the differences in the development of research questions, theoretical frameworks, and generalizability. The research questions in action research arise from the day-to-day experiences of teachers and not from educational theory, as in university research. The limited generalizability of action research has been used to discount the value of the research method. Although generalizability may be limited, there is value in the findings of classroom teachers in actual teaching situations,

which can be defended as more authentic than decontextualized studies conducted by university researchers (Feldman & Minstrell, 2000).

With the limited generalizability in mind, action research was employed in this study to explore the research questions regarding the use of concept mapping by a pre-service teacher. Subjects of this research study included the researcher and her students. All students received a concept map lesson described by Ruiz-Primo, Schultz, Li, and Shavelson (2001a), and completed a construct-a-map assessment question at the end of a unit that was paired with a corresponding traditional test short-answer question. Half the classes received three additional concept mapping exposures in class, in order to evaluate whether the additional exposures to concept mapping were a beneficial use of time, as determined by scores and by student feedback. The two conditions in this study were divided by classroom; one class received extra training (“E” condition) on concept mapping and the other class was the received on the initial concept mapping lesson (“C”



*Figure 3: Study Flow Chart Showing E and C Conditions*

condition). See Figure 3 for a visual representation of the study.

Data collected in the study was both qualitative and quantitative in nature.

Throughout the study the teacher researcher maintained notes on the implementation of concept mapping in the classroom, time requirements, as well as on the challenges (such as scoring the tests), and modifications for next steps. Student surveys were administered post-test to discern student opinions on the test, concept mapping in general, and feedback related to concept mapping instruction. Teacher journaling reflections were used to determine whether ease of implementation changed over time, if the time needed to design and score concept map questions was prohibitively high, and whether the additional class time spent practicing concept maps yielded results other than any observed in the test scores.

### **Participants**

Study participants included the researcher, who is a Masters' degree student at a public urban university in the Pacific Northwest and was continuing practicum field observations in the 2010-2011 school year, and her students who were recruited from two Foundations in Physics and Chemistry (FPC) classes at "A" High School<sup>1</sup> (AHS), a public high school near the downtown of a moderate sized city in the Pacific Northwest.

AHS is adjacent to the urban hub of the downtown area and serves a higher income area of the city, as measured by average property values. Enrollment at AHS is about 1400 students out of 11,000 high school students in the district based on information from the school district. While there is limited ethnic diversity at AHS compared to the other high schools in the school district (see Table 1), the ethnic ratios

---

<sup>1</sup> "A" High School has been changed by the researcher to maintain confidentiality.

seen at AHS are largely in line with that of the city, as reported in the 2010 U.S. Census Data (see Table 2), with a slightly lower percentage of Hispanics and African Americans at AHS than one would expect given the city's statistics (United States Census Data, 2009).

*Table 1. Ethnicity of student populations at AHS and high schools in the district*

	Asian	African American	Hispanic	Native American	White	Multiple Races
<b>AHS</b>	8.7%	3.8%	7.7%	0.4%	75.3%	4.2%
<b>District</b>	10%	15%	13%	1%	57%	4%

*Table 2. Ethnicity of city and state where AHS is located (United States Census Data, 2009).*

	Asian	African American	Hispanic	American Indian, Alaskan Native	Native Hawaiian, Pacific Islander	White, not Hispanic	Multiple Races
<b>City</b>	6.3%	6.6%	6.8%	1.1%	0.4%	77.9%	4.1%
<b>State</b>	3.0%	1.6%	8.0%	1.3%	0.2%	86.6%	3.1%

As would be expected at a high school serving a higher average income level than the other schools, AHS has the lowest rate of students receiving Free and Reduced (FRL) lunch of all of the high schools in the public school district serving the city. AHS has 13% of students receiving FRL compared to a district average of 40%. AHS also has a lower percentage of English Language Learners (ELL) compared to the other high

*Table 3. Public high school characteristics at AHS and high schools in the district.*

	SPED	TAG	FRL	ELL	Total Students
<b>AHS</b>	4%	27%	13%	1%	1410
<b>District</b>	12%	19%	40%	5%	10,978

schools in the district, with 1.1% of AHS students classified as ELL while 5% of all

students in the district are ELL. 12% of the district students receive Special Education (SPED) instruction; 4.3% of the students at AHS are SPED. About about 27% of the student body at AHS is designated as Talented and Gifted (TAG) compared to 19% at the district level.

AHS holds classes on a modified block schedule with 7 classes, or “periods”,

*Table 4. AHS Bell Schedule. Table shows the bell schedule on days where all classes meet (“A”) and block days (“B”), depending on the day.*

Monday/Tuesday/Friday “A” Schedule		Wednesday “B” Schedule		Thursday “B” Schedule	
Period	Time	Period	Time	Period	Time
1	8:10-9:00	2	8:10-9:45	1	8:10-9:45
2	9:05-10:00	Flex	9:50-10:35	Break	9:45-9:55
Break	10:00-10:10	Break	10:35-10:45	3	10:00-11:30
3	10:15-11:05	4	10:50-12:20	Lunch	11:30-12:05
4	11:10-12:00	Lunch	12:20-12:55	5	12:10-1:40
Lunch	12:00-12:35	6	1:00-2:30	7	1:45-3:15
5	12:40-1:30	Flex	2:35-3:20		
6	1:35-2:25				
7	2:30-3:20				

offered (see Table 4). The school week has days designated as “A” days, where all 7 periods meet for 50 minutes each on Monday, Tuesday, and Friday, and “B” days, which are block days and periods meet for 95 minutes. On Wednesday “B” days, periods 2,4,and 6 meet, along with 2 “Flex” periods that provide students freedom to work in classes of their choice. Thursday’s “B” schedule consists of periods 1, 3, 5, and 7.

The cooperating classroom teacher for this research study is a veteran teacher with a background in Physics, Biology, and Chemistry and a M.Ed. She has 14 years of

high school instructional experience, all of which has been at AHS. She is currently teaching 3 classes of FPC and 2 classes of Physics. As a master teacher with years of experience, the cooperating classroom teacher's opinion was solicited and her ideas were implemented by the researcher at several stages during the research project.

Two classes were selected to participate in the research project based on the researcher's practicum experience with the 2 classes during the Fall 2010 school term. All students in these classes were invited to participate in the study, with parental and student consent required prior to participation. While the average class size at AHS is 29.3 students, higher than the district average of 25.5, both of the classes in this study had higher enrollment than the AHS average: one class had 34 and the other had 31. Of the 65 students eligible for enrollment, 43 students enrolled in the study. Of the 43 students who enrolled, data from only 35 students was included in the study due to several student absences on the day of testing resulting in 17 students in the E condition class and 18 students in the C condition class completing the study. Class demographic data was not collected due to additional consent requirements of the school district, but the gender ratio was close to even, as would be anticipated in a typical public high school. Ethnicity of the students in the two classes is assumed to reflect that of the overall student population at AHS as was previously described.

The decision about which class to assign to which condition was made in consultation with the cooperating classroom teacher. When generalizing the overall class characteristics, one class was considered "motivated" while the other was "challenging," as described by the cooperating classroom teacher. The class considered to be

challenging was selected for the E condition. The rationale for providing the more challenging class with additional concept map experience was based on the idea that if the more motivated class received additional mapping experience, any differences seen between the two classes would be magnified by the fact that the class receiving extra instruction was already higher performing, as a whole. Additionally, it was the opinion of the cooperating classroom teacher that extra exposure to class material, through the concept mapping practice, would benefit the challenging class more and that the motivated class did not need the extra review as much. The selection of the more challenging class for the E class was made with the acknowledgement that the composition of the class might result in diminished observable differences.

### **Treatment**

This research project was designed to be inserted into a regular classroom curriculum unit with minimal disruption to the planned unit, so that the C instructional activities would remain as close a possible to what is normally taught in the astronomy unit and the E activities would complement the lessons without adding significant conceptual instruction. See Appendix A for copies of the instructional materials used in this unit, including the concept map lesson developed by the researcher.

During the first day of the new curriculum unit for astronomy, all students received a lesson on concept mapping, as described in Ruiz-Primo, et al. (2001a). The lesson consisted of a 50 minute training divided into four parts in one class session. Because of the familiarity of the students with PowerPoint delivery of instruction, a PowerPoint slide show was used for the concept mapping lesson. The first part of the

lesson was the introduction of concept maps, including what they are, what they are used for, what the parts of a concept map are, and examples of different maps. The second part of the lesson emphasized the construction of concept maps, concentrating on the relationships between concepts, recognizing what a good map looks like, and revising a map to improve it. Then the students worked as a group to construct a map using a familiar concept, “seasons.” The third part of the lesson required students to construct a map individually using familiar concepts. Based on the opinion of the cooperating classroom teacher, the concept selected for this segment of the lesson was “Planets of the Solar System.” It was her opinion that the students would have learned about the planets in middle school, would be successful with the map, and it would serve as a good introduction to the new unit of astronomy. The fourth part of the lesson was a class discussion about concept maps to address any student questions. The effectiveness of this method of instructing students on concept mapping has been established in multiple studies by evaluating the use of concepts from the list, use of labeled links, and the accuracy of the propositions (Ruiz-Primo, et al., 2001a).

Class E received three additional concept mapping activities as part of the research project. The number of extra concept mapping exposures was selected based on the amount of material being taught and the short length of the unit, which covered 2 chapters in the textbook and took 2 weeks. The lessons for the C class did not have additional instruction beyond the planned classroom activities. Table 5 depicts the overall unit outline for the instructional unit on astronomy taking place over a 2 week time frame in April.



The cooperating classroom teacher taught the unit material as normal for C and E classes; the researcher taught the introductory concept mapping lesson for C and E classes and provided the additional concept mapping instruction to class E.

*Table 5: Overview of Astronomy Unit Lessons. C and E class lessons are identical except for days noted with an asterisk (\*), which indicate an additional concept map exposure for the E class. 95-minute block days are indicated with [B].*

Unit Day	Date	Lesson
1	4/13	Concept Map Lesson, Hubble Telescope Computer Lab Activity [B]
2	4/15	Big Bang Theory, Origin of the Universe
3*	4/18	Life Cycle of Stars <i>*E: Galaxy Concept Map, partners</i>
4	4/19	Origin of the Solar System
5*	4/20	Moon & Phases of the Moon [B] <i>*E: Star Life Cycle Concept Map, groups</i>
6	4/22	“Mt. Nose” Activity for Moon Phases
7*	4/25	Eclipses, Asteroid Impact Activity in Computer Lab <i>*E: Moon, Sun, Earth, Time Concept Map, individual</i>
8	4/26	Class review of worksheet, Jeopardy
9	4/27	Test [B]

## Instruments

See Appendix A for copies of all instruments described here.

### Concept map lessons.

Additional concept map lessons were created by the researcher using Inspiration 9 concept mapping software. After reviewing the cooperating classroom teacher’s unit plan, curriculum materials, PowerPoint slides for class lectures (Appendix A), and the corresponding chapters in the class text book (Chapters 25 and 26 in *Physical Science: Concepts in Action*, by Wysession, Frank, and Yancopoulos), the main concept mapping lesson for both classes and the additional concept mapping activities for class E were planned.

To develop the extra concept maps for the E curriculum, the researcher reviewed the content of the each planned lesson and selected possible topics to cover using concept maps based on reference concept maps located online and topics that seemed to lend themselves well to concept mapping techniques. The final selection of concept map topics was made with the cooperating classroom teacher based on the instructional importance of the idea in the classroom; for example, the star systems translated to a good concept map, but they were not considered important by the cooperating classroom teacher, so it was not selected for a concept mapping lesson. The selected additional concept maps for the E class were: galaxies, the life cycle of stars, and the relationship between the Earth, moon, sun, and time. The concept mapping instructional tools were designed by the researcher, with input from the cooperating classroom teacher regarding content and difficulty. Activities were designed to employ a variety of teaching techniques to reach students of all learning styles. Individual, group, and partner concept mapping lessons were included.

### **Tests.**

The concept map test question was created to pair with a question on the teacher-made unit test. The traditional test contained a series of recall/identification questions and short conceptual problems, requiring a one or two word answer from the student. An example conceptual problem follows: “The phase of the moon that should be halfway between its highest point and setting at 9:00 am,” requiring the student to work out the answer using the techniques taught in class (the answer is “3rd Quarter Moon”). An example recall/identification problem: “A large, glowing ball of gas in space fueled by

nuclear fusion in its core, requiring the student to recall and identify the correct item being described (the answer is “Star”). The test also contained a series of short answer questions requiring a brief written response, which is the question format the concept map question was paired with. An example short answer question: “Discuss the nebular model for the formation of the solar system and show how it explains 4 important characteristics of our solar system.” For the short answer section of the test, students choose 4 short answer questions to complete out of the 8 offered. So as to not interfere with the format of offering the student the choice of question, it was decided to offer 3 concept map questions and allow students to choose one that they had already completed a paired short answer response. In order to encourage all students to complete the traditional question and the concept map to the best of their abilities, the higher score on the question pair was awarded for the overall test grade.

The three concept map questions for the test were selected from the eight short answer test questions based on which concepts would work well as a concept map. The researcher created a question sheet with 8-12 key concepts and appropriate propositional linking phrases, as demonstrated in the sample concept map in Figure 1. The concepts and linking phrases were reviewed by the classroom cooperating teacher, and they were printed on the test forms for students.

A scoring guide for the concept map question was used, following the Novak and Gowin Scoring System (1984), see Table 6. Each concept map was grouped by class, then scored independently and in random order by the researcher using the scoring

*Table 6. Novak and Gowin Concept Map Scoring System (1984)*

<b>Component</b>	<b>Description</b>	<b>Score</b>
Propositions	Is the meaning relation between the two concepts indicated by the connecting line and linking word(s)? Is the relation valid?	1 point for each meaningful, valid proposition shown.
Hierarchy	Does the map show hierarchy? Is each subordinate concept more specific and less general than the concept drawn above it (in the context of the material being mapped)?	5 points for each valid level of the hierarchy
Cross-links	Does the map show meaningful connections between one segment of the concept hierarchy and another segment?	10 points for each valid and significant cross-link. 2 points for each cross-link that is valid but does not illustrate a synthesis between concepts or propositions.
Examples	Specific events or objects that are valid instances of those designated by the concept level.	1 point for each example.

system. In order to translate the concept map score from the rubric to a test score comparable to the paired short answer question, the students' scores were grouped by question, such that the highest scoring map for a question and class became a perfect score. See Figure 4 for a flow chart of the scoring process as described. By grouping by class then question, students were only graded against students answering the same question and in their own class. Because the short answer questions were decided to be worth 8 points each by the cooperating classroom teacher, the concept map percentage score was then calculated to a point score out of 8 points, rounding to the half point. The traditional tests were scored by the master teacher using her own scoring system. The test was graded on a point scale, with the short answer questions worth 8 points each, and the test worth a total of 45 points.

