The Effect of Concept Mapping On Student Achievement in An Introductory Non-Majors Biology Class

Julie Low Brinkerhoff\textsuperscript{1} and Gary M. Booth\textsuperscript{2}

\textsuperscript{1} Dept. of Plant and Wildlife Sciences, Brigham Young University, Provo, UT 84604

\textsuperscript{2} Dept. of Plant and Wildlife Sciences, Brigham Young University, Provo, UT 84604

*corresponding author, gary_booth@byu.edu, 801-422-2458

Introduction

Interest in studying science and in science related careers has been declining steadily since the 1970’s (Markow & Lonning, 1998). The pace of scientific discovery and the ability of American students to learn and perform science have lagged behind their international peers (Mason, 1992). Science education has, in many cases, become teacher centered, based on rote memorization, and focused on test scores (Heinze-Fry & Novak, 1990; Huai, 1997; Kinchin, 2001; Mason, 1992). “Most students consider science to be boring, a list of big words and facts, intimidating, and not relevant to their lives” (Mason, 1992). Negative attitudes towards the study of science are also fostered as students experience no connection between their study and their real lives (Roth, 1994).

In addition to this lack of interest in the sciences, an overall lack of critical thinking skills is becoming more evident in the general population. There is an increased inability of people to think through problems; individuals often fail to see solutions to problems that in retrospect, appear simple (Novak, 1998). The social and economic consequences of this trend are alarming, and measures have been taken to try to reverse these trends. Increases in per-pupil expenditures (300% from 1955 to 1985, after the adjustment for inflation) have been made, but have been largely unsuccessful (Novak, 1998).

During the 1970’s, educators and psychologists became aware of this trend and began movements towards more meaningful learning. David P. Ausubel developed an assimilation theory of cognitive learning; a theory from which other researchers developed models (Ausubel et al., 1978). Ausbel’s work became the basis for the work of Joseph Novak, who developed concept mapping at Cornell University in 1972 as a way for teachers and students to move away from rote memorization and regurgitation of facts towards more meaningful learning (Novak & Musonda, 1991). Novak’s work is the most frequently cited in concept mapping literature.

A concept map is a schematic device for representing the relationships among a set of concepts. A concept is a perceived regularity in events or objects, or records of events or objects, designated by a label (Novak, 1998). Concept mapping is a technique for externalizing one’s understanding of a conceptual framework (Novak & Gowin, 1984). It is not focused on individual concepts, but on the organization of a
set of concepts in a conceptual framework; it emphasizes organization of the whole. In its simplest form, concept mapping is a representation of knowledge organized into cognitive structures.

Knowledge is a human creation, where new ideas are constructed by creative people on the basis of their existing concepts and theories and of a search for new patterns or regularities in events or objects they observe (Novak, 2004). From birth to death, individuals construct and reconstruct the meaning of events and objects they observe (Kinchin & Hay, 2000; Novak & Musonda, 1991). Cognitive structure is a hypothetical construct referring to the organization of the relationship between concepts in long term memory; thus, it is difficult to measure (Huai, 1997).

Concept mapping differs from other knowledge representations in that it utilizes ideas from constructivist epistemology and Ausubel's assimilation theory of cognitive learning and that it places high value on prior knowledge in the process of acquiring new knowledge (Ausubel et al., 1978; Novak & Gowin, 1984; Novak & Musonda, 1991).

**Literature Review**

Because concept mapping was designed to promote meaningful learning, it has been used for a wide variety of specific research purposes. Educators have identified an area of difficulty in a class and then have examined concept mapping as a solution. Concept mapping has been used to:

2. Develop critical thinking skills (Able & Freeze, 2006; Briscoe & LaMaster, 1991; Kinchin, 2001).
7. As a Communication tool between teacher and student (Roth, 1994).
8. As a Compensation for different learning styles (Huai, 1997).
11. Evaluation/testing (Kinchin et al., 2005).

As indicated, concept mapping has been used as both an instructional and an assessment tool. Subjects in which students have used concept maps are varied as well, but are largely nested in sciences such as chemistry (Regis et al., 1996), microbiology (Kinchin et al., 2005), chemical engineering (Muryanto, 2006), nursing (Able & Freeze, 2006), and medical education (Rendas et al., 2006). Their use is most concentrated in the United States and Canada, but they have been used in the UK (Kinchin, 2001), China (Huai, 1997), Nigeria (Horton et al., 1993; Jegede et al., 1990), and Taiwan (Horton et al., 1993).

The extent to which concept maps have been used varies widely among the research literature. If examined closely, virtually all fit along a continuum of concept mapping ranging from maps used strictly as a visual aid (Laight, 2004) to completely self-directed map construction (Hardy & Stadelhofer, 2006; Regis et al., 1996). Intermediates include students completing skeletal or partially completed maps (Hardy & Stadelhofer, 2006; Rendas et al., 2006), and maps where students are given terms from which to construct their own map (Kinchin et al., 2005; Regis et al., 1996). Only one study examined combined each of the techniques and claimed the need for all in order to have effective concept mapping (Hardy & Stadelhofer,
2006). In contrast, Horton et al. (1993) found that there was no significant difference between the teacher prepared concept map used as a visual aid and the student prepared concept maps in improving student achievement.

Concept mapping is always used in conjunction with other teaching techniques. One of the most common pedagogies it is combined with is cooperative learning. Various researchers have used cooperative learning and have made claims as to the role it plays in concept mapping. Santhanam and associates (1998) found that cooperative learning made the maps too difficult to assess individual students. Another specifically examined the combined effect of concept mapping and collaboration, but was unable to distinguish if one component had a greater effect than the other (Okebukola, 1992a). Some researchers employed cooperative learning without much comment on the cooperative element (Muryanto, 2006; Roth, 1994), while others claim that collaboration is necessary for the success of concept mapping (Kinchin et al., 2005).

