# Testing Computer Based High School Science Modules in a Classroom Setting

James W. Bartholomew Jr. Department of Geological Sciences University of South Carolina, Columbia South Carolina 29208 J.Bart@sc.edu

John R. Carpenter Center for Science Education and Department of Geological Sciences University of South Carolina, Columbia South Carolina 29208 Jpawleys@AOL.com

Thomas J. Owens Department of Geological Sciences University of South Carolina Columbia South Carolina 29208 owens@sc.edu

## ABSTRACT

An investigation of the effectiveness of inquiry based teaching modules for instructing a group of high school physics students was conducted in 5 classrooms. Four teaching modules were tested on junior and senior level high school students. All the modules were in accordance with the South Carolina High School Science standards. Three modules utilized computer manipulation of seismic data and earth physics principles to teach concepts in physics and physical science. All students showed a significant increase of instruction knowledge for all modules. Test scores were unreliable for the modules individually. Combining the four individual tests into a single post-test for all four modules proved to have acceptable internal consistency and to show a significant increase for knowledge level as shown as a difference score between pre- and post-test results.

Keywords: Inquiry, Learning Cycle, High School Science, Computer Based Learning

## INTRODUCTION

The establishment of the South Carolina Earth Physics Project (SCEPP)(Hamburger, 2001) as a South Carolina Commission on Higher Education Center of Excellence (CHE) at University of South Carolina allowed South Carolina to implement and adapt a growing national program at a critical time when our state's grade 9-12 science curriculum was being reexamined and revamped. The SCEPP project has four overall objectives that are being addressed.

Objective 1: High school science teachers will have an increased knowledge and perceived importance of earth physics so as to be better prepared to utilize SCEPP resources to address appropriate science education standards.

Objective 2: High school science teachers will have increased knowledge of and confidence in their ability to use state-of-the art instructional and research technology.

Objective 3: High school students will have increased knowledge of natural science concepts so as to be able to meet or exceed the appropriate science education standards.

Objective 4: High school students will have increased knowledge of and ability to use research technology so as to be able to meet or exceed the appropriate science education standards

SCEPP was designed so the objectives can be used to guide the actions of the project staff. The actions include:

Creating teaching modules that are tied to the appropriate high school science curriculum standards,

Training of teachers to use these modules to increase the awareness of earth physics concepts in teachers,

Having trained teachers use the modules to increase the knowledge level of students, including their ability to make use of state of the art technology.

Testing to evaluate objective 3, to see if students could learn natural science concepts with the use of the SCEPP modules is what this study attempted to answer.

The Princeton Earth Physics Project (PEPP) first implemented the concept of using seismic data to teach high school science (Clouser et al., 1994, Hamberger et al., 2001). While PEPP was successful in encouraging teachers in several states to use seismic data in teaching earth science concepts in their states on a limited basis, it was handicapped by the technology that was available at that time. Downloading data was relatively difficult for teachers and the number of participating sites was limited. SCEPP uses a "high-density" approach by installing digital seismographs in high schools in 32 of the state's 46 counties, thus providing the connection between students' familiar environment and earthquakes that occur throughout the world on a daily basis. These individual high schools are linked, in near real time, via the Internet to a central resource center at The University of South Carolina so that students and teachers from any high school in the state can access SCEPP data and share experiences with other participants. SCEPP is also developing an integrated curriculum from which teachers can utilize data from the SCEPP network and provide pre-service and in-service teachers throughout the state with the training and support necessary to make optimum use of this unique educational resource.

The integrated curriculum available to teachers currently consists of four teaching modules that use the principles of geophysics to teach physical science concepts. All modules use a discovery-based or learning cycle-based instructional format (Saunders, 1992). The module entitled "Introduction to the Carolina Earthquake Explorer" is used as an introductory module, prior to using any other module. This module covers basic seismology, familiarizes the student with seismic data in the form of earthquake traces and the use of the Carolina Earthquake Explorer, a computer program (CEE) that is the central part of the SCEPP approach. The