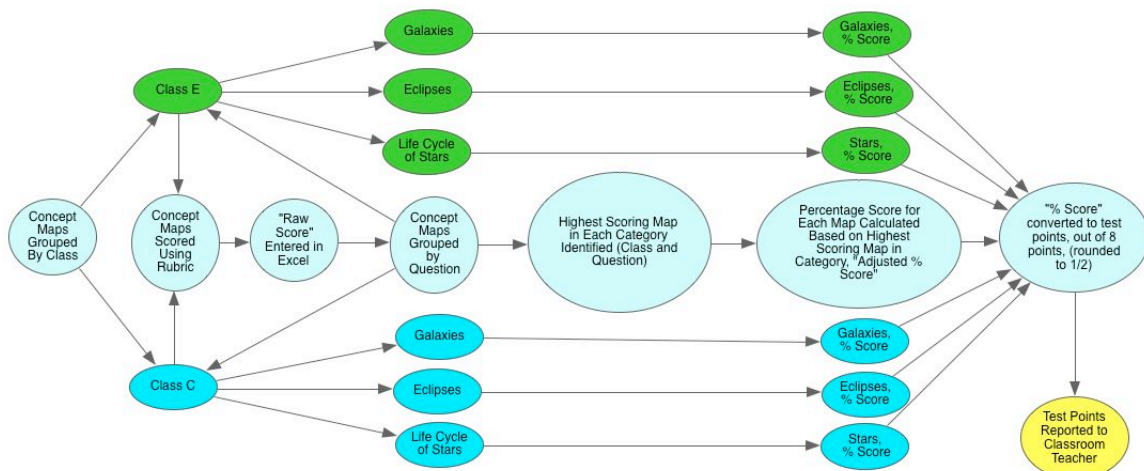


Figure 4. Flow Chart of Concept Map Scoring System.

### Post-test survey.

A short survey of 4 questions was completed by all students at the end of the test to gather student feedback. The questions were aimed at gathering student opinion on question format preference, class time spent on concept mapping, and how prepared the student felt for the concept map test. The survey was developed by the researcher, with input from the cooperating classroom teacher and the researcher's research group. A detailed review of research on the use of student surveys to rate classroom experiences and instruction shows that student surveys can be both reliable and valid, though there might be correlations between positive ratings and student perception of personality traits of the instructor (Costin, Greenough, & Menges, 1971). With this in mind, the survey was designed to gather student opinion on the test methods rather than on the researcher's instruction on concept maps.

### **Teacher journaling.**

Teacher journaling was used to collect technical reflection and reflection-on-action observations regarding the implementation of concept mapping tests in the classroom (Del Carlo, et al., 2010). A journal entry was made for each day of the instructional unit, with special attention paid towards concept mapping days. The entries used the following guidelines for reflective writing (Francis, 1995):

1. Summary: a brief summary of the key points.
2. New Learning: major new insights that come out of the session.
3. Questions: questions that emerge from the topics, issues or strategies.
4. Personal Reaction: a personal reaction to the context, content or strategies used.

Additionally, the journal recorded data on time spent by the researcher related to this research project, including: time spent preparing for concept map lessons, time spent in class on concept mapping, and time spent scoring maps.

### **Procedure**

This was an action research study that took place over a 2 week period in April 2011. It was carried out in two mixed-grade classes of Foundations of Physics and Chemistry (FPC) at a public high school near the urban center of a moderately sized city in the Pacific Northwest.

During the first day of the new curriculum unit for astronomy, both classes received the concept mapping lesson previously described. After the concept map training, both classes received identical curriculum for a 2 week unit in astronomy that was developed and taught by the cooperating classroom teacher. One class was

designated to receive extra concept mapping time (E) while the other served as a control (C) and only received the initial concept mapping lesson and the traditional curriculum. The E classes had 3 extra concept mapping activities included in lesson plans throughout the unit. All students completed “traditional format” unit tests previously created by the cooperating classroom teacher along with a construct-a-map test question created by the researcher that was paired with one test question.

Question numbers and topics for the three paired concept map questions were posted on the chalkboard prior at the beginning of the test. Students were reminded to select at least one of the short answers with a paired concept map to complete. Because the concept map question provided a list of concepts and propositions to use which are not traditionally provided with the short answer format question, students completed the traditional test first and turned it in before selecting the concept map question of their choice to complete. As students handed in their tests, a sticker coded with the class period and a number was placed on the test and an identical sticker was placed on the concept map question and survey packet to maintain anonymity. Students were instructed not to write their names on the concept map/survey packet. All students completed the post-test survey.

The cooperating classroom teacher scored the traditional test and the researcher scored the concept map test. Concept maps were grouped by class and scored in no particular order within the group. Concept map scores were reported back to the classroom teacher using the number codes so students could receive the higher score on the paired question. A spreadsheet coded by student number was used to track scores on

the concept map question, the paired question, the overall test score, and survey responses. Using the main test as the linking list, study participation was determined by cross referencing consent forms, and participation or not was noted in the spreadsheet. Concept maps and surveys for students not participating in the study were set aside and those scores were deleted from the spreadsheet. Photocopies of student responses to the paired short answer question were made and marked by code number for further comparisons.

Teacher data was collected longitudinally during the study by keeping a teacher journal. A journal entry was made for each day of the unit, with special attention paid towards concept mapping days. Along with the teacher journal, a teacher generated chart was used to track all time spent related to concept mapping during the unit.

### **Data Analysis**

Given the small sample sizes of this research study, descriptive statistics were utilized to review the data gathered. Given the small number of samples, ANOVA analysis was not used. The qualitative data (student surveys and teacher journals) was coded and summarized using descriptive statistics.

Student understanding and the ability of the pre-service teacher to measure student understanding were measured using the concept map scores and paired short answer question scores. Because the concept map and short answer responses were scored independently, a sample of student response pairs were reviewed to ensure the scores were comparable. These student response case studies of paired responses were examined for inclusion of similar content details demonstrating student knowledge, any



demonstration of student misconceptions, as well as for scoring similarities and irregularities between the two questions. Additionally, the amount of time needed to review the short answer response was compared to the amount of time needed to score the concept map, as recorded in the original data.

All student responses were reviewed on a cumulative basis using descriptive statistics to determine whether the concept mapping scores could be used by an inexperienced teacher to form broader conclusions on the students' understanding.

Feasibility of using concept mapping as an assessment tool by a novice teacher was examined by reviewing the data records and the researcher's journal for time requirements for concept map preparation and scoring, researcher's reflections on implementation and on concept maps.

## **Findings**

This action research study was guided by two research questions. The following section addresses findings for each of previously stated research question separately.

### **Research Question 1**

How feasible is it for a pre-service teacher to use concept maps to assess student understanding and what are the advantages or possible barriers of using this method?

To answer the first research question posed in this project, a two part analysis to quantitatively and qualitatively and examine student scores was followed by a qualitative review of teacher journal entries.

#### **Student response case studies.**

To determine whether scores were reflective of understanding, and the validity of the independently assigned concept map and short answer scores, a sample of student responses were reviewed to compare content of student responses and scores. These case studies were 3 randomly selected paired questions from each class; each case study pair was examined for display of student conceptual knowledge and to compare student performance and determine how well the concept map score reflected understanding compared to the information contained in the short answer result and the short answer score. The random number generator function was used in Excel to randomly select 3 students from each class for review. Students 427, 404, 416 were selected from class E

and students 216, 218, and 202 were selected from class C. Copies of these student responses can be found in Appendix B.

***Student 216.***

Student 216 selected the Eclipse concept map question to answer. This student was one of the few who received a higher score on the concept map than the short answer question with a score of 6.5 on the concept map and 4 on the short answer, and an overall test score of 33 out of 45. The student answered the short answer version of the question by drawing a table for lunar and solar eclipses and put 3 facts below each type of eclipse: the length of each eclipse, that the corona is visible during solar eclipses, and implies that there is something about location that is important to the visibility depending on the eclipse type, but does not fully realize this thought to demonstrate knowledge of the differences between lunar and solar eclipses and what the location on Earth has to do with seeing them. The cooperating classroom teacher had the comments “More detail please”, “what blocks what?” and “phase? frequency?” along with an emphatic “More!” next to the student response. The concept map scored well on all categories, and the researcher note calls this map “clear, detailed” and said the map “left out some important details about position of earth, moon, sun, but still managed to include many details”.

Compared to the short answer, the concept map is very complete. Among other details left out of the short answer, it shows that the student understands that solar eclipses are rare to see at a given location on earth, though they occur as frequently as lunar eclipses. The propositions missing from the concept map that would have improved the score on propositions have to do with the relative positions of the Earth,

Moon, and Sun during eclipses, though the student does include that the Sun is blocked by the Full Moon during a solar eclipse, so the sun is not visible other than the corona. The concept map for this student shows that the student does have a solid understanding of the similarities and differences between lunar and solar eclipses while the short answer implies that student does not. The time needed to assess the concept map for this student was more (3 minutes) than the time to assess the short answer (1 minute). The higher score on the student's concept map compared to the short answer is accurate given the detail included.

***Student 218.***

Student 218 selected the Galaxy concept map question to answer. This student scored a 5 on the concept map, a 6 on the short answer question, and had a total test score of 39 out of 45. The time needed to assess both answers was about the same (2 minutes). The concept map for this student was noted as “simple, lacking meaningful cross-links” by the researcher at the time of scoring. The lack of cross-links between concepts contributed to the low score on the map; this student scored 0 points for cross-links using the scoring guide. Additionally, there are very few linking phrases used by the student. The information contained on the map is organized correctly for the most part, with minor omissions, with information grouped with the appropriate galaxy type, but it seems to show that the student lacks skills in concept mapping. The short answer response lacks the description of “pinwheel arms” that was included in the concept map, but the short answer includes that elliptical galaxies are “circular or oval shaped” and that irregular galaxies have “no defining shape.”

Comparing the short answer and concept map answer pair shows that answers are also missing information regarding barred spiral galaxies, among other small details. Both answers reveal the student's basic knowledge of galaxies, though the short answer did have slightly more detail overall. Given the greater detail on the short answer and the lack of cross-links on the concept map, the higher score on the short answer is justified.

***Student 202.***

Student 202 was the only student in class C to choose the Star Life Cycle Concept map. This student scored 8 out of 8 on both the concept map and the short answer and 45 out of 45 on the test. Both answers were very detailed and demonstrate that the student has a strong understanding of stars. The short answer states "All stars start out as nebula, a large cloud of dust and gas, then the dust and gas through gravity form a protostar, this protostar gains mass and starts to fuse hydrogen into helium which transforms the protostar into a real star." On the concept map, some of this detail regarding the fusion of hydrogen to helium as a step from changing from a protostar to a star is missing. The short answer includes the statement that a neutron star is "super dense" and this not part of the concept map.

Overall, it seems that the concept map for this student shows slightly less detail than the short answer. However, the concept map for this student is more legible than the short answer response and took about the same amount of time to assess (4 minutes). The student has created a concept map response that was noted as "excellent" by the researcher and a short answer response noted as "nice!" by the cooperating classroom

teacher. Even with slightly less detail on the concept map, the perfect scores on both responses are justified given the attention to detail on both questions.

***Student 427.***

Student 427 chose the galaxy question to complete and scored a 6 out of 8 on both the concept map and the short answer, and a 42 on the test overall. Both answers are detailed, but just not quite as complete as they could be. The notes on the concept map score by the researcher are “Basic map, missing cross-link opportunities” and the cooperating classroom teacher noted “what could they do” and “caused by?” in response to the lack of information given about irregular galaxies. The concept map is also missing this information, even though “collisions” was included as a concept on the question. The student instead made the proposition that “spiral [galaxies] have many collisions”. While this is somewhat valid, the idea that is missing on both answers is that irregular galaxies are formed from collisions. That this idea is missing on both responses is an indication of a lack of understanding by the student for this idea. The time needed for both responses was about the same (2 minutes for the concept map, 1 minute for the short answer). The scoring for this student’s responses was valid given the missing content on both replies.

***Student 404.***

Student 404 selected the eclipse question and scored a 3.5 on the concept map and a 4 on the short answer, with an overall test score of 21 out of 45. The researcher note on the concept maps stated “Some misconceptions apparent” and the cooperating classroom teacher noted “when are they formed” and “more!” on the short answer. The short

answer was composed in table form and the student listed 3 ideas for each eclipse, including that solar eclipses are dangerous to look at, location is important for viewing each type, and included frequency. The concept map answer is difficult to read and is a jumble of concepts and linking phrases. The student does have the map organized into halves, such that there is a cluster of thoughts for solar eclipses and a cluster of thoughts for lunar eclipses, with some of the concepts generally correct; for example, the student has “7 minutes,” and “corona” with solar and “can be seen from everywhere” and “not dangerous” with lunar. But several of the propositions with linking phrases are incorrect: “solar [eclipse] results from umbra” and “lunar [eclipse] in front of position”. Additionally, many links were made without linking phrases at all, making a judgement of student understanding difficult. Scoring the concept map took about 2 minutes and reading the short answer took about 1 minute. The slightly lower score on the concept map compared to the short answer was justified given the lack of correct propositions on the concept map and lack of information included in the short answer.

***Student 416.***

Student 416 answered the Eclipse question and had a score of 8 on both the concept map and the short answer, with an overall test score of 36 out of 45. Researcher notes for the concept map state “detailed map, good details and cross-links” and the cooperating teacher noted “good!” next to the short answer response. The student elected to create a table for the short answer response, and included 6 details under each eclipse type. The concept map response is clearly organized and contains all of the information in the short answer plus details that were not included in the short answer, such as the

“the corona of the sun causes eye damage.” Additionally, the student used complex cross-links to indicate the position of the Earth, Moon, and Sun for each eclipse type, information that is not present on the short answer. Scoring the concept map took about the same amount of time as scoring the short answer table (2 minutes and 1 minute, respectively). The perfect score on both questions was justified given the amount of information the student included in both. However, it is of note that the concept map response contained more information than the short answer response, and it was only the scoring conventions used for the purposes of comparison in this research project that limited the concept map score to an 8.

#### **Student scores.**

Once the random sample review ensured that the independent scoring of the concept map and short answer were accurate given the content in the student responses, all student responses were reviewed on a cumulative basis to determine whether the concept mapping scores could be used by an inexperienced teacher to form broader conclusions on the students’ understanding. Overall test scores were compared by class, then question scores were compared by class.

The astronomy unit test scores from the E and C class reveal little difference between the two classes. Class C had an average score of 84.2% and Class E had an average score of 77.1%, as seen in Figure 5. Error bars indicate the standard deviation from the average.

A review of paired test scores by class revealed the average score was higher in the C class than the E class for both question formats, as might be expected given the



overall higher test scores for class C. The average concept map score was 7 out of 8 in the C class and 6 out of 8 in the E class. The average short answer score was 7 of 8 in the C class and 6.5 of 8 in the E class.