Sample sizes used in concept mapping research vary widely, ranging from a very few (about 6 students (Briscoe & LaMaster, 1991), to 150 (Roth, 1994)). These differences are largely due to the differences in the measurement methods used to determine the effectiveness of concept mapping, and the amount of time and money required for each selected measurement type.

Researchers have used several approaches to measure the effectiveness of concept mapping. Some have relied solely on a comparison of pre and post scores on a multiple choice test (Jegede et al., 1990; Okebukola, 1992a). Others have used a combination of multiple choice tests and interviews (Markow & Lonning, 1998). The most effective studies have used interviews to capture the cognitive structure of students (Briscoe & LaMaster, 1991; Heinze-Fry & Novak, 1990; Markow & Lonning, 1998; Mintzes & Wallace, 1990; Nicoll et al., 2001).

Origins of the interview date back to the nineteenth century work of psychoanalysts, although forms of systematic questioning were used in early Greek and Roman times or before. It was through Piaget’s work in the 1920’s and 1930’s (Cooney et al., 1993) that the interview was developed as a tool with strategies for use with children. Part of Piaget’s genius was in devising the specific interview events and objects and the questions that made it possible to observe regularities in children’s responses (Novak & Gowin, 1984; Novak & Musonda, 1991).

Interviews used throughout concept mapping research vary in their structure and in the disclosure of the researchers about the interview protocol; often the specific information about how the interviews were conducted is omitted. Formats also vary from highly flexible to highly structured. Novak and Gowin (1984) advised that interviewing for the purpose of examining the effectiveness of concept mapping should not be Socratic Teaching. Interviewers are not to ask questions that will steer students towards understanding; a temptation that has not been properly avoided in the past. Novak also recommended that interviewers be familiar with the subject matter in order to conduct an effective interview (however, his interviewing techniques involved the interviewer building a concept map based on student responses, rather than having the students themselves construct a map) (Novak & Gowin, 1984; Novak & Musonda, 1991).

The advantages of concept mapping can be significant. Students and teachers often remark that they recognize new relationships and hence new meanings or at least meanings they did not consciously hold before mapping. In this sense, concept mapping can be a creative activity, and may help foster creativity (Novak & Gowin, 1984). Concept mapping has been shown to improve test results, student attitudes, and overall enjoyment of the subject matter (Kinchin & Hay, 2000). When concept mapping is incorporated as a tool for deeper learning, it has the potential to move students away from surface learning into a deeper mode of learning for understanding.
Several problems are associated with the implementation of concept mapping in a classroom setting. Historically, assessment using concept maps has been difficult. Concept maps are difficult for many teachers to score because they are unfamiliar with them (Kinchin, 2001). Some researchers began using the maps as an assessment tool, but ended up discontinuing their use for assessment because of these difficulties (Regis et al., 1996). Concept maps often lack reliability and validity. A call for research in these areas has been made, with minor attempts to answer (Mintzes & Wallace, 1990). Because of these and other reasons, the studies examined often lacked statistical significance (Heinze-Fry & Novak, 1990; Markow & Lonning, 1998).

The time and effort required to implement concept mapping in a classroom is seen by many teachers as excessive. For a pedagogy that is generally unfamiliar, implementation can be problematic and the benefits appear small (Santhanam et al., 1998). Regis et al. (1996) effectively used concept mapping and implemented it over a four year period of study; the amount of time invested was sufficient to make the mapping experience successful for the students.

Training in concept mapping for students, instructors, and teaching assistants is also seen as a difficulty. The time and effort involved in training is considered, by some instructors, not worth the effort. Studies that were successful in implementing concept mapping were careful in their training (Regis et al., 1996).

Concept Mapping often lacked incentives in the classes where it was used. Teachers struggled to make students care enough about the maps to spend the necessary time and care to make them effective. One study attributed failure of concept mapping to such lack of incentives (Santhanam et al., 1998). In other studies, sufficient incentives were provided. If there was enough impact on grades, students were provided with an extrinsic motivation to complete the maps and to do them well (Kinchin et al., 2005). Brisco and LaMaster (1991) allowed the maps to be used as aids in the class exams. The most successful approach was a study which required concept mapping as part of a four year curriculum; students knew that they would need to know how to effectively map for their entire course of study at the school (Regis et al., 1996).

Another significant and common problem with concept mapping in the classroom is that it is often peripheral. Concept mapping is most often tacked on to lessons rather than being indicative of a teacher’s overall approach or underlying epistemological belief (Kinchin, 2001). Most often, the underlying philosophy of the curriculum is not in line with the principles behind concept mapping, and so it is out of place amongst the other strategies of the class.

There is little emphasis on long term use of concept maps. (Santhanam et al., 1998) Santhanam et al. (1998) examined the extent to which students continued to use concept mapping after the semester was over; the results indicated that none did. When concept mapping is not indicative of a deeper approach to learning by the instructor, it is insufficient to produce results on its own. Concept mapping is not a quick fix for educational problems, but it is a move towards more meaningful learning (Heinze-Fry & Novak, 1990).

Using essays to measure connected understanding after a concept mapping treatment has not been examined. Virtually all of the measurements have been done as interviews and/or multiple choice tests. Additionally, there has been little effort to look at lower quartile learners: the struggling students in a class. Novak has said that “we have found that many students classified as ‘learning disabled’ are really bright children who lack the skill and/or motivation for rote mode learning” (Novak & Gowin, 1984).

**Research Hypothesis**

Freshman Biology students who regularly use concept mapping will have greater connected understanding of the conceptual domains of enzymes, energy, proteins, and immunity than students who did not use concept mapping. More specifically, the mean scores of the concept mapping group will be
significantly greater than the mean scores of the non concept mapping group. This hypothesis is extended to the students in the lower quartile.

- $H_0: \mu_T = \mu_C$
- $H_A: \mu_T > \mu_C$

**The Population**

**Freshman Academy**

All of the students for this study came from an organization known as Freshman Academy. Freshman Academy is a learning community option that is given to all incoming freshman at Brigham Young University. Students who choose to participate in Freshman Academy take an envelope of classes together, live in close proximity to each other, and have peer mentors to help them take advantage of university resources and aid in their adjustment to university life. Students in Freshman Academy have professors who are dedicated to making their first year experience a good one, and who are generally known for working well with students.

The Biology classes formed from the entering freshman class of 2006 were examined for equality in gender ratios, entering high school GPA, and entering ACT score (Table 1). T-tests revealed no difference in between classes in entering high school GPA or ACT scores. Chi-Square analysis revealed no difference in gender ratios. The AM and PM classes were slightly different in areas of housing, but most were still in on-campus dorms. Considering that this is educational research, the similarities were greater than any usually seen in such studies.

### Table 1

**Freshman Academy Biology 100 student statistics**

<table>
<thead>
<tr>
<th></th>
<th>AM Class</th>
<th></th>
<th>PM Class</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
</tr>
<tr>
<td>Gender</td>
<td>75</td>
<td>40</td>
<td>50</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>65.20%</td>
<td>34.80%</td>
<td>54.30%</td>
<td>45.70%</td>
</tr>
<tr>
<td>HS GPA</td>
<td>3.77</td>
<td>3.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACT</td>
<td>27.44</td>
<td>28.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Housing</td>
<td>Non-Helaman</td>
<td>Helaman Halls</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Methods**

The type of learning a student engages in is influenced by the type of assessment used in a particular class. When the assessments are designed to test the inter-relatedness of concepts, students will learn on deeper levels (Briscoe & LaMaster, 1991). If students are to learn deeply, they must be assessed deeply and then instructed deeply. The methodology of this research is rooted in the goal of helping students gain a greater connected understanding of biology, rather than memorizing a myriad of disconnected facts.

**Control**

Because both classes came from Freshman Academy, the samples examined came from the same population. Measures were taken to ensure that other elements of the classes (besides concept mapping and
article portfolios) were kept consistent. There are several components to Biology 100 that were incorporated in both the control and treatment classes in this study. These are detailed below:

**Service Learning.** Service Learning is a teaching strategy that was incorporated in both control and treatment sections. One of the class goals was to encourage civic engagement and awareness; opportunities were provided for students to go out into the community and serve in biology-related areas. Four areas of service were available to the students of each class: 1) The Huntsman Senior World Games, 2) Adaptive Aquatics, 3) Elementary Education, and 4) Natural Resource Conservation. Students in both classes were required to write a reflection paper about their experiences in service learning.

**Instrumentation.** Four tests were given throughout the semester to both classes. The objective portion of each exam was identical between the two sections. These exams consisted of 90 objectively graded multiple choice/matching items, and were completed by the students in the BYU testing center (there were a few exceptions made for students with conflicts, in which case, the exams were taken out of the testing center). The written portion for each class differed slightly (the treatment class receiving components on concept mapping and the control class receiving components on current events), and was therefore, part of the treatment.

**Labs.** Each of the large class sections were divided into smaller sections of 30 students each, although some sections in the afternoon class had fewer than 30 students). All sections met once weekly with their teaching assistants in a mandatory lab class. These lab classes were taught by graduate and undergraduate TAs (out of the 8 lab sections, 5 were taught by graduate TA’s and 3 were taught by undergraduate TA’s). The treatment class had three sections taught by a graduate student and one by an undergraduate student; the control class had two sections taught by a graduate student and two sections taught by undergraduates.

The purpose of these labs was to review and clarify material that had been taught during the previous week. Labs began with a short quiz that contributed to the overall grade in the class. Homework assignments and service learning arrangements were also discussed during week labs.

**Instruction.** Dr. Gary M. Booth was the class instructor for both the control and treatment sections. His instruction was intended to be constant between the two sections. Dr. Booth has taught biology for 34 years and has developed a teaching style that has become quite consistent.

Dr. Booth also incorporated pair share learning in both classes. At the beginning of each lecture period, he gave the students a brief quiz, usually covering the material from the previous class period. At the conclusion of this quiz, he allowed the students to share their answers, to encourage peer instruction. Finally, he provided an opportunity for students to share their answers in front of the whole class.

Although it could be argued that no teacher ever teaches the same material twice in the same way, Dr. Booth’s methods were consistent enough that they generally did not vary between the morning and afternoon classes.

**Research Paper.** Each student was required to write a 5-7 page research paper, regardless of which class he/she was in. Students were instructed to select a topic relating to their service learning. Students from both classes had access to TA assistance, the BYU writing lab, and to Dr. Booth. These research papers were graded by their individual lab TA’s using a grading rubric to decrease grading biases (Appendix A).

**Office Hours.** Each TA (four per class for a total of eight) was required to have two office hours per week. TA’s were allowed to pick a time that worked with their own schedules, but they were highly encouraged to select hours that would allow most of their students to attend if needed. These office hours were designed to provide extra feedback on homework assignments and clarifications on difficult material. The only difference in the office hours between the two classes was the type of homework the students were
seeking help with. Because each class had a different type of homework, the treatment class members were encouraged to speak to their TA’s for clarifications and extra points on their concept maps and subject matter, while those in the control sections were encouraged to seek help on their article portfolios and general subject matter.

**Homework.** Although the nature of the homework assignments between the two classes was different, the amount of time required by each student per week remained relatively equal. The treatment class’ assignments consisted of concept mapping; they completed a variety of assignments ranging from the completion of skeletal maps (partially completed maps) to building their own maps from a list of given terms. Thus, some weeks required more time than others.