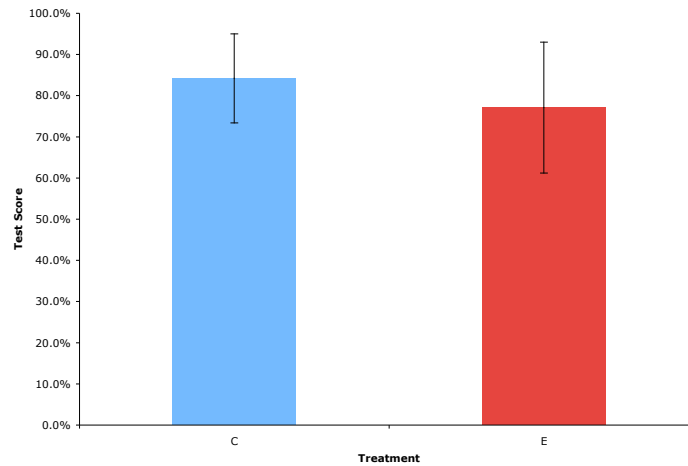


Figure 5. Overall Student Test Score by Class Condition

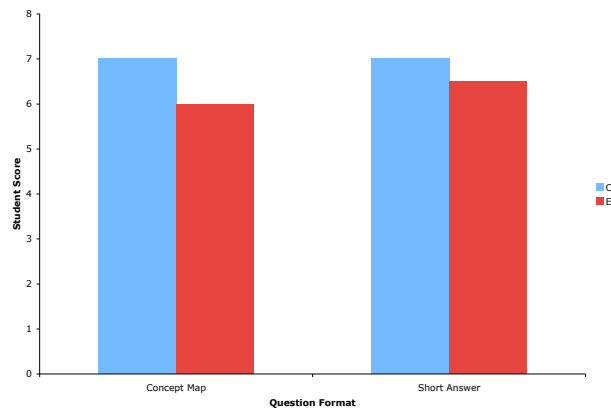


Figure 6. Average student test score on paired question by class condition

Because students were able to answer the question of their choice, and because each concept map was scored against “like” concept maps in the same class, it was necessary to review the differences in score by question format on a per question basis.

This analysis was done in Excel by calculating the average score for each question and question format, calculating the standard deviation of the average, and creating a chart to depict the results.

The student scores on the Star Life Cycle question are presented in Figure 7, though with only 1 student in the C class completing the question makes comparisons difficult. An additional student completed the Star Life Cycle concept map with a score of 6/8, but did not complete the paired short answer question, so that score is not included in the comparison. The remaining student in the C class scored 8/8 on the concept map question and the paired short answer question, while the average score in the E class on this question pair was 5/8 on the concept map and 6/8 on the short answer (n=4). However, the range of scores on the concept map was greater than the paired short answer question, giving the concept map scores a greater standard deviation than the short answer scores for class E.

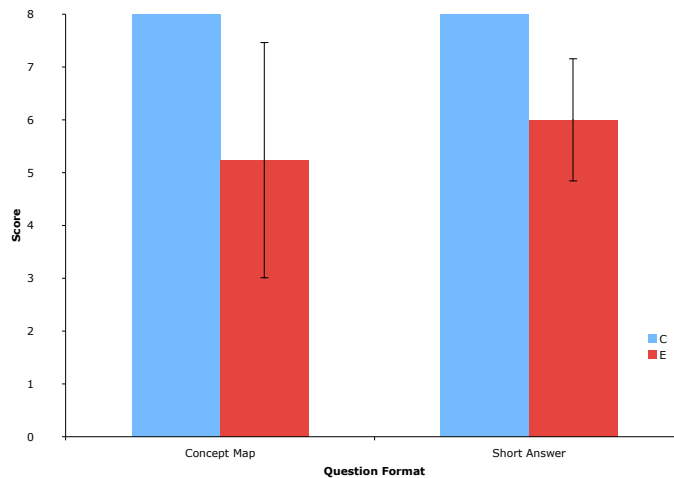
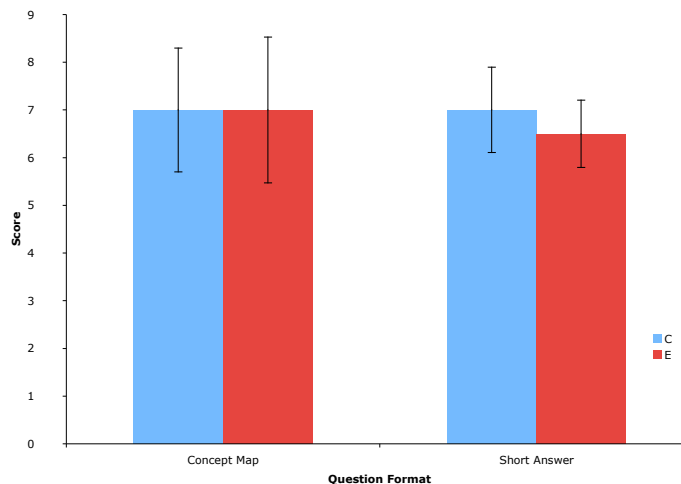


Figure 7. Student Test Score on Star Life Cycle Test Question by Question Format.

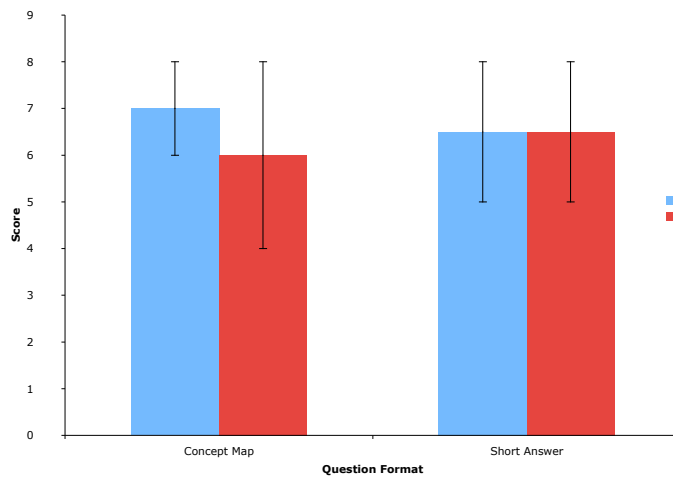
The Galaxy question was selected by more students in class C than class E, with 6 students in C completing the question compared to 2 in the E class. Figure 8 shows the mean score for both classes on the concept map was 7/8, with a slightly greater range in scores seen for the concept map scores in class E compared to C. The average score for the short answer question was 7/8 for class C and 6.5/8 for class E, with a greater range for the short answer in the C class than the E class. There was one student in C who completed the concept map with a score of 5/8, but did not complete the paired short answer so those scores are not included. There was one student in E who chose the Galaxy concept map question and was scored a 0/8 (there was essentially no response to score) but did not complete a paired short answer response so those scores were also removed from the comparison.



*Figure 8. Student Test Score on Galaxy Test Question by Question Format.*

Most students chose to answer the Eclipse question on the astronomy test, with 11 students in class C and 10 students in class E selecting this question. The average score

for the concept map question was 7/8 in C and 6/8 in E, with greater range of scores in the E class compared to C. The lowest scoring concept map in E was 3.5 and while the lowest scoring map in C was 5, and both classes had at least one map score 8/8. The average score on the short answer question was 6.5/8 for both classes, with an identical range of scores observed. Class C had a mode of 8 compared to class E with a mode of 7 for the short answer score, but both had a median score of 7 on the eclipse short answer question. Figure 9 shows the average student scores on the Eclipse question.



*Figure 9. Student Test Score on Eclipse Test Question by Question Format.*

### **Journal review.**

#### ***Preparation.***

Creating concept maps for students to use with the lessons was time consuming and often “frustrating,” as stated in the researcher journal. The researcher noted frustration with trying to find example concept maps on the internet as references, and

had limited success locating sample concept maps for inspiration and ideas. The researcher spent approximately 15 hours over different days researching existing concept maps for the subjects covered in the astronomy unit and drafting and refining materials for the 3 concept maps for student use in the E class; preparation time did diminish as the researcher became more comfortable with concept maps. Planning and creating the concept map training session took about 3 hours, not including a 1 hour meeting with the cooperating classroom teacher to discuss the lesson and research project. Creating the concept map test questions took about 2 hours. Preparing for concept map instruction was not measured outside of the time needed to create the concept map materials.

### ***Implementation.***

Despite the researcher's journal reporting being "nervous" about starting the research project, the teacher journal reports students in class C being "attentive" and it was noted that they "asked questions, and seemed to be interested." One student proposed an alternative to the sample concept map, which was a positive response and recorded as "exciting." It was noted that initially many students had trouble getting started, but "once they got going they seemed to think it was easy." Final thoughts for the class C lesson note that it "was a successful training session" and "my cooperating teacher made several comments about how well she thought it went" and "said she was looking forward to using concept mapping for her other classes." The positive experience with class C resulted in the initial nervousness being replaced by a confidence and feeling of relaxation going into the E class lesson. However, the journal entry shows that class E had much lower engagement with the concept mapping lesson, and there

were “blank faces and stares” and “overall the training session did not seem to go as well as with the other class.” The journal also notes that the “teacher commented that ‘sometimes it is hard to get them engaged’ without me saying anything, so I feel like she really noticed the difference as well.”

Despite the rough start, the additional days of concept mapping activity in the E class went off without any issues noted in the teacher journal. The first extra concept mapping day was noted as “successful” with “most students having good practice,” “enthusiastic,” “good questions,” “some students have extensive cross-links.” It was noted that 2 students, one who struggles in class and one that is traditionally a difficult student to engage, were not working on the concept map assignment at all. The student reported not knowing where to start. The questions that students had were about the mechanics of mapping, and how to organize the information. The 2nd concept mapping exposure in class E was noted as “went well,” though “students had difficulty with gravity as a linking phrase.” The 3rd concept mapping exposure noted that “students indifferent,” and the journal reported that there was “some groaning” in response to the introduction to the concept mapping activity. There was one student who commented to the researcher in private that he “can’t stand concept mapping.” One student did not take the worksheet with him/her at the end of class, which is not necessarily unusual in the classroom, but the journal notes that “I feel like the kids are getting sick of concept mapping/this research project. I sense they are annoyed.” Other than the negative concept mapping experience on this day, the class seemed “normal for them.”

On the test day, the journal reported that there was “some chaos in both classes” not normally present on a testing day, and inferred that it was a result of the “different testing routine,” requiring “the explanation of the coding stickers” and “needing to select specific questions on the short answer in order to have a concept map pair.” In class C, “several [students] had questions about how to do it” and there were several students who needed help regarding the use of the provided concepts and linking phrases. In class E there were three questions about the use of the concept map for a grade compared to one in class C, and also four questions about how to select a pairing question compared to two in class C. The most difficult part of testing implementation was clearly explaining the need for students to select a short answer that had a paired concept map and how the concept map/short answer question would contribute to the score on the test.

***Scoring concept maps.***

Faced with the responsibility of scoring the concept maps, the researcher noted initially feeling “intimidated” and “not sure where to start.” The use of the scoring guide was “awkward at first” and the researcher referred to the rubric multiple times while scoring the first maps, needing to re-read the scoring criteria. After the first 10 maps, the

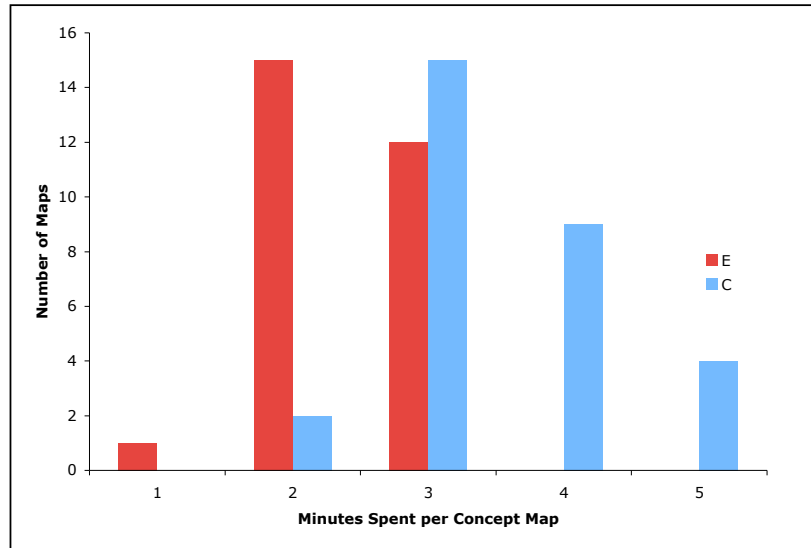


Figure 10. Histogram of minutes spent scoring concept maps sorted by class.

journal reports that the researcher was “getting the hang of it.” Scoring the concept maps was not prohibitively time consuming, and as the researcher scored more maps, the time spent per student response dropped. The first class scored was class C, which took an average of 2.5 minutes per map, for a total of 75 minutes to score. Class E took an average of 1.4 minutes per map, and a total of 40 minutes to score. Figure 10 is a histogram showing how many maps required 1 minute, 2 minutes, 3 minutes, 4 minutes or five minutes. No map took longer than 5 minutes to score. The histogram is broken up by class condition to show the difference in scoring time by class. Class C, which was scored first, had only 2 maps that took 2 minutes, with the rest needing 3-5 minutes. Class E, which was scored second, had 1 map that took 1 minute, 15 maps that took 2 minutes, and 12 maps that took 3 minutes to score. Overall, the scoring of maps for class E took less time than class C.



## **Research Question 2**

Is a single lesson on concept mapping sufficient for students to prepare them for a unit test based on concept maps? If not, how much practice is necessary?

To answer the second research question of whether a single class period is sufficient training for students to prepare them for a unit test based on concept maps, student surveys were reviewed for general trends and opinions, then student survey responses were reviewed in conjunction with test scores.

### **Student surveys.**

The student surveys revealed that there was a range of opinion on the amount of time spent in class on concept mapping that was not clearly linked to the class condition. As seen in Figure 10, the majority of students in both classes felt there was “just enough time” spent in class on concept mapping. There were 5 students in each class that thought there was “not quite enough” time. 3 students in the E class felt there was “a little too much time” spent on concept mapping, compared to only 1 student in the C class.

Student self assessment of preparedness revealed that most students felt prepared for the concept map question. 26 students reported feeling “just prepared enough” or better. 4 students in the E class felt “not quite prepared” despite the extra concept mapping instruction they received, with 1 student in the C class reporting the same feeling. One student in each class self-reported feeling “no-where near prepared enough.” Survey results for this question are seen in Figure 11.