The control class was also designed to help students make connections between biology and their everyday lives. Each week, the students were to survey the current literature and media, looking for articles that involved biology. These articles could be found in scholarly journals or in articles found in the popular media (to increase their awareness and ability to critique biology in their “every day lives”). Each student was required to find three articles per week, and to write a paragraph to accompany each article.

**TA Review Sessions.** Before each exam, all eight TA’s (four for each class) scheduled a test review session. These sessions were one to two hours long, and were made available to any of the student in either section. Thus, TA’s from the morning class gave reviews to students in both the morning and afternoon classes and visa versa. In this way, potential biases were eliminated.

**Treatment**

**Concept Mapping.** The morning Biology 100 class received the treatment in this design while the afternoon class was designated as the control. The students were blinded to the treatment, and were not aware that concept mapping was part of research being done. An emphasis on connected understanding was given in both the morning and afternoon sections. The treatment class had concept mapping as an assignment to reinforce that emphasis, while the control class prepared current event article portfolios. It was hypothesized that concept mapping would be a more effective method for helping students gain a greater connected understanding between biological concepts and between biology and themselves. Many students enter this type of biology class (general education, non majors) with the intent of simply surviving, or of getting a grade and moving on. The goal of this class was that students would be able to see the bigger picture of biology while still understanding how the details fit together; to minimize the “memorize and regurgitate” phenomenon.

Concept mapping strategies focused on: (1) training, (2) homework assignments, and (3) testing. The training took place during the lab in the first week of class and was ongoing throughout the semester. A handout outlining concept mapping was given with simple examples the students could relate to (Appendix B). The first few maps the students completed were “skeletal maps” (Appendix C), in that they had some of the concepts and links provided. Students were to complete the maps in the structure they were given. TA’s graded these maps according to a detailed protocol and gave feedback to the students; additional concept map training specific to each student was achieved in this way.

As students became more familiar with concept maps (within the first three weeks of class), we then introduced them to CMAP Tools software (Cañas, n.d.) they could use to construct their own maps from a list of terms we gave them (Appendix D). Student homework assignments were done individually, and included a combination of skeletal and self-constructed maps each week for a total of 20 maps throughout the semester. On each exam (excluding the final), the students were given two concept mapping questions to complete (Appendix E). Students in the control section were given questions regarding their article summaries in place of concept mapping questions.
**Reflection.** Reflection was an important part of the concept mapping treatment. After the completion of each map, students would meet with their TA during office hours to assess their performance. Incentives were provided for these meetings by offering returned points for those they had missed; if a student missed a link in their concept map and then discussed it with the TA (finally being able to accurately describe that link), they could receive those points back. This component of the design encouraged students to think more about the connected understanding than the concept map itself; the map became a tool of understanding rather than a mechanism of punishment if they did not understand how to map well.

**Grading.** Each week the four TA’s of the treatment class would meet and discuss the concept mapping assignments for the upcoming week. Master maps had been created by a graduate student the prior summer, and skeletal maps were constructed from these with the input of the TA’s. The TA’s would then complete the maps by themselves. Basic training on rating maps was given prior to any student assignments (Appendix F). To create a specific rubric for each mapping assignment, each TA’s would grade the same students’ map and then as a group, discuss why points were or were not given. In this way, a consensus of expectation was reached and reliability increased. The same process was used for the creation of a student-constructed mapping assignment.

Although the grades of the concept maps did not directly impact this study, they did insofar as they served to motivate the students to spend time and thought creating them. The maps were weighted heavily enough in the course’s grading system (15% of their total grade) that students had to participate in order to do well in the class.

**Measures**

**Essays**

Carefully constructed restricted response essay questions were used as a measure of connected understanding in both the treatment and control classes. These essay questions were designed by two graduate students (Julie Brinkerhoff and Ken Plummer) and supervised by Dr. Richard Sudweeks (Department of Instructional Psychology – BYU). The questions were designed specifically to assess connected understanding between concepts, and were closely aligned with the interview questions. The essay questions used in this study were administered as part of each exam (excluding the final). Two essay questions were given on both midterm exams.

Rubrics for the essays were discussed/debated by the TA’s and the graduate students. Sample essays were graded by each and a consensus reached as to the expectation for points. The scores used to compare the two classes were not the actual scores given to the students during the semester, but were scores obtained from ratings done by the TA’s that were independent of the class. The TA’s who did the ratings for the study had two years of training in grading connected understanding, and spent more time and effort on a fewer number of maps in order to be more accurate. Each essay (and interview transcript) was rated by four TA’s on two separate occasions.

**Interviews**

Interviews were scripted and designed to be strictly student-response. Interview questions that aligned with the essay questions from the exams were designed; interviewers were instructed to read the questions and provide no help in answering them. The interviewers could read the questions as many times as the student needed, but they were not to offer hints of any kind (as recommended by Novak (1984)). Several questions were combined to make up a “task”; each of the two interviews had two tasks for a total of four interview tasks.

The interviews for the entire semester were conducted by three people: two biology TA’s (one a female sophomore, the other a junior pre-dental student) and a nursing student (the spouse of one of the
TA’s). All had a background in biology, and all were responsible for transcribing the interviews he/she conducted. From the transcriptions, the TA’s designed rubrics and rated each transcript (and essay) on two different occasions.

Sampling

This study examined a random sample of students within the two classes. Stratification was used after testing to be sensitive to the possibility that concept mapping may have different effects on students in different achievement quartiles (for example, that the “C” and “D” students may benefit more from concept mapping than the “A” and “B” students, etc.). Due to resources, a sample of 68 students was taken throughout the semester (34 students on two different occasions). Each sampling consisted of 24 students from the morning class and 10 students from the afternoon class; a different group of students was sampled after each exam (therefore, no multivariate tests could be conducted examining student progression over time). The difference in sample sizes between the treatment and control sections was due to an additional study that was coupled with this research (this study examining concept mapping as an instructional tool, the other examining it as an assessment tool); additional money contributed by the other researcher provided for a larger sample of the treatment group.

The three stratification categories (above average, average, and below average) were determined after each exam was completed by the students. The determination of above average, average, and below average students was done by a qualitative sorting. After the completion of each exam, 2 TA’s sorted the essay questions based on their initial appearance. One would sort the exams into three piles, representing the three categories. The piles were marked to indicate the selection of that particular TA. Following this procedure, a second TA also sorted the piles (being unaware of the selections made by the previous TA). The second TA’s sorting was also marked. The tests were finally divided into the three categories, however, only those tests upon which both TA’s had agreed were included in the final selection pool. An online random number generator was then used to select which exams were extracted from the three categories.

Students that were randomly selected signed an informed consent form allowing the use of their essay scores and interview transcripts for research purposes. Each received a $10 gift card to the BYU Bookstore as compensation for their time to interview (which was not required as part of the class).

Statistical Analysis

The data were analyzed using 2-way ANOVA for treatment and gender. As each rater graded each essay and interview task twice, the averages of the rater’s scores over both grading periods were used to make comparisons between the treatment and control sections. If the treatment gender interaction was not significant, one-tailed results were examined. If there was a significant interaction between the treatment and gender, two tailed results were looked at. For analysis of the lower quartile, a modified Levene’s test was conducted (with two factors: treatment and gender). Examination of the left tail was used to determine differences between the treatment and control groups. Significance was $\alpha \leq 0.05$.

Results/Discussion

From Pilot to Actual Study: Changes in Methods

The design of this experiment evolved as understanding deepened (Santhanam et al., 1998). A pilot study the year prior to the actual study (during Fall semester 2005), provided essential information that improved the experimental design. Changes were made in the training, grading, implementation, and emphasis of concept mapping between the two years of study; the result was a more refined study in which stronger inferences could be made.

Second Year Study Results & Discussion

Each rater rated each essay and each interview on two different occasions (with three exceptions for absences). Inter-rater reliability was estimated by computing a mean for each rater across the two rating
occasions. The mean ratings for the various raters were then analyzed by the SPSS Reliability procedure to assess the internal consistency of the average ratings. Table 2 displays the internal consistency coefficients for each of the essays and each of the interviews. The resulting inter-rater reliability coefficients are generally high. That is, 14 of the 16 exceed .90.

Table 2
Inter-rater Reliability Coefficients

<table>
<thead>
<tr>
<th>Inter-rater Reliability Coefficients</th>
<th>AM</th>
<th>PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essay Task 1</td>
<td>.98</td>
<td>.95</td>
</tr>
<tr>
<td>Essay Task 2</td>
<td>.92</td>
<td>.69</td>
</tr>
<tr>
<td>Essay Task 3</td>
<td>.94</td>
<td>.98</td>
</tr>
<tr>
<td>Essay Task 4</td>
<td>.94</td>
<td>.93</td>
</tr>
<tr>
<td>Interview Task 1</td>
<td>.90</td>
<td>.95</td>
</tr>
<tr>
<td>Interview Task 2</td>
<td>.91</td>
<td>.97</td>
</tr>
<tr>
<td>Interview Task 3</td>
<td>.98</td>
<td>.98</td>
</tr>
<tr>
<td>Interview Task 4</td>
<td>.97</td>
<td>.85</td>
</tr>
</tbody>
</table>

Analysis of the essays and interviews revealed that there was a general trend between the control and treatment groups that was consistent with the research hypothesis (that the treatment class would have higher scores). The degree of significance varied between essays and interviews; interviews had more statistically significant differences than the essays.

In order to analyze the lower quartile students of each class, an examination of the lowest 6 students (of the sampled 24) in the treatment class and the lowest 3 students (of the sampled 10) in the control class was made. Table 3 details the results for the Essay Tasks by domain; Table 4 details results for the Interview Tasks.