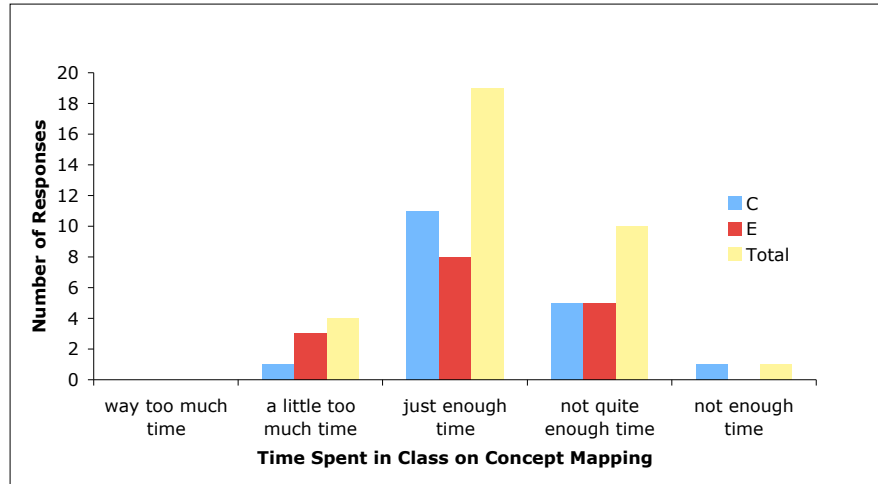


Figure 11. Student Assessment of Time Spent in Class on Concept Mapping.

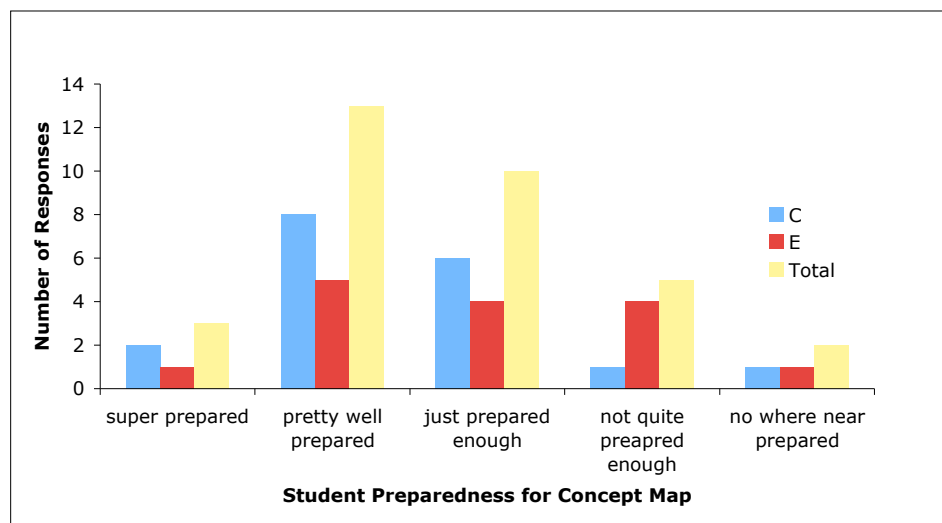


Figure 12. Student Self-Assessment of Preparedness for Concept Map Question.

Comparing the student opinion on the amount of time spent on concept mapping to average concept map question scores did not reveal a clear relationship between student opinion on the amount of time spent concept mapping and student scores, as seen in Figure 12. However, there was a relationship observed when comparing the students' self assessment of preparation level to the average score on concept map as seen in

Figure 13. In general, the students were accurate in their self assessment of preparation. Students who reported feeling “prepared” or better had higher average scores on the concept map than those who did not feel prepared.

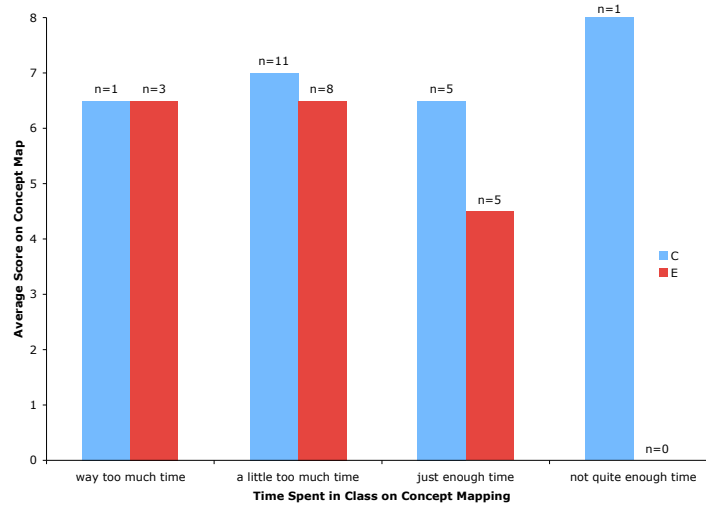


Figure 13. Student opinion on amount of class time spent concept mapping and average concept map scores by class. Sample size per category indicated by “n”.

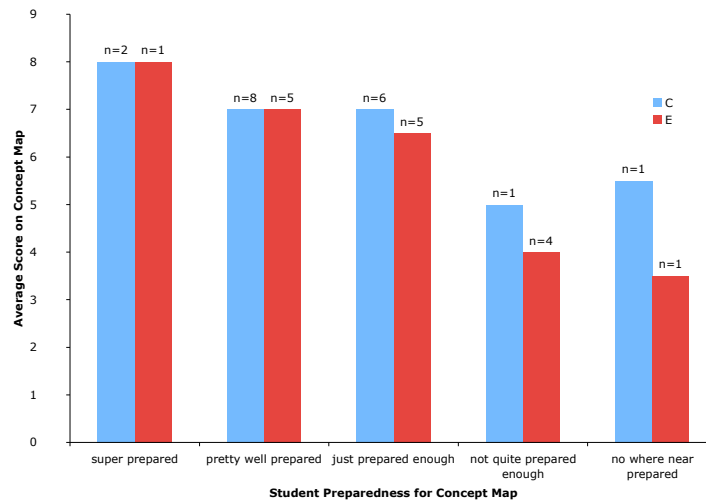


Figure 14. Student self-assessment of preparation for concept map and average concept map score by class. Sample size per category indicated by “n”.

While student preference for question format was not part of this investigation, the question was asked on the student survey and revealed that more students preferred the short answer format to the concept map format, as seen in Figure 15.

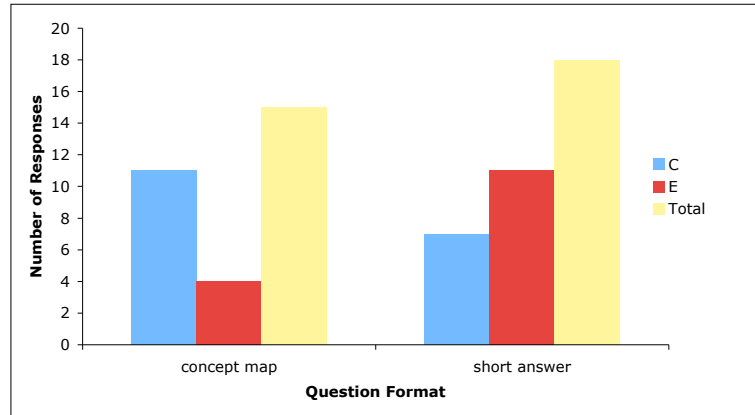


Figure 15. Student preference for question format.

## **Discussion**

This study was guided by two research questions. The following section addresses each of the research questions.

### **Research Question 1**

How feasible is it for a pre-service teacher to use concept maps to assess student understanding and what are the advantages or possible barriers of using this method?

#### **Case studies.**

In the comparing the short answers to the concept maps in the student responses, it seemed the independent scores assigned by the researcher for the concept map and the cooperating classroom teacher for the short answer were comparable based on the content and detail of each response. Point discrepancies observed between student scores for 404, 218, and 216 were valid given the differences noted between answers; either the student provided more detail on the concept map compared to the short answer resulting in a higher score (e.g. student 216), or the student provided less detail on the concept map compared to their short answer responses (e.g. students 404 and 218). The concept map scores seem to be an accurate reflection of the amount of detail and student knowledge demonstrated in the students' maps, as would be expected given the research and the use of the scoring guide (Ruiz-Primo, Schultz, Li, & Shavelson, 2001b; Novak & Gowin, 1984). This finding suggests that the scoring guide used to score concept maps was

implemented accurately and fairly by the researcher despite a lack of experience with the scoring guide rubric.

### **Student scores.**

Quantitative review of student test scores on an overall level and a paired question basis by class condition revealed several interesting findings. The overall test scores by class condition revealed that the additional concept mapping lessons did not appear to have an effect on the test score, as might be predicted given the extra instruction E received on certain topics. The difference in overall test scores observed in class E and class C suggests that there are differences between class E and class C, though with a small sample size and limited data, the difference cannot be quantified in the context of this research.

The other interesting finding was in comparing the individual question scores; not only did class E not score higher than C on the overall test, but class E did not have higher average scores on the concept map questions compared to class C. This finding suggests that the additional concept mapping exposure in the E class did not translate to scores higher than the control. This is a difficult comparison to make, given the unbalanced number of students selecting the questions as well as the few students and selecting the Star Life Cycle question in class C and the Galaxy question in class E. However, the eclipse question had the most students choose it, and almost equal numbers of students complete it, and it shows that class E averaged a full point lower than class C on the concept map. Based on the overall average scores in each class on concept mapping and the average scores on the eclipse question, it does appear that class E did

not seem to benefit from the extra time spent concept mapping, but this assumption will be further discussed in the “Limitations and Recommendations” section of the Discussion.

The comparison between the paired scores (the average concept map score to the average short answer score) by question and class condition did reveal an interesting and unanticipated finding. Even though the average scores between short answer responses and concept map responses were close, there was a greater range in scores on the concept map questions compared to the short answer scores. The greater range in concept map scores compared to the short answer scores, but close overall and paired average scores could be an indication of the sensitivity of the concept map score as a demonstration of student conceptual understanding. This outcome is supported by the literature (Ruiz-Primo, Schultz, Li, & Shavelson, 2001b) but was not expected to be observed in this study given the novelty of concept mapping to both the researcher and the students. That the concept map question scores seemed to be more a more sensitive reflection of conceptual understanding despite the lack of experience with concept maps of the researcher and the students is an important finding that should be shared with other teachers, and not necessarily just novice teachers, who might be intimidated by the idea of using concept maps as assessments in the classroom.

#### **Advantages and barriers.**

Having established that concept mapping is a potentially more sensitive technique for demonstration of student understanding, possible advantages and barriers to using

concept mapping in the classroom were evaluated, while keeping in mind that perhaps the additional time spent with the E class was not necessary to achieve the result.

One source of data for determining advantages and barriers was the teacher's journal. The journal recalls the struggles of engaging class E with the concept mapping lesson, followed by the subsequent interest, then concept map fatigue that the class demonstrated. The problem of engagement is a constant consideration for new teachers and not just limited to the use of concept maps or other novel assessment or instructional techniques. However, the fatigue that class E exhibited by the 3rd extra concept mapping lesson, as described in the journal, seemed to be confined to concept mapping itself and not the other activities of that day. This finding suggests that student disengagement and boredom are barriers that were not thoroughly considered in designing the additional concept mapping lessons for the E class, and might remain barriers for the future use of concept maps on a regular basis in the classroom by the researcher, depending on the class.

Time requirements for concept mapping were something the researcher considered a possible barrier prior to the start of the research study; anticipating time as a potential barrier resulted in the researcher collecting data on time spent with the goal of conducting a review of time spent on the various stages of the research study. However, review of the time data collected show that while initial concept map lesson preparation took a considerable amount of time, it is difficult to form conclusions about the true time required in lesson planning given there is no comparison unit planned by the researcher as part of this study. It was noted that concept map lesson preparation took less time as



the unit progressed and as the researcher became more comfortable with concept mapping. The researcher recognized that as a novice teacher, there is likely to be a large learning curve for any lesson planning in the classroom, especially lesson planning involving a new instructional technique.

The same overall decrease in time trend was noted in looking at the time needed to score the concept maps. The second class that was scored shows the dramatic drop off in time per map as the researcher became more comfortable using the scoring guide, reading student work, and assigning point values to that work. Overall, what was considered by the researcher as a likely barrier prior to the research study was not found to be a barrier during the study. Given the importance of selecting test formats with reasonable time requirements for scoring (Roediger & Marsh, 2005; Yin, et al., 2005), this outcome turned out to be an extremely positive finding of this research. It also confirms that the research by Yin, et al. (2005), which claimed the modified construct-a-map concept map test format was teacher friendly with regards to scoring ease and time requirements, is applicable even to novice teachers with limited experience with concept maps. This finding also supported the use of the Novak and Gowin concept map scoring guide as a method for quickly and easily assigning student scores to the concept maps by a novice teacher with limited experience with scoring guides (Novak & Gowin, 1984).

Several advantages for using concept mapping as an testing format as a novice teacher were identified. The case study comparisons of student answers revealed that the concept map answers were, in most cases, at least as detailed as the short answer, but in several cases, more detailed. The visual format of the map made it easy to see student

ideas, and compared to some difficult to read short answer responses, poor student penmanship became less of an issue for the researcher when scoring the maps. Finally, the presumed advantage of concept mapping being a superior method of testing student conceptual knowledge did seem to manifest, as previously discussed.

### **Research Question 2**

Is a single lesson on concept mapping sufficient for students to prepare them for a unit test based on concept maps? If not, how much practice is necessary?

To answer the second research question of whether a single class period is sufficient training for students to prepare them for a unit test based on concept maps, student surveys were reviewed for general trends and opinions, then student survey responses were reviewed in conjunction with test scores.

#### **Student surveys.**

While there was no prediction made as part of this research regarding what students would think about the time spent on concept mapping, it was surprising that there was no difference in the number of replies for “not quite enough time” on the student survey between the two class conditions. Additionally, there were more students in the control class, C, who received no additional class time concept mapping, who believed that they had received enough concept mapping with just the one concept mapping lesson.

Given the apparent concept map fatigue of students in class E, it was unexpected to find that only 3 students in class E believed that the class had spent “a little too much time” and no one selected “way too much time;” it was also surprising to see that one student in class C selected “a little too much time,” even with only one lesson spent learning concept mapping class C. Assuming the student survey results are accurate, this finding on student opinion of class time spent concept mapping would indicate that the researcher observations during the final extra concept mapping lesson in class E, recording students’ negative reactions to concept mapping, were not necessarily accurate reflections of student opinions towards concept mapping.

The results from the student surveys compared to test scores show that while there is a lot of variation in student preference for how much class time to spend learning concept mapping, there was no observed relationship between opinion of class time spent on concept mapping and scores. This suggests that student opinions on how much time to spend concept mapping are not a strong indicator of score; that is, a student who felt more time was needed on concept mapping did not necessarily get a low score on the concept map, and conversely, a student who selected “too much time” on the survey did not necessarily get a high score on the concept map. If students are able to self-assess whether more class time spent on concept mapping would be of benefit and the survey results are accurate, as suggested by research on student surveys (Costin, Greenough, & Menges, 1971), then the results showing that there is no relationship between time spent in class and concept map score further supports the previous finding that additional classroom instruction does not seem to result in higher concept map scores.

The survey question asking students about preparedness for the concept map did show that students were generally accurate at self-assessing preparedness, when preparedness rating was compared to individual concept map scores. The accurate self-assessment by students of their preparedness for a concept map question suggests that the larger predictor for student performance on concept mapping questions is more likely related to individual preparation efforts than instructional time spent with class concept mapping activities.

### **Limitations and Recommendations**

This study had several limitations which are important to consider as part of the discussion. Perhaps most significantly, any findings this research study are limited by the lack of generalizability due to this being an action research study. While the results are informative and valuable, given the nature of action research as a result of the specific classes the research is conducted in, it is not possible to predict future results based solely on the results found in this study (Feldman & Minstrell, 2000). The two classes participating in this study were unique, and if this study were to be repeated in two more classes, there might be different results observed between the E and C condition.

Another major limitation of this research was a result of the study design. The study design presented difficulties in that there was no pre-test administered to the E class to determine gains in concept mapping ability based on the extra concept mapping lessons. If this study were to be repeated, it would be advisable to add a pre-test to both classes for comparison purposes. It is possible this proposed change to study design would produce more meaningful findings.

A further limitation of the study design was introduced by the researcher by allowing students to choose with concept map question to complete. While this was done to follow the classroom conventions in place surrounding student freedom of choice, this resulted in very small sample sizes for two of the classes on two of the questions. If all students had been required to complete the same map, conclusions based on class condition comparisons would be more robust due to the larger sample comparisons available, though still overall limited by the action research study design.

The researcher's knowledge of the class assignment during the scoring of the concept maps should also be considered a limitation. It is unknown to what degree this knowledge of class assignment introduced a scoring bias. One way of eliminating a scoring bias would be the random assignment of study identification numbers, using a master list to keep track of the student and the study number, and assigning all of the concept map scores blinded to class assignment. The student class assignment could be unblinded at that point if needed, to group scores by class, assign test points, and transfer concept map scores to the test scores.

Additionally, the comparison of the short answer scores to the concept maps scores is problematic because of the two different scoring methods used as well as the two different people scoring the answers. In order to have a more meaningful comparison between the two scores, it would be best to have one person do all of the scoring for uniformity, and it would be better to use a rubric or scoring guide for both types of responses. Ideally, an alteration to the Novak and Gowin scoring guide used for the concept maps to include applicability to short answer responses could be made, so

that the same scoring guide could be used for all scoring, regardless of response format. By using the same scoring guide and having the same person assigning scores, the comparisons for the scores would become more robust and meaningful.

The lack of randomization in this study, which is a constraint of most educational research, also presents a limitation to the research. However, because of the difficulty in conducting a true randomized, controlled experiment in the classroom, this limitation would be difficult to eliminate in any future follow up studies (Cochran-Smith & Lytle, 1990). The two classes of students were not randomized into treatment conditions, and there were clear differences in the composition of the two classes at the start of the study. It is impossible to predict how the E class would have done on the concept mapping or on the content without the extra exposure, so the assumption that the extra lessons had no effect on concept map scores is based on the limited information collected in this study. Including a pre-test in the study design as previously discussed would alleviate, but not eliminate, the randomization problem of this study.

The small number of classes and students participating is something that could have been improved in this study and might have had an impact on study outcomes. There was a third class of FPC taught by the same cooperating classroom teacher that was not included in this study. By including an additional class in the study as another E class, it would have provided additional insight into the role of the class composition on the extra concept mapping; that is, was it this particular class E that did not show improvement in scores on concept mapping as a result of extra lessons, or would similar findings be observed in another class? Also, the participation rate within the classes

might have been increased had the researcher attempted enhanced student recruitment methods: by increasing the student participation within each class, a more accurate portrait of the class outcomes would have been possible. Future action research studies should carefully consider the inclusion of all possible classes and students within the classes as a way of increasing the teacher learning outcomes. Classroom researchers should also anticipate limited student participation and respond by building a fair, ethical, incentive into the research study that would inspire students to return participation forms.

One way of minimizing the limitations of action research findings is to accept that two goals of action research are to improve the learning of students and increase teachers' knowledge of instructional methods (Feldman & Minstrell, 2000). As this study has demonstrated, the novice teacher has gained confidence in using concept mapping as an instructional tool and an assessment tool, and has documented the learning of students with the tool. The fact that every class and school year will be slightly different despite covering the same general curriculum as a teacher is not a limitation in the opinion of this novice teacher/researcher; the continual challenge of meeting individual student needs is one of the most attractive qualities of a career in education.

## References

- Bruner, J. (1977). *The Process of Education*. Cambridge, MA: Harvard University Press.
- Cochran-Smith, M., Lytle, S.L. (1990). Research on teaching and teacher research: The issues that divide. *Educational Researcher*, 19(2), 2-11.
- Costin, F., Greenough, W.T., Menges, R.J. (1971). Student ratings of college teaching: reliability, validity, and usefulness. *Review of Educational Research*, 41(5), 511-535. Retrieved from <http://www.jstor.org/stable/1169890>
- Del Carlo, D., Hinkhouse, H., Isbell, L. (2010). Developing a reflective practitioner through the connection between educational research and reflective practices. *Journal of Science Educational Technology*, 19(1), 58-68.
- Dietel, R. (2004). *Why test students? Assessment Brief. Number 7*. Oakland, CA: Center for Assessment and Evaluation of Student Learning (CAESL).
- Feldman, A., & Minstrell, J. (2000). *Action Research as a Research Methodology for the Study of the Teaching and Learning of Science*.
- Francis, D. (1995). The reflective journal: A window to preservice teachers' practical knowledge. *Teaching and Teacher Education*, 11(3), 229-241.
- Gonzalez-Espada, W. (2009). Detecting Gender Bias through Test Item Analysis. *Physics Teacher*, 47(3), 175-179. Retrieved from ERIC database.
- Klassen, S. (2006). Contextual assessment in science education: Background, issues, and policy. *Science Education*, 90(5), 820-851. Retrieved from Academic Search Premier database.
- Kohn, A. (2000). High-Stakes Testing as Educational Ethnic Cleansing. *Education Digest*, 66(4), 13. Retrieved from Academic Search Premier database.
- Kuo, T.-M., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology*, 109(3), 451-464.
- Lawrenz, F., Huffman, D., Welch, W. (2000). The science achievement of various subgroups on alternative assessment formats. *Science Education*, 85(3), 279-290. Retrieved from ERIC database.
- Linn, R.L., Baker, E.L., Betebenner, D.W. (2002). Accountability Systems: Implications of Requirements of the No Child Left behind Act of 2001. *Educational Researcher*, 31(6), 3-16. Retrieved from JSTOR Archive.



- Mintzes, J.J., Wandersee, J.H., Novak, J.D. (Eds.) (1998). *Teaching science for understanding: A human constructivist view*. San Diego, CA: Academic Press.
- Novak, J.D., & Gowin, D.B. (1984). *Learning how to learn*. Cambridge, UK: Cambridge University Press.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).
- Portland, City Data. (2011) Assessor, Tax Lot Information. Retrieved from <http://www.portlandmaps.com>
- Portland Public Schools (2010). *Enrollment Summary Comparison. Enrollment by Program Type, October 2010*. Retrieved from [http://www.pps.k12.or.us/files/data-analysis/2010\\_Enrollment\\_Summary.pdf](http://www.pps.k12.or.us/files/data-analysis/2010_Enrollment_Summary.pdf)
- Roediger, H., & Marsh, E. (2005). The Positive and Negative Consequences of Multiple-Choice Testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 1155-1159. Retrieved from ERIC database.
- 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.
- Ruiz-Primo, M.A., Shavelson, R.J., Li, M., & Schultz, S. (2001b). On the validity of cognitive interpretations of scores from alternative concept-mapping techniques. *Educational Assessment*, 7(2), 99-141.
- Ruiz-Primo, M. (2004). Evaluating students' science notebooks as an assessment tool. *International Journal of Science Education*, 26(12), 1477-1506.
- Shavelson, R.J., Ruiz-Primo, M.A., & Wiley, E.W. (2005). Windows into the mind. *Higher Education*, 49(4), 423-430.
- The White House. (2001). *No Child Left Behind*. Washington, DC: The White House.
- Yin, Y., Vanides, J., Ruiz-Primo, M.A., Ayala, C.C., Shavelson, R.J. (2005). Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use. *Journal of Research in Science Teaching*, 42(2), 166-184.
- United States Census Bureau: State and County QuickFacts. (2009) Data derived from *Population Estimates, 2000 Census of Population and Housing, 1990 Census of Population and Housing, Small Area Income and Poverty Estimates, County Business Patterns, 2002 Economic Census, Minority- and Women-Owned Business, Building Permits, Consolidated Federal Funds Report, Census of*

*Governments*. Last Revised: Wednesday, 08-Jul-2009 18:31:25 EDT Retrieved from <http://quickfacts.census.gov/qfd/states/41/4159000.html>

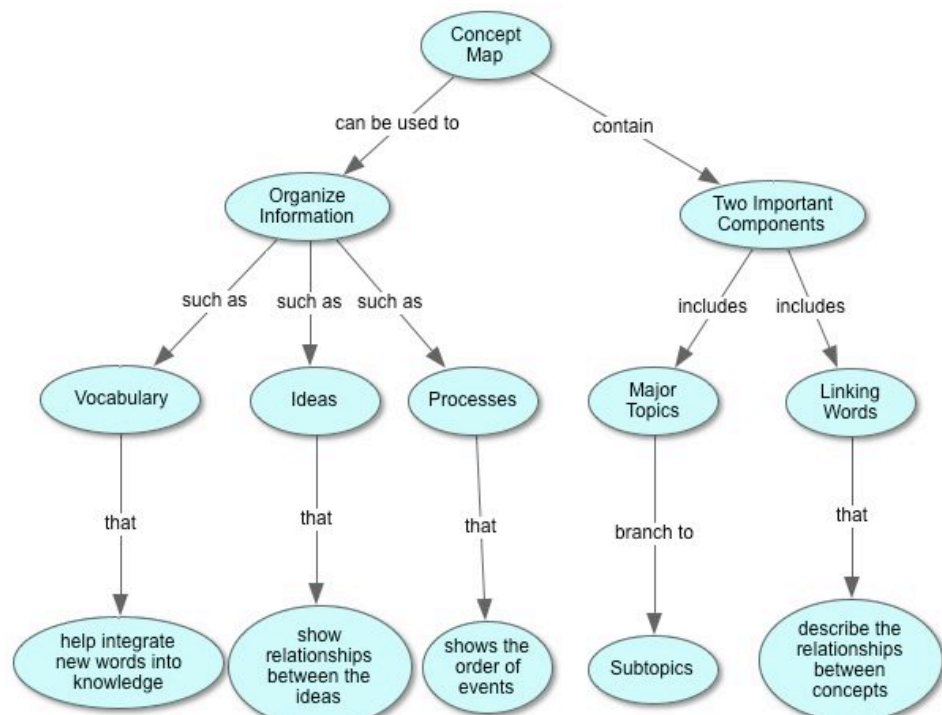
- Waters, F., Smeaton, P., & Burns, T. (2004). Action Research in the Secondary Science Classroom: Student Response to Differentiated, Alternative Assessment. *American Secondary Education*, 32(3), 89-104. Retrieved from ERIC database.
- Wiggins, G. (1993). Assessment: authenticity, context, and validity. *The Phi Delta Kappan*, 73 (3), 200-214. JSTOR (accessed February 23, 2010).
- Zurher, R. (1998). Issues and trends in culture-fair assessment. *Intervention in School & Clinic*, 34(2), 103. Retrieved from Research Starters - Education database.

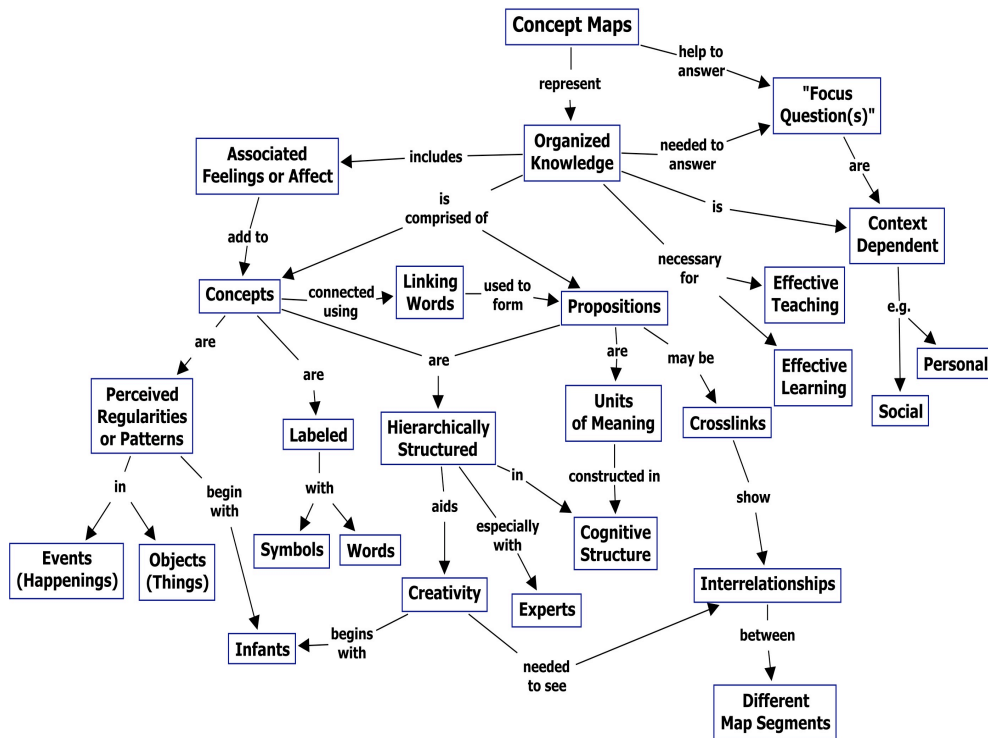
## Appendix A: Lesson Materials

### Concept Map Lesson

#### What is a Concept Map?

- ◆ A visual diagram of ideas from most general to specific (**hierarchy**)
- ◆ Each idea is enclosed in a **node** (usually enclosed in a circle or rectangle)
- ◆ It *shows* the relationship between ideas using lines with **arrows (cross-links)**
- ◆ It *describes* the relationship between ideas with **linking phrases**