Table 3
Essay Means and Standard Errors for the Treatment and Control Groups

<table>
<thead>
<tr>
<th>Domain</th>
<th>Gender</th>
<th>Treatment</th>
<th>Control</th>
<th>Significance (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SE</td>
<td>M</td>
</tr>
<tr>
<td>Enzymes</td>
<td>Males</td>
<td>8.80</td>
<td>0.67</td>
<td>8.15</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>8.07</td>
<td>0.47</td>
<td>7.58</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>8.44</td>
<td>0.45</td>
<td>7.87</td>
</tr>
<tr>
<td></td>
<td>Males</td>
<td>6.57</td>
<td>0.56</td>
<td>5.65</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>6.84</td>
<td>0.42</td>
<td>6.53</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>6.70</td>
<td>0.40</td>
<td>6.09</td>
</tr>
<tr>
<td></td>
<td>Males</td>
<td>2.72</td>
<td>0.74</td>
<td>4.36</td>
</tr>
<tr>
<td>Domain</td>
<td>Gender</td>
<td>Treatment</td>
<td>Control</td>
<td>Significance (p-value)</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>-----------</td>
<td>---------</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SE</td>
<td>M</td>
</tr>
<tr>
<td>Enzymes</td>
<td>Males</td>
<td>4.57</td>
<td>0.30</td>
<td>4.75</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>4.48</td>
<td>0.26</td>
<td>3.18</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>4.53</td>
<td>0.24</td>
<td>3.97</td>
</tr>
<tr>
<td>Energy</td>
<td>Males</td>
<td>7.60</td>
<td>0.56</td>
<td>6.15</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>7.87</td>
<td>0.30</td>
<td>6.32</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>7.73</td>
<td>0.31</td>
<td>6.23</td>
</tr>
<tr>
<td>Proteins</td>
<td>Males</td>
<td>2.72</td>
<td>0.71</td>
<td>3.64</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>4.97</td>
<td>0.45</td>
<td>2.84</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>3.84</td>
<td>0.44</td>
<td>3.24</td>
</tr>
<tr>
<td>Immunity</td>
<td>Males</td>
<td>3.02</td>
<td>0.85</td>
<td>3.34</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>4.61</td>
<td>0.42</td>
<td>3.34</td>
</tr>
<tr>
<td></td>
<td>Combined</td>
<td>3.81</td>
<td>0.45</td>
<td>3.34</td>
</tr>
</tbody>
</table>
Analysis of Interview Task 1 (domain of enzymes) showed a significant difference between the regular treatment and control class (p=0.03; Table 4, Figures 5 & 6).
Analysis of the lower quartile on Interview Task 1 also displayed a significant difference between the two groups (p=0.03). The treatment class clearly scored significantly higher on the interview questions than the control class (Table 4; Figures 7 & 8).
Interview Task 2 (domain of energy) was highly significant; p=0.01 in the regular class (Table 4; Figures 13 & 14). The lower quartile with a p=0.51 (Figures 15 & 16) was not statistically significant. The differences between the treatment and control classes were most apparent in the regular class on Interview Task 2.
Interview Task 2 covered the domain of energy. The significant results obtained here could possibly be due to the nature of the content material. This energy domain includes the processes of glycolysis, Krebs Cycle, and oxidative phosphorylation. It is much more complex than the domain of enzymes; it requires students to understand processes that many of them (as non science majors) have not thought about previously and may therefore have been more discriminating between the treatment and control groups.

In Essay Task 3 (domain of proteins), there was no significant difference between the entire sections; however, there was a gender/treatment interaction that was significant (p=0.04). Females in the treatment class scored significantly higher than the females in the control class.
Essay Task 3 - AM Class

Points Awarded

Figure 17: Essay Task 3 – Treatment Class

Essay Task 3 - PM Class

Points Awarded

Figure 18: Essay Task 3 – Control Class

Essay Task 3 - AM Class - Lower Quartile

Points Awarded

Figure 19: Essay Task 3 – Treatment Class – Lower Quartile

Essay Task 3 - PM Class - Lower Quartile

Points Awarded

Figure 20: Essay Task 3 – Control Class – Lower Quartile
The Interview Task 3 also displayed a gender/treatment effect (p=0.04); the females of the treatment group scored significantly higher than the females in the control group.

Figure 21: Interview Task 3 – Treatment Class

Figure 22: Interview Task 3 – Control Class

Figure 23: Interview Task 3 – Treatment Class – Lower Quartile

Figure 24: Interview Task 3 – Control Class – Lower Quartile
Essay Task 4 (domain of immunity) yielded the most perplexing data of the study. There was a significant difference, but in the opposite direction of that hypothesized. With a p-value of 0.045, the control class was significantly higher in scores than the treatment class (Table 3; Figures 25 & 26). A gender/treatment interaction here showed that males in the treatment group scored lower than males in the control group (p=0.04).

Figure 25: Essay Task 4 – Treatment Class

Figure 26: Essay Task 4 – Control Class

Figure 27: Essay Task 4 – Treatment Class – Lower Quartile

Figure 28: Essay Task 4 – Control Class – Lower Quartile
The males in the control class scored higher than the males in the treatment class in Essay Task 4 (domain of immunity). This result was contrary to expectations and could have resulted from differences not related to concept mapping. The control class was writing current event articles throughout the semester; in essence, they were practicing essay skills that the treatment group was not. Essay Task 4 was given later in the semester, when the pace of the material tends to be more rushed than that covered earlier in the semester. Some research has indicated that there is a difference in male and female achievement in science (Lupart et al., 2004; Schober et al., 2004). Perhaps the control class (with a higher percentage of male students) was more able to handle the quicker pace. There could have been a difference in the review sessions taught by various TA’s in preparation for the exam as well. It could also be that the question was given late enough in the semester that the afternoon class had had adequate practice in writing essays about connected understanding; in other words, a learning effect had taken place.

Figure 29: Interview Task 4 – Treatment Class

Figure 30: Interview Task 4 – Control Class

Figure 31: Interview Task 4 – Treatment Class – Lower Quartile
Excluding Essay Task 4, analysis of all Essay and Interview Tasks show that the treatment class scored higher than the control class. Statistically significant results were obtained in seven of the sixteen measurements. All of the significant measurements (except Essay Task 4) showed that the treatment class scored higher than the control class.

The results of this study could have been influenced by the small sample size and the difference in treatment and control sample sizes. A larger, more equal sample could yield results with more statistical power and potentially, more significant differences.

Small sample sizes are common when interviews are used as a measure in educational research (Briscoe & LaMaster, 1991; Hay, 2007; Mintzes & Wallace, 1990). Often, small sample sizes lack statistical power; however, they can still be meaningful. An interview sample size of 34 is above average compared to the above mentioned studies, but with only 10 interviews in the afternoon class, this study could also have more statistical power if replicated with a larger, equal sample between the two groups. As with other studies, despite a potential lack of statistical power, the general trends of the data are meaningful.

The overall trend of the data supports the assertion that concept mapping makes a difference in student learning (Able & Freeze, 2006; Hay, 2007; Horton et al., 1993). In all but one of the sixteen measurements taken, the treatment class scored higher than the control class; in five of the sixteen measurements, the treatment class (either as a whole or the females within the treatment class) scored significantly higher than the control class. Even those results that were not statistically significant could indicate meaning and a need for additional study to strengthen inferences.