## So why use a concept map?

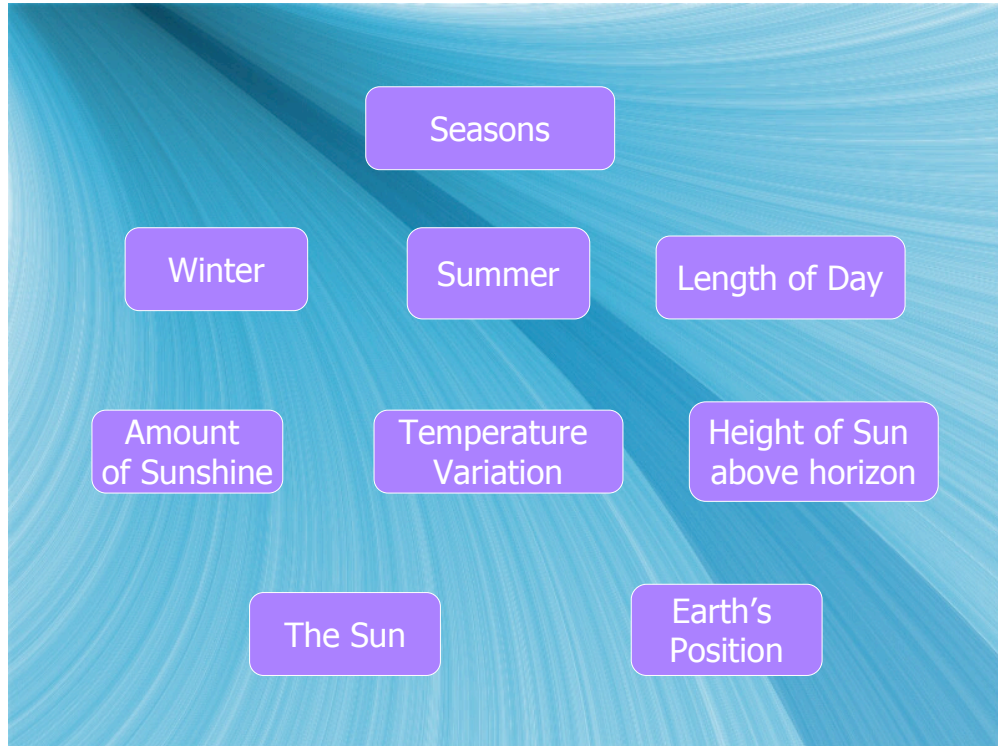
- ◆ Powerful learning tool: creating a concept map helps you organize your learning and structure new ideas in your memory
- ◆ Effective evaluation tool: a concept map can be used by a teacher to see your understanding (or misunderstanding) of subject matter

## Let's Make a Concept Map!

- ◆ Start with a **focus question**.
- ◆ **Why do we have seasons?**

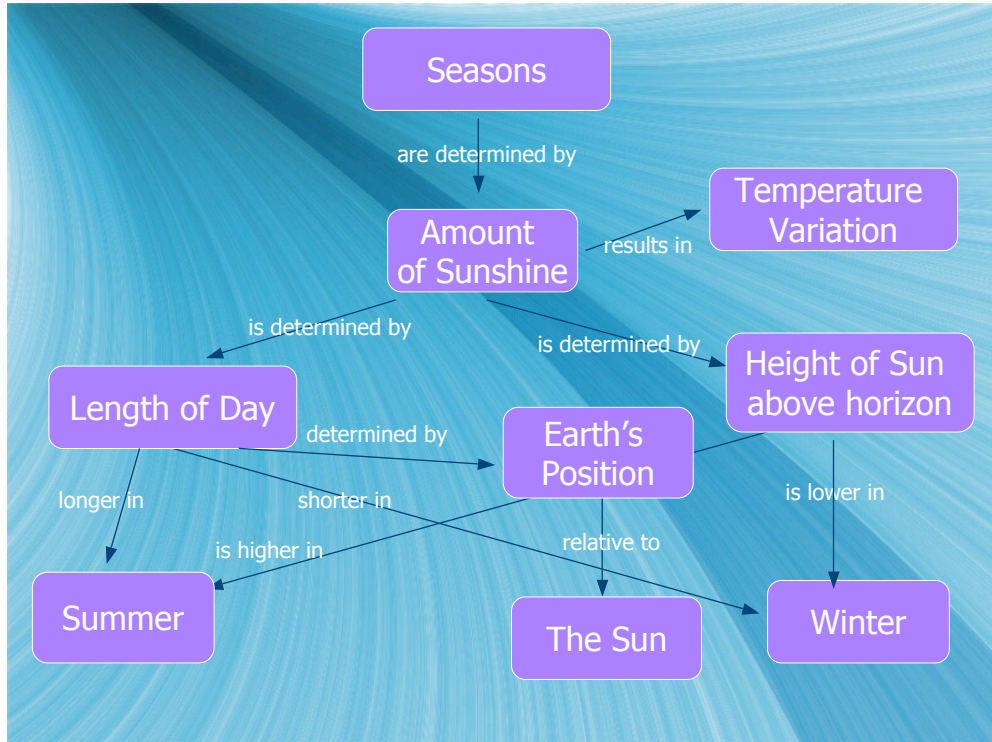
Seasons

- ◆ Brainstorm ideas/words that address the focus question.
- ◆ These ideas form the concept map **bank** that you can build from.



## Next Steps

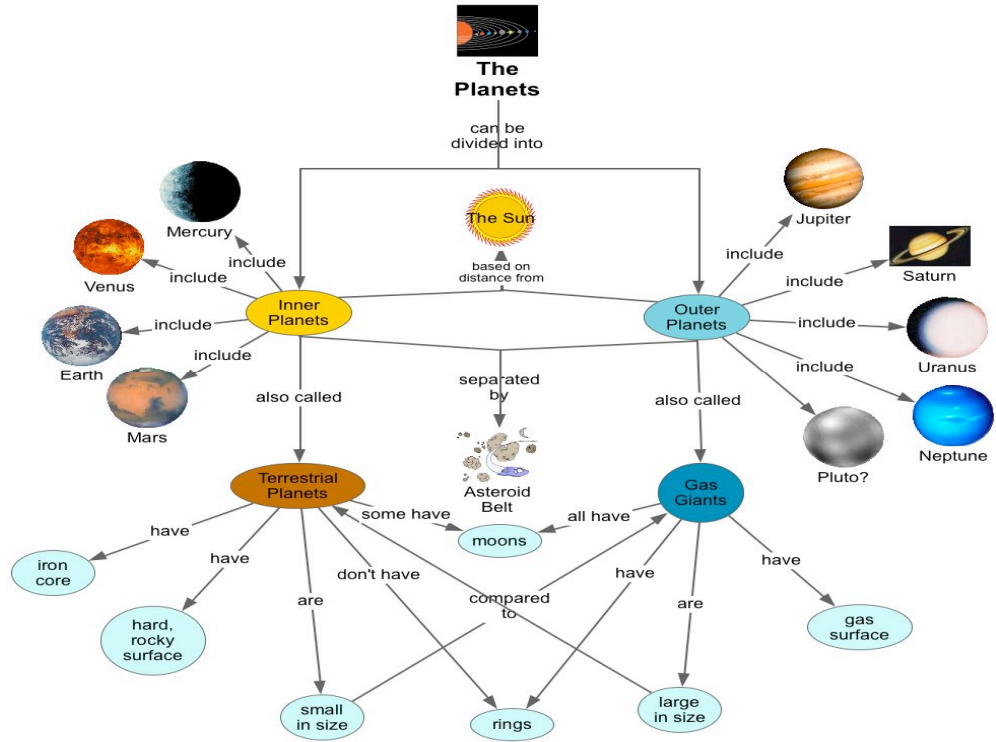
- ◆ Organize the ideas spatially
- ◆ Add cross links and linking phrases



## Practice

- ◆ Focus question: **How are the planets in the solar system organized?**
- ◆ Brainstorm ideas
- ◆ Organize spatially
- ◆ Create cross-links and linking phrases





## The Origin of the Universe

- Edwin Hubble: American astronomer observed galaxies during the 1920's from Mt. Wilson Observatory in the LA area
- Observed that the spectra of most galaxies were "red-shifted"
- This means that the galaxies are moving away from us
- If most galaxies are moving away from us, then the universe must be expanding!!

## The Big Bang Theory

- The universe came into being at a single moment, in the “big bang”
- All the matter and energy were concentrated in a very small hot region
- About 13.7 billion years ago, this region exploded and sent matter and energy expanding outward
- After several hundred thousand years of cooling, atoms formed
- Later, gravity pulled atoms together into stars and galaxies

## Further Evidence for the Big Bang

- 1965, two physicists, Arno Penzias and Robert Wilson, discovered “cosmic background radiation” from every direction
- This background radiation is believed to be energy from the Big Bang still traveling throughout the universe

## Fates of the Universe

- 📁 Continued expansion forever??
- 📁 Continued expansion, but only closer and closer to a certain limit??
- 📁 Expansion up to a point, then gravity pulls back and results in a “Big Crunch”??
- Fate depends on the amount of mass in the universe
- Hard to determine the mass because some mass is “dark matter” and cannot be seen directly

## Star Systems

- Most stars occur in groups of 2 or more, most commonly, binary systems
- Star clusters (small groups of stars): open clusters, associations, and globular clusters
- Galaxies (huge groups of stars, dust, and gas bound together by gravity): billions of galaxies in the universe, each having billions of stars (some even have trillions!!)

## Types of Galaxies:

- Spiral (like ours, the Milky Way): central bulge with arms extending like a pinwheel, younger stars in the arms, much dust and gas, older stars in the bulge, rotate slowly
- Elliptical: spherical or oval, no arms, many sizes, very little dust and gas, tend to have older stars, rotate slower than spiral or not at all
- Irregular: disorganized appearance, could be from collisions with other galaxies or just larger galaxies "pulling" on them
- Images

## Life Cycle of Stars, Unit Day 3

### Life Cycles of Stars

- Stars tend to form in the densest regions of nebulae
- A nebula is a large cloud of gas and dust spread out over a huge region of space
- Gravity pulls more and more dust and gas together until a protostar is formed
- The protostar contracts and its internal temperature rises
- When the internal temp and pressure are high enough, nuclear fusion begins and the star is born (in nuclear fusion, hydrogen atoms form helium atoms and release much energy)

### Main Sequence

- Stars spend about 90% of their time in this main phase
- More massive stars burn hotter (bluer) and faster, and spend less time on the main sequence
- Low mass stars burn cooler (redder) and may spend 100 billion years on the main sequence!!
- The sun is a medium mass yellow star and has a main sequence life span of about 10 billion years

## Deaths of Stars



- Low to Medium mass like the sun:
  - hydrogen fuel in the core starts to run out
  - forces become unbalanced, gravity overcomes outward push of fusion
  - outer hydrogen fuses and outer layers expand (Red Giant)
  - helium starts to fuse, forming carbon and oxygen

## Origin of the Solar System

- The Nebular Theory: the solar system formed from a rotating cloud of dust and gas (formed from material expelled by previous stars!!)
- The solar nebula started to collapse and as it did it started to rotate
- As it contracted further due to gravity, it rotated faster
- It began to flatten out forming "protoplanetary disk"

## Planet Formation

- 99.9 % of the mass stayed in the center forming the sun, while small asteroid-like bodies called "planetesimals" revolved around the center
- Some planetesimals grew in size due to accretion (adding mass through collisions)
- Planetesimals combined and eventually formed planets
- Some planetesimals became the moons of planets



## Evidence for Model

- All planets revolve around sun in same direction
- Almost all planets rotate in the same direction (exception? \_\_\_\_\_)
- Planets and moons are on the same plane (the ecliptic) with some exceptions (Pluto!)
- Denser planets stayed close, less dense gassy planets were flung outward

## Problem with Model

- Recently, scientists have said that solar systems like ours are quite rare
- In similar solar systems, planets the size of Jupiter are much closer to their stars
- Generally, scientists have been very optimistic that there must be many many Earth-like planets out there, but now we're not so sure
- What kind of strange solar system are we living in?
- Link:

## The Moon!!

- About 384,400 km from Earth
- $\frac{1}{4}$  the radius of Earth and a gravitational pull  $\frac{1}{6}$  as strong as the Earth's
- No atmosphere, so extreme temps
- Surface features:
  - maria: flat plains of dark lava flows
  - highlands: light-colored rough mountains
  - craters: many! Show that the moon is geologically dead



## Formation of the Moon

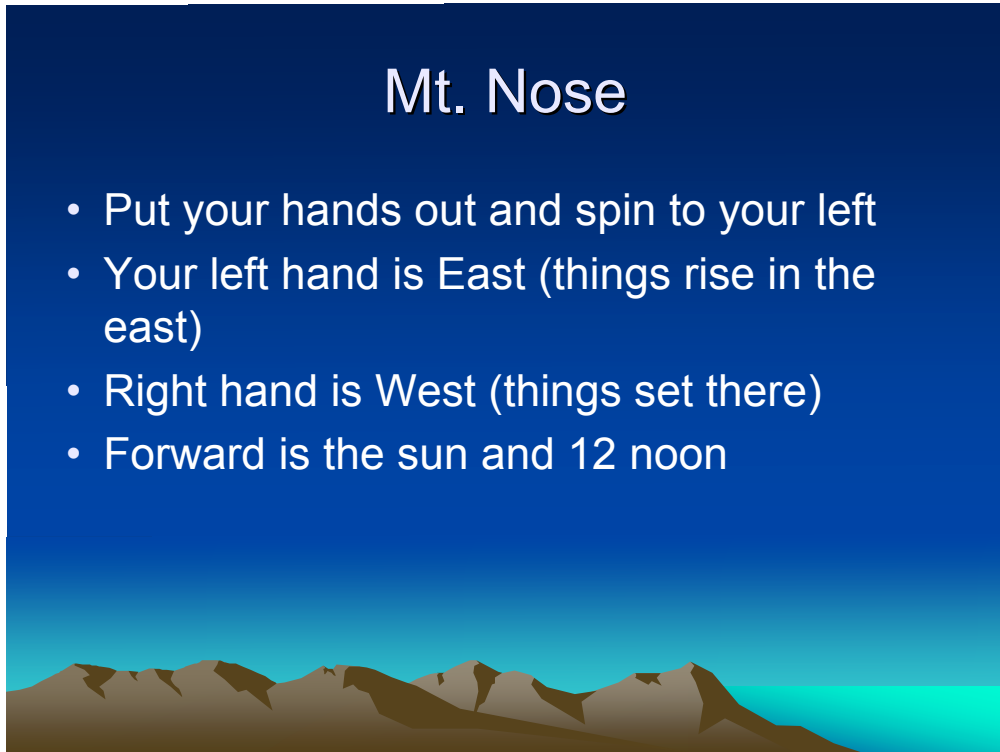
- Hypothesized that the moon was formed after a great collision early in Earth's history
- The collision ejected a lot of material, which eventually became the moon
- The moon revolves around the Earth in 29.5 days
- The same side always faces the Earth, so it also rotates once per revolution

## The Phases

- Phases are caused by changes in the relative position of the moon, sun, and Earth as the moon goes around Earth
- From Earth, we can only see the lit part of the moon that is facing us
- 8 phases:
- Draw them:

## Mt. Nose

- Put your hands out and spin to your left
- Your left hand is East (things rise in the east)
- Right hand is West (things set there)
- Forward is the sun and 12 noon



## Eclipses

- An eclipse occurs when the shadow of one body in space, such as a moon or planet, casts a shadow on another
- Lunar: occurs when the Earth casts a shadow on the moon
  - phase: full moon
  - duration: several hours
  - visibility: can be seen from anywhere on Earth, so fairly common
  - frequency: every year or so

# Solar Eclipse!!

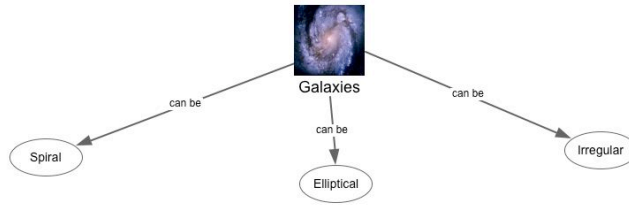
- Moon casts a shadow on the Earth
  - phase: New Moon
  - duration: no more than 7 minutes!
  - two shadows—umbra and penumbra, umbra causes total solar eclipse and penumbra partial
  - very rare at any one location, approx once every 370 years!
  - total solar occur somewhere ~every 18 mos.
  - sun's corona only visible during total solar
  - be careful of your eyes!!!

Pictures:

**Concept Map Lessons, Class E**

Galaxy Worksheet, Unit Day 3

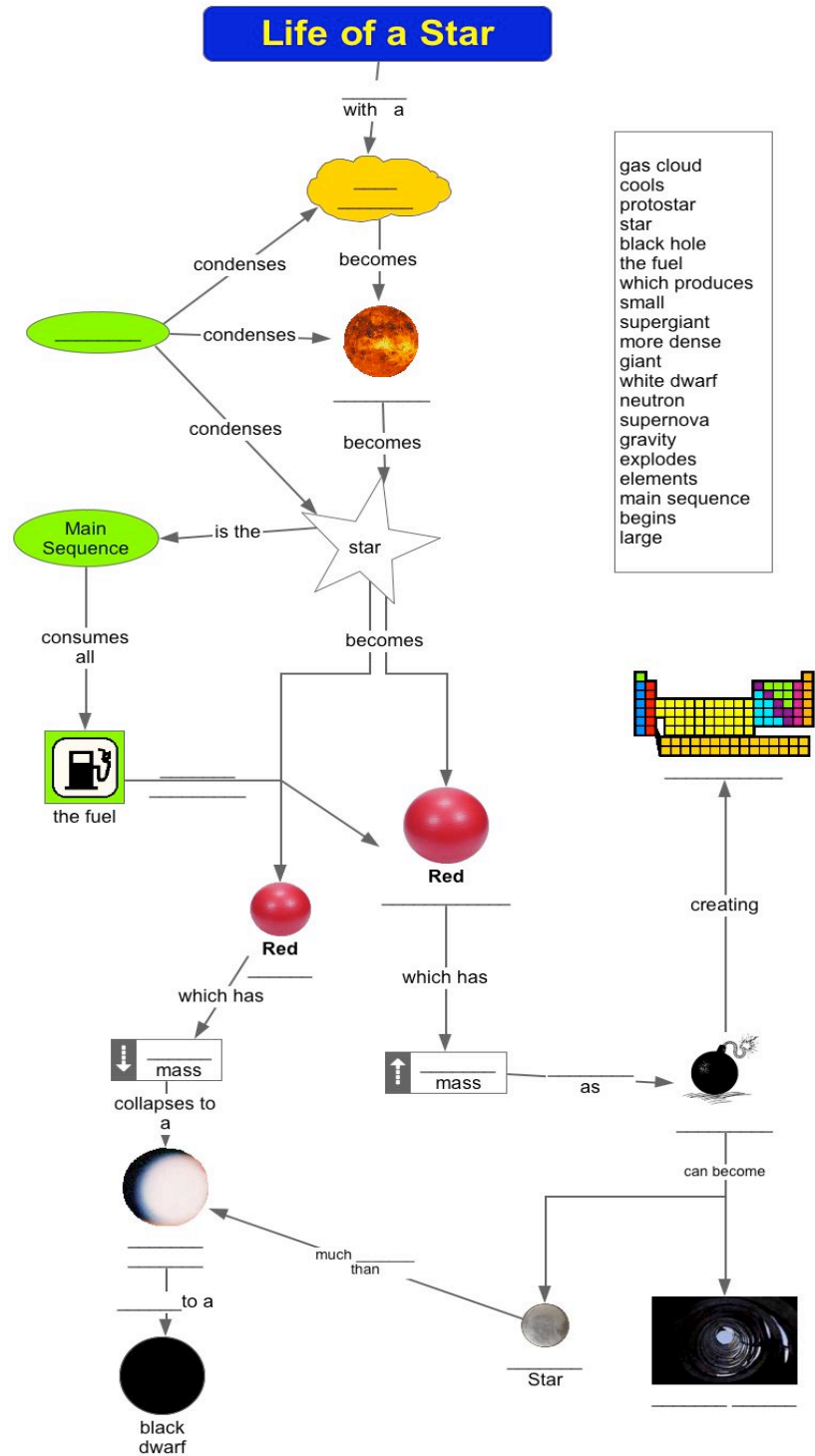
Focus Question: What are the differences between the types of galaxies?



- 
- collisions
- disorganized appearance
- Spiral Barred
- central bulge
- younger stars
- pinwheel arms
- older stars
- no arms
- rotate
- huge groups of stars
- younger stars
- dust and gas

can be	tend to have	don't have	include	rotate	very little	have lots of	have	slower than	billions of	are	with	formed by	
--------	--------------	------------	---------	--------	-------------	--------------	------	-------------	-------------	-----	------	-----------	--

Star Life Cycle<sup>2</sup>, Unit Day 5



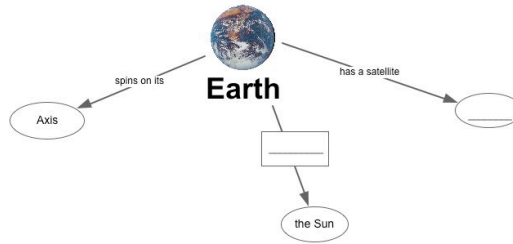
<sup>2</sup> Star Life Cycle Concept Map adapted from Inspiration Software Website, <http://www.inspiration.com/inspiration-science-examples>



# Earth, Moon, Sun, and Time, Unit Day 7

Focus Question: Describe the relationship between the Earth, the Moon, and the Sun, and time.

- 29.5 days
- the same direction
- sphere
- 8 phases
- orbits the Earth
- the Moon
- collision with Earth
- year (or 365 days)
- the Sun seems to rise in the east and set in the west
- day (or 24 hours)
- we have day and night
- 



which	once every	is shaped like a	causes	once every	because it spins as it	always faces	is a crater covered	orbits
results in	which is why	such as	formed by a	spins	visible as	revolves	rotates	

# Concept Map Test Questions

## Star life cycle

Draw a concept map that shows the life cycle of a typical star from its formation to its death. Show and describe the different endpoints for stars of low-to-medium mass (like our sun), large mass, and very large mass. Include as many of the provided concepts as you can and add your own if you need more. Use the linking phrases provided (the boxes) or create your own to show relationships.

### CONCEPTS

supernova

black dwarf

gravity

protostar

white dwarf

fuel

red giant

large mass

star

neutron star

main sequence

black hole

small mass

nebula

red supergiant

elements

has	affects	collapses to a	much denser	which produces	less dense than	becomes	creates	condenses
begins with	can become	causes	is the	consumes all	contains	cools to	produces	explodes as

# Eclipses

Draw a concept map for eclipses. Make sure that you make it clear how they are different and similar. Include as many of the provided concepts as you can and add your own if you need more. Use the linking phrases provided (the boxes) or create your own to show relationships.

**CONCEPTS**

- eclipse
- 7 minutes
- solar
- daytime
- several hours
- penumbra
- lunar
- common
- sun
- shadow
- nighttime
- full moon
- rare at one location
- umbra
- new moon
- moon
- corona
- every 18 months
- position
- earth

has	affects	can be seen from	occurs	which produces	are	are visible	creates	casts
results from	can become	causes	is the	behind	in front of	no more than	about	duration

# Galaxies

Draw a concept map showing the differences between the 3 main types of galaxies. Include looks, composition, and relative age. Include as many of the listed concepts as you can and add your own if you need more. Use the linking phrases provided (the boxes) or create your own to show relationships.

**CONCEPTS**

- rotate slowly
- larger galaxies
- many sizes
- dust and gas
- no arms
- huge groups of stars
- rotate
- younger stars
- disorganized appearance
- spiral barred
- central bulge
- older stars
- pinwheel arms
- collisions

has	affects	rotates	occurs	which produces	are	don't have	billions of	include	have	have lots of
results from	formed by	causes	is the	behind	tend to have	no more than	with	about	slower than	very little

## Concept Map Survey

Please take the time to read and answer each question below. Your opinions are a very important part of the research project. If you are not participating in the research project your responses will not be used for research purposes, but they are still important instructional feedback.

Your responses will be **anonymous** and will not affect your grade on this test!

1. Which test question format gave you a better opportunity to show what you learned:
  - a. concept map
  - b. short answer

2. Explain your answer for question #1 above:

3. How did you feel about the amount of class time we spent on concept mapping?  
Circle your choice below.

1	2	3	4	5
way too much time	a little too much time	just enough time	not quite enough time	not nearly enough time

4. How prepared did you feel for the concept map test question? Circle your choice below.

1	2	3	4	5
super prepared	pretty well prepared	just prepared enough	not quite prepared enough	no where near prepared

5. Is there anything else you would like to say about concept maps as a testing format or as a learning tool in science class?

Thank you for helping me with my research project!

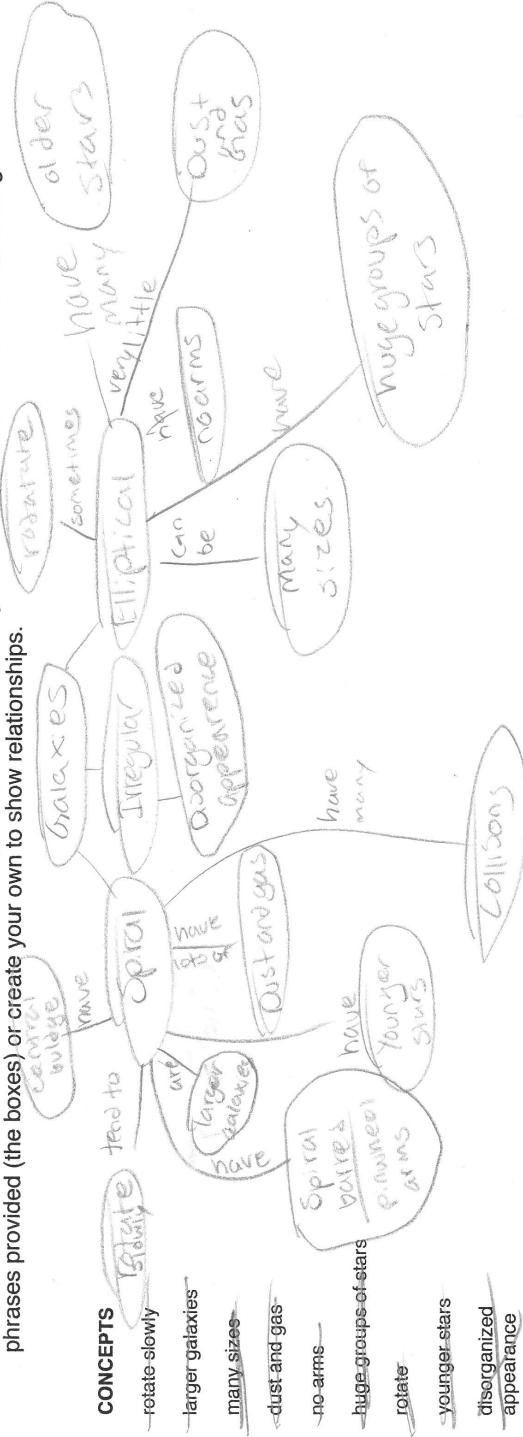
## **Appendix B: Case Study Student Samples**

### **Class E Case Study Samples**

# 427 Concept Map

FPC4/27

Draw a concept map showing the differences between the 3 main types of galaxies. Include looks, composition, and relative age. Include as many of the listed concepts as you can and add your own if you need more. Use the linking phrases provided (the boxes) or create your own to show relationships.



has	affects	rotates	occurs	which produces	are	don't have	billions of	include	have	have lots of
results from	formed by	causes	is the	behind	tend to have	no more than	with	about	slower than	very little

FPC4/16

Draw a concept map for eclipses. Make sure that you make it clear how they are different and similar. Include as many of the provided concepts as you can and add your own if you need more. Use the linking phrases provided (the boxes) or create your own to show relationships.