Analysis of Essay and Interview Task 3 showed that females in the treatment scored higher than females in the control. These findings suggest that concept mapping could be especially helpful for female students. Laight in 2004 indicated that there were no learning styles that were more or less likely to prefer concept mapping; however, our data support the inference that there could be preferences and advantages to females using concept maps. Other researchers have also found that gender differences exist in the use of concept maps at the college level, often favoring females (Martin et al., 2000; Okebukola, 1992b; Pearsall et al., 1997; Thompson & Mintzes, 2002).

One significant result was found in the examination of the lower quartile students (Interview Task 1; Table 4, Figures 7 & 8). This finding supports evidence from other studies that claim that concept mapping is beneficial to students of average or lower achievement (Guastello, 2000; Snead, 2007). The lower-scoring students in our study who concept-mapped scored significantly higher on this interview task than did their non-concept mapping counterparts. Further research focused completely on the lower quartile students could provide additional evidence for this inference.
Interviews have been commonly used throughout the concept mapping literature (Briscoe & LaMaster, 1991; Heinze-Fry & Novak, 1990; Markow & Lonning, 1998; Mintzes & Wallace, 1990; Nicoll et al., 2001; Novak & Musonda, 1991). The findings of our study provide additional evidence that interviews can be an effective measure of connected understanding. Four of the seven statistically significant results were found in the interview process. Many students find it easier to talk meaningfully about a subject than to write meaningfully about it. Gender could also be a consideration; the treatment class sample had a higher ratio of females to males. Perhaps females were more effective than males at expressing themselves verbally.

Essays have not been used in the literature to measure connected understanding when used with concept maps. Our study suggests that it is perhaps difficult to measure the effectiveness of concept mapping through essay questions. Possible explanations could include the fact that many entering freshman students struggle to express themselves in writing (Lackey, 1997; Rooney, 2000). Many have a negative affect towards essay questions and writing in general. Essay questions, if carefully constructed, can measure connected understanding; however, many students (especially new freshman) are not accustomed to reading essay questions carefully enough to answer exactly what is being asked. Many assume they understand a question when in reality, they do not. Lack of experience and ability to express ideas and understanding in writing could also have contributed to the lack of effectiveness in measurement in the present study.

One of the evidences used to establish a case for causation is that there is a dose-response relationship that is controllable. When applied to this study, a positive dose-response relationship would indicate that the treatment class scores should become increasingly higher than the control class scores. In science education research, there is often a learning factor, or a skill build-up that needs to be considered. Even though the control class was not concept mapping throughout the semester, they were being tested on connected understanding on all of their exams. Students in the control class likely increased in their familiarity and ability to do the tasks on the exams over time. Because of this, the results are not simply measures of the students’ connected understanding, but of their learning abilities as well. This could also explain why there was not necessarily a consistent growth in the gap between the treatment and control classes.

Gender was not a factor in the sample selection process; we did not select based on gender in either the treatment or the control class (Table 19). Arguments could be made for differences between genders in verbal abilities and science aptitude. Historically, the sciences have been male dominated; males tend to be more successful in school science settings (Lupart et al., 2004; Schober et al., 2004). If this is indeed true, the fact that the treatment class (with a higher percentage of females) scored higher than the control class adds impetus to the claim that concept mapping improves learning. It is possible that concept mapping helped to overcome a gender effect. Difference in verbal expression abilities and/or patterns could have been a factor as well.

| Table 19 |
| Class and sample gender statistics |
| Gender Distribution | Treatment | Control |
| Females | Males | Females | Males |
| Regular Class | 65.2% | 34.8% | 54.3% | 45.7% |
| Sample Taken | 75.0% | 25.0% | 60.0% | 40.0% |
BYU Freshman Academy Biology 100 is not typical of most general education biology classes. Concept mapping was only one part of an epistemological approach designed to involve students in the learning process. Even in the control group, pedagogies of service learning, pair share learning, and daily interactive assessments were included. Essentially, concept mapping was added to a class filled with other innovative pedagogies. As such, concept mapping may have had less of an impact than if it had been employed as a solitary intervention. Differences between the two classes were seen even with these other techniques; thus, the impact of concept mapping may be even greater than that documented. Conversely, there are claims that concept mapping is most successful when used in combination with other pedagogies designed to move students towards meaningful learning (Kinchin, 2001). Teachers that are most likely to use concept mapping in their classes are those who will use them in combination with other innovative techniques.

Conclusions

This study provides additional evidence that concept mapping really can help students to learn more meaningfully (Able & Freeze, 2006; Briscoe & LaMaster, 1991; Heinze-Fry & Novak, 1990; Kinchin, 2001). Most entering college freshman do not think in terms of connected understanding; they have been accustomed to memorizing and regurgitating bits of information and recallable facts (Briscoe & LaMaster, 1991). Because of the different emphasis concept mapping brought to the treatment class and because of weeks of practicing a new way of thinking, the treatment class developed skills in making connections between various biology concepts.

One piece of data supports that assertion that concept mapping is particularly helpful to students in the lower quartile. In Interview Task 1 (domain of enzymes) the lower quartile students in the treatment class scored significantly higher than the lower quartile students of the control. Additional investigation into this sub-population in a future study could further substantiate this claim.

The evidence in this study suggests that concept mapping may be most helpful to females in non major’s science classes. Two statistically significant pieces of data demonstrated that females who concept mapped scored significantly higher than their non concept mapping peers. The overall trend of these students in the treatment and control groups show that the students who used concept mapping were able to score higher than their non-concept mapping peers.

A combination of pedagogies designed to involve students in the learning process is most effective. Concept mapping is effective when it is embedded within a larger curriculum based on an epistemological philosophy centered in meaningful learning. Teachers that incorporate concept mapping in their classes in combination with other innovative pedagogical tools will likely have higher student achievement than their more “traditional” counterparts.