- CONCEPTS**
- eclipse ✓
  - 7 minutes ✓
  - solar ✓
  - several hours ✓
  - penumbra
  - lunar ✓
  - common
  - sun
  - shadow
  - full moon
  - rare at one location
  - umbra
  - new moon
  - moon
  - corona
  - every 18 months
  - position

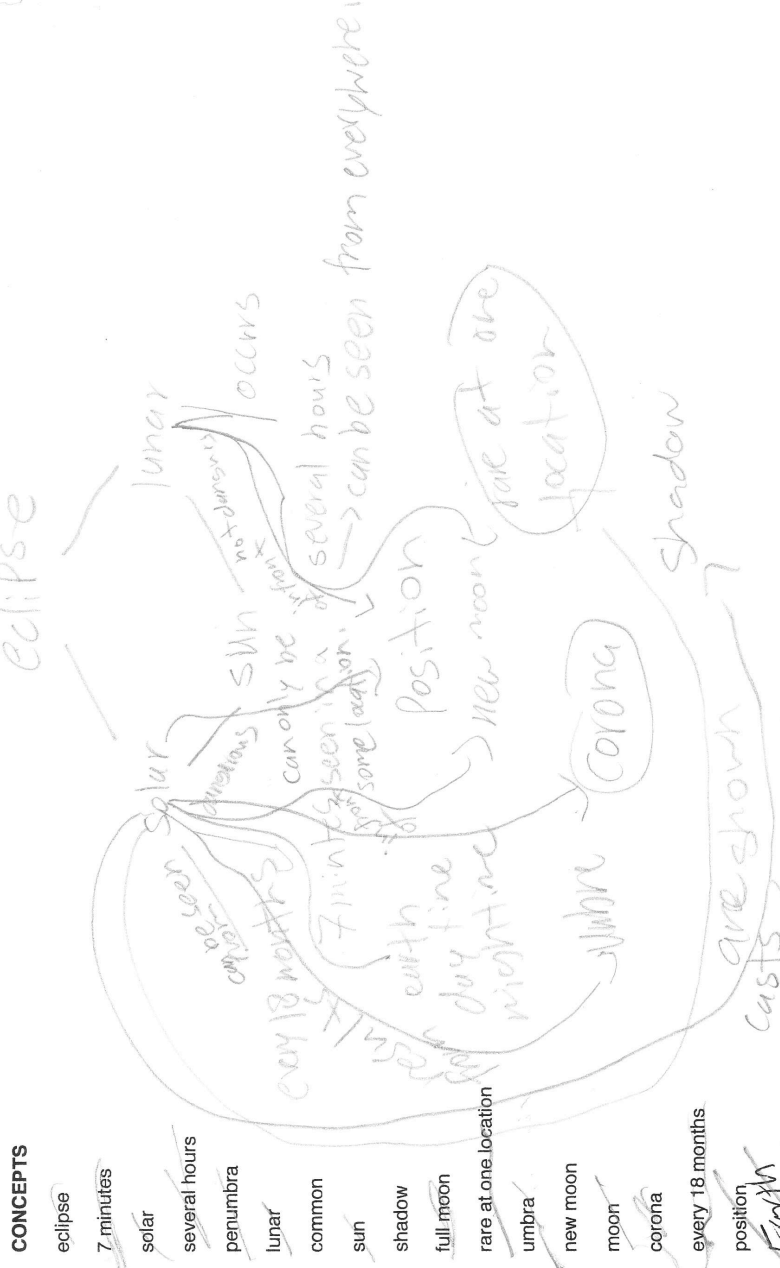
*Earth  
daytime  
night*

has	affects	can be seen from	occurs	which produces	are	are visible	creates	casts
results from	can become	causes	is the	behind	in front of	no more than	about	duration



FPC4/4



Draw a concept map for eclipses. Make sure that you make it clear how they are different and similar. Include as many of the provided concepts as you can and add your own if you need more. Use the linking phrases provided (the boxes) or create your own to show relationships.



has	affects	can be seen from	occurs	which produces	are	are visible	creates	casts
results from	can become	causes	is the	behind	in front of	no more than	about	duration

427 Short Answer

16) Galaxies: 3 types of galaxies, Spiral, Elliptical, Irregular.

Spiral: look like a disc with central bulge in the middle,  newer stars in the ends of the arms/ <sup>top</sup>  ligaments. have lots of gas and dust. Rotate slowly. They are younger galaxies.

Elliptical: older galaxies, shaped like an egg, oval, older stars, have little dust and gas, have no arms.

Irregular: have no particular shape, have some dust and gas.  $\rightarrow$  what could they do caused by?

how do we know so much?  $\rightarrow$  they do

416 Short Answer

	Solar	Lunar
13)	1) Duration is no more than 17 min.	1) Duration is several hours.
25/32)	2) Happens during the Day.	2) Happens at night.
8	3) Moon phase is the new moon.	3) Moon phase is the full moon.
grad	4) happens once somewhere on the earth about every 18 months.	4) happens every year.
	5) You can see the corona of the Sun.	5) The moon turns a kind of orange color.
	6) Can only see in small portions of the earth.	6) You can see anywhere on earth.

404 Short Answer

(1752)

(13)

	Solar	Lunar
more!	- is dangerous for the eyes.	- is not dangerous for the eyes.
4/8	- Only can be seen in some locations.	- you can see it from everywhere.
when are they formed?	- every 17 months	- every year or so.

## Class E Case Study Samples

### 202 Short Answer

15) All stars start out as a nebula, a large cloud of dust and gas, then the dust and gas through gravity form a proto star, this proto star gains mass the starts to fuse hydrogen into helium which transforms the proto star into a real star. With every star they go on this phase for about 1/2 their life, then when the inside runs out of fuel for low to medium mass stars the outside expands and fuses helium into carbon and oxygen, this is called a red giant. High and super massive stars they expanded into red supergiants and the start to fuse up to iron, when the red phase is done in low to medium mass stars the outside turns into a planet and nebula and the core turns into a white dwarf which will when cool become a black dwarf. For high mass stars they go supernova, a large explosion which fuses all the other elements and their core becomes a super dense neutron star while super massive stars also go super nova their cores get so massive they vibrate through spacetime to create black holes.

Nice!

### 218 Short Answer

16) Spiral (ours): has a central bulge with gas, dust and stars. The younger stars are in the arms, while the older ones are in the bulge. It rotates fairly slowly.

17) Elliptical: Is circle or oval shaped filled with mostly older stars. It has no arms, and rotates slowly or not at all. <sup>dust?</sup>

18) Irregular: It has no defining shape because it was hit by other galaxies. <sup>what else? how do we know so much?</sup>

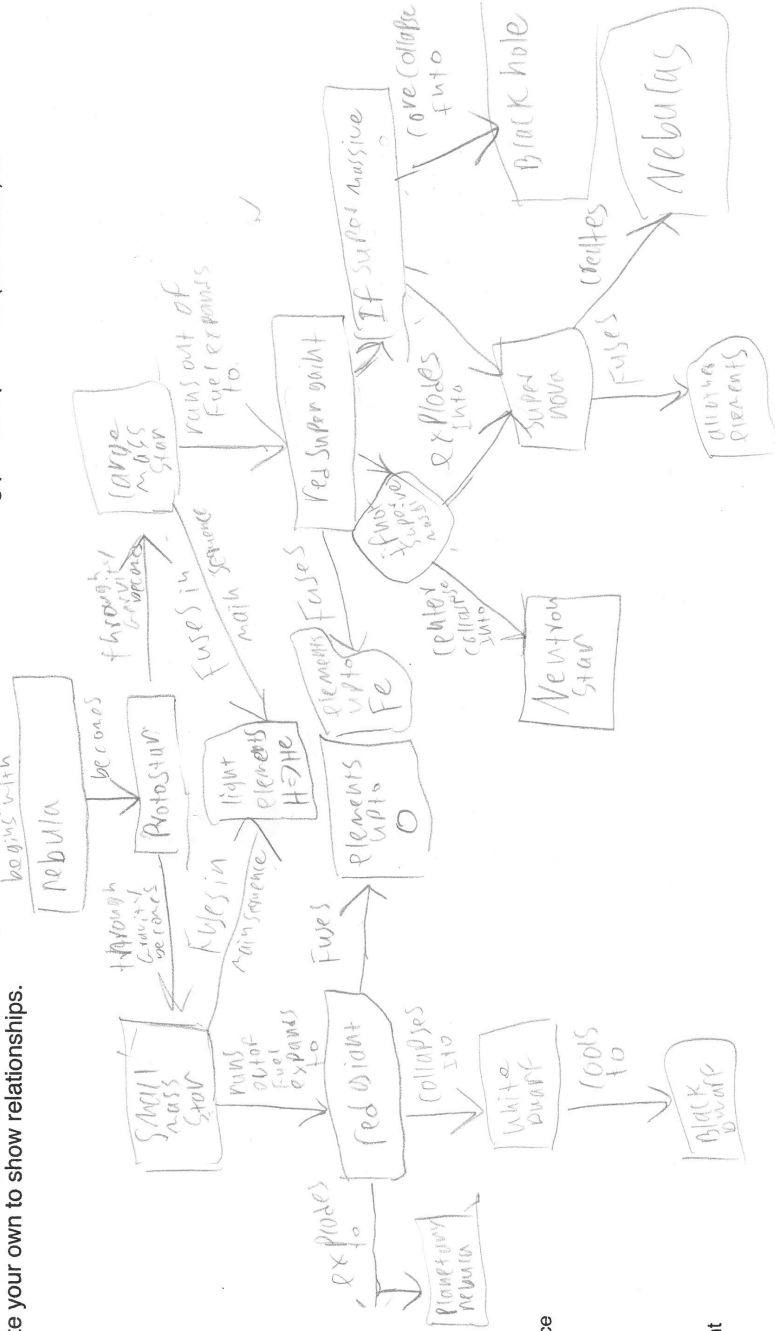
216 Short Answer

planets. More detail please  
what blocks what? phase? frequency?

13.) solar eclipses	lunar eclipses
More! last about 7 minutes	last several hours.
4/8 can see corona	can't see it (corona)
can't always be seen	can be seen from everywhere

FPC2/2

Draw a concept map that shows the life cycle of a typical star from its formation to its death. Show and describe the different endpoints for stars of low-to-medium mass (like our sun), large mass, and very large mass. Include as many of the provided concepts as you can and add your own if you need more. Use the linking phrases provided (the boxes) or create your own to show relationships.



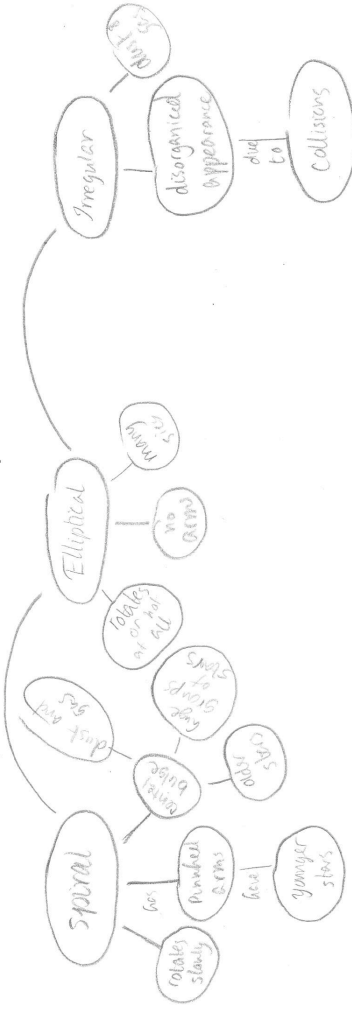
**CONCEPTS**

- supernova
- black dwarf
- gravity
- protostar
- white dwarf
- fuel
- red giant
- large mass
- star
- neutron star
- main sequence
- black hole
- small mass
- nebula
- red supergiant
- elements

has	affects	collapses to a	much denser	which produces	less dense than	becomes	creates	condenses
begins with	can become	causes	is the	consumes all	contains	cools to	produces	explodes as

FPC2/18

Draw a concept map showing the differences between the 3 main types of galaxies. Include looks, composition, and relative age. Include as many of the listed concepts as you can and add your own if you need more. Use the linking phrases provided (the boxes) or create your own to show relationships.



**CONCEPTS**

- ✓ rotate slowly
- ✓ larger galaxies
- ✓ many sizes
- ✓ dust and gas
- ✓ no arms
- ✓ huge groups of stars
- ✓ rotate
- ✓ younger stars
- ✓ disorganized appearance
- ✓ spiral barred
- ✓ central bulge
- ✓ older stars
- ✓ pinwheel arms
- ✓ collisions

has	affects	rotates	occurs	which produces	are	don't have	billions of	include	have	have lots of
results from	formed by	causes	is the	behind	tend to have	no more than	with	about	slower than	very little





## Appendix C: Research Approval Documents

### Portland State University Institutional Review Board Approval Memo



Human Subjects Research Review Committee

Post Office Box 751  
Portland, Oregon 97207-0751

503-725-4288 tel  
503-725-3416 fax  
hsrrc@lists.pdx.edu

November 18, 2010

To: Nancy Mitchell

From: Mary Oschwald, HSRRC Chair

Re: Approval of your application titled, "An Action Research Study to Determine the Feasibility of Concept Maps as Alternative Assessments by a Novice Teacher" (HSRRC Proposal # 101537).

Dear Nancy,

In accordance with your request, the Human Subjects Research Review Committee has reviewed your proposal referenced above for compliance with DHHS policies and regulations covering the protection of human subjects. The committee is satisfied that your provisions for protecting the rights and welfare of all subjects participating in the research are adequate, and your project is approved. Please note the following requirements:

**Changes to Protocol:** Any changes in the proposed study, whether to procedures, survey instruments, consent forms or cover letters, must be outlined and submitted to the Chair of the HSRRC immediately. The proposed changes cannot be implemented before they have been reviewed and approved by the Committee.

**Continuing Review:** *This approval will expire on November 18, 2011.* It is the investigator's responsibility to ensure that a Continuing Review Report (available in ORSP) of the status of the project is submitted to the HSRRC two months before the expiration date, and that approval of the study is kept current.

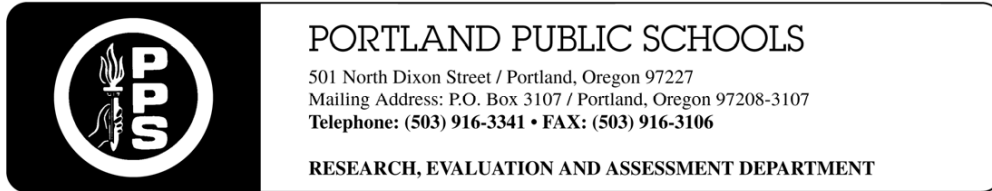
**Adverse Reactions:** If any adverse reactions occur as a result of this study, you are required to notify the Chair of the HSRRC immediately. If the problem is serious, approval may be withdrawn pending an investigation by the Committee.

**Completion of Study:** Please notify the Chair of the Human Subjects Research Review Committee (campus mail code ORSP) as soon as your research has been completed. Study records, including protocols and signed consent forms for each participant, must be kept by the investigator in a secure location for three years following completion of the study.

If you have questions or concerns, please contact the HSRRC in the Office of Research and Sponsored Projects (ORSP), (503) 725-4288, 6th Floor, Unitus Building, 4th & Lincoln.

Cc: Michael J. Flower

## Portland Public Schools Research Approval Memo



March 3, 2011

Ms. Nancy Mitchell  
4204 SE Clinton  
Portland, OR 97206

Dear Ms. Mitchell:

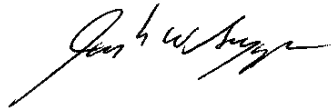
The Portland Public Schools Research & Evaluation Department has reviewed and approved your request to conduct survey research on *Feasibility of Concept Maps as Alternative Assessments by a Novice Teacher*. The study is consistent with Board policy and professional research practices. Per your most recent e-mail, we agree that you will not be collecting and using existing data from our student information system for your research. If you change your mind, and feel that you need existing data, you will need to provide copies of signed release forms to authorize such data releases.

I have communicated our approval of this survey to Peyton Chapman, Lincoln High School's principal. Participation in the study by the schools is voluntary; District staff and students are not obligated to participate in outside research, regardless of approval by the Research & Evaluation Department.

Please note that paid researchers must have worker compensation coverage while on school property and, because they will be working directly with students, they must have completed a PPS criminal background check. Background checks take about five working days and must be completed prior to accessing the school. Researchers meeting with students when a PPS employee is not present, must also be fingerprinted by our Security Services department. The current cost of fingerprinting is \$82.

The District would be interested in receiving information on the results of this study when they become available. Please submit a copy of the final report to this office. We wish you every success in this study.

Sincerely,



Joe Suggs  
Director  
Research, Evaluation & Assessment  
Portland Public Schools

c: Peyton Chapman, Lincoln Principal