Finally, further research would strengthen the inferences made based on the results of this study. Future investigations could examine a larger, more equal sample size, and thus have more statistical power regarding the effectiveness of concept mapping in developing connected understanding. Research following one sampled group of students throughout the course of an entire semester would also likely yield additional evidence to validate the use of concept maps in classrooms. If one group of students could be examined longitudinally and gains measured throughout a course, support from a repeated measures model would be added.
References


Appendix B
Dr. Booth’s Biology 100 Concept Map Training

1. Purpose of Concept Mapping

In order to “really” understand a subject like biology, you need to know

- Biological Facts – Watson and Crick discovered the structure of DNA.
- Biological Concepts – DNA or mRNA
- How biological Concepts are Interrelated – mRNA is a mobile transcription of DNA

In Dr. Booth’s class this Fall you will learn facts, concepts, and how these concepts interrelate with one another. We will use Concept Mapping as a way to see how the concepts you learn this semester are interrelated.

2. How to Concept Map

A concept map has –

- Concepts - that you write with a word in a circle

- Linked Concepts – which communicate a “complete thought”. The arrow shows the direction of the thought.

3. Linking Phrase Training

Kinds of linking phrases:

There are two kinds of linking phrases you will encounter in your future assignments

- Link #1. One that is true and informative (because you might not have known it before)
  
  For example,
• **Link #2.** One that is true but not very informative (because you probably already knew it)

  *For example,*

  Many of these obvious links tend to be those links that give the concept map its structure. For example, organs are made up of tissues which are made up of cells which are made up of molecules. Notice that this information is fairly obvious, but it does provide the organizational structure of the map, like an outline. After you have the organizational structure in place (link #2), then you may add concepts that create more informative linking phrases (link #1).

**What makes a helpful linking phrase:**

The linking phrase should communicate the *essential relationship* between two concepts with *as few of words as possible*. For this to happen, you usually will not add any new concepts in the linking phrase, unless absolutely necessary. Read the following examples:

**Good example**

Notice that the concept “electrons” is in the linking phrase, because it is a critical part of the relationship between the concepts “bonds” and “covalent”.

**Bad example**

This link may be true, but it could be a bit more informative.

**Bad example**

This linking phrase could be shorter and still communicate the essence of the relationship between the two concepts. It might be better to break this link apart and make connections between concepts like “octet rule”, “outer valence”, “bonds”, and “covalent” in a larger concept map as shown below:
Read the following paragraph and then fill in the blanks below. Note that a concept (circles below) or linking phrase (lines below) may be blank. All matter is composed of compounds. Compounds are made up of a combination of different atoms, held together by bonds. Bonds come in a variety of forms, depending on how the electrons of the atoms interact. In an ionic bond, the electrons from one atom are transferred to another atom. In a covalent bond, the electrons are shared. In hydrogen bonding, weak bonds are formed through the attraction of hydrogen atoms. The number of electrons that are available to form bonds are found in the valence shell of the atom. The valence is determined by the number of electrons in the outer shell. Electrons (negatively charged), protons (positively charged), and neutrons (no charge) are all particles contained in an atom. An element is composed of identical atoms.
Appendix D – Concept Mapping Homework from a list of terms.

Lecture 7 – Concept Mapping Homework Assignment

Construct a concept map including each of the concepts listed below. You are only accountable for making connections based on the information written in LECTURE 7 of the syllabus. Remember that some connections will be very basic and simple and other connections will require careful wording. Please refer to previous mapping assignments if you need help remembering what should go in a link.

You may find the following approach helpful.

1. BEFORE YOU MAP – read the syllabus
2. BEGIN MAPPING
a. Put the most general concepts at the top of your C-Mapping document
b. Put more specific concepts below more general concepts throughout the map
c. Link each concept with a line
d. Write linking phrases (with arrows) on each line

3. REFINING YOUR MAP
   a. Reread the syllabus, specifically checking the wording of each linking phrase.
   b. Restructure your map until you feel like it clearly communicates each relationship as taught in Lecture 7.

List of Concepts

ADP
ATP
Cellular Respiration
Cytoplasm
Energy
Glucose
Glycolysis
Krebs Cycle
Inner Mitochondrial Membrane
Mitochondria
Organism
Oxidative Phosphorylation
Pyruvates
Terminal Phosphate
Appendix E

Exam 1 Concept Map Task 1 (Constrained Constructed Map)

**Instructions:** Construct a concept map showing how the concepts in the list below are interrelated. You are required to use every concept in the list below. **Do not add additional concepts.**

Activation Energy
Active Site
Anabolic Process
ATP
Catabolic Process
Cell
Competitive Inhibitors
Energy
Enzymes
First Law of Thermodynamics
Metabolism
Molecules
Noncompetitive Inhibitors
Second Law of Thermodynamics

Remember make as many connections as you like, but be sure that each linking phrase is as accurate as possible to avoid being penalized for an incorrect linking phrase. *(Total possible points 10)*

*Each linking phrase will be scored individually using the following accuracy scale:*

<table>
<thead>
<tr>
<th>Quality of Proposition</th>
<th>Descriptions and Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accurate and Essential (2 points)</td>
<td>Communicates accurately the essential relationship between two closely related concepts without adding extra information.</td>
</tr>
<tr>
<td>Accurate but Incomplete (1 point)</td>
<td>Communicates accurate information between two closely related concepts but is missing some essential information.</td>
</tr>
<tr>
<td>Accurate but weak connection (0 point)</td>
<td>Communicates accurate information between two concepts that are not closely related.</td>
</tr>
<tr>
<td>Don’t Care or Inaccurate (-.5 points)</td>
<td>Does not communicate an understanding of the relationship between the two concepts or information is inaccurate.</td>
</tr>
</tbody>
</table>