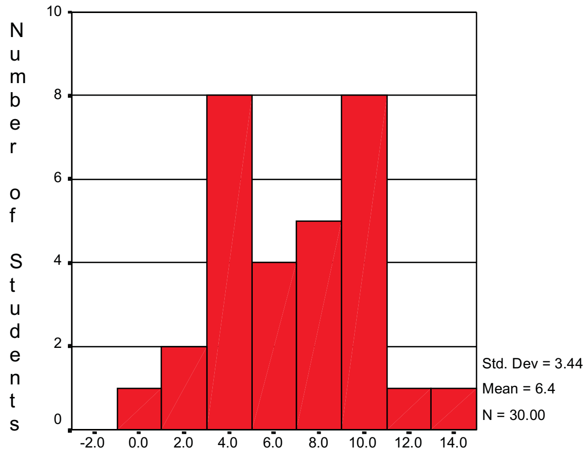


**Figure 1 Difference scores for post-test minus pre-test for the combination of all modules combined. Only data from those students who completed all pre- and post-tests for all four modules were included.**

other three modules can be used in any order depending on how the modules fit into the class curriculum. The modules are entitled, "Detecting the Layered Earth", "Electromagnetism and detecting Earthquakes", and finally "Longitudinal and Transverse Waves". All modules are aligned with appropriate high school science, mathematics and inquiry standards. Among the beauties of integrating seismology and real-time seismic observations into the high school science curriculum is that it provides a foundation from which teachers can address basic principles of both physical science and earth science using a common "horizontally- integrated" resource. Thus, this project has the potential of having a profound impact on the way natural science concepts are taught in South Carolina. While the SCEPP Center of Excellence program will focus on the integration of SCEPP resources and activities into the Grades 9-12 science curriculum, the potential impact of this project on science education in South Carolina is far broader. Through use of the Internet, SCEPP will provide easily accessible seismic data and educational resources that could be utilized throughout K-16 education in the state. Thus, SCEPP also has great potential to become a "vertically-integrated" educational resource for the State of South Carolina.

## METHODS

This study reports on data gathered in connection with Objective 3, i.e. increase student knowledge of usual science concepts. For this study, we selected one school as a pilot site. The school was located in Pickens County near Greenville, South Carolina, and consisted of grades 9 through 12 with a of population approximately 3,000 students. The study had available approximately 110 students in four classes of honors physics and one class of "tech-prep" physics for students not intending to attend college but may want to attend a technical or vocational school. Classes consisted of primarily 12th grade students, with a small number of 11th grade students. The students were given four SCEPP teaching modules to use during the last 5 weeks of the school year. All the modules were taught at the same time to all classes consecutively. The experimental design used in this study was the One-Group pre-/post-test design of Campbell and Stanley (1963). Using this design, students of an experimental instruction treatment group (Experimental Group), are given both a pre-test on the subject to be covered by the treatment and an identical (or equivalent) post-test on the subject matter covered. In this design there is no Control Group of students, on whom a traditional treatment of the same subject matter is used. For our study, a control group was not used because it would double the number of class preparations the cooperating teacher would have to prepare and was considered unacceptable by the cooperating teacher in the study. With no control group, we could not compare the effects of the experimental approach and a more traditional approach on any gains in knowledge. Because of this, the modules were done after the students had already completed subject matter instruction and the school year was nearly over. This design is testing how much additional learning students accomplish using the modules. A set of four pre-tests, one for each module, was given prior to instruction. A post-test was given at the end of each module, before beginning the next modules instruction. Pre-test and post-test instruments were identical and consisted of 8 to 10 multiple choice questions for each module. Each test was taken using university scan sheets to allow the tests to be graded, item analysis to be done and a reliability coefficient to be calculated. The total number of test scores was 400. Item analysis would allow us to identify questions that were not "performing" appropriately. We could then correct or discard inappropriate questions in the next round of testing. The reliability coefficient was calculated to evaluate how well the tests were functioning overall and as a measure of how reliable our conclusions using data from the tests would be. Low reliability would cast doubt on any conclusions about our materials regardless of the scores or other statistical analysis.

| | Mean | N | Std. Dev. |
|---|---|---|---|
| DW Daniels Pre-test | 17.40 | 30 | 4.17 |
| DW Daniels Post-test | 23.83 | 30 | 5.25 |

**Figure 2 Combined pre-test and post-test data statistics. Students who completed all of the pre- and post-tests given were selected out of the overall data. Those tests were combined and re-scored to create the above statistics.**

## RESULTS

The study suffered from a high degree of student absenteeism due to end-of-year activities, testing, etc., causing many students to miss part of a module, a pre-test or a post-test. The analyses, therefore, are based

| | Paired Difference | | | | t | df | Sig (2-tailed) |
|---|---|---|---|---|---|---|---|
| | Mean | Std. Deviation | 95% Confidence Interval of the Difference: Lower | 95% Confidence Interval of the Difference: Upper | | | |
| DW Daniels pre-test - DW Daniels post-test | -6.43 | 3.44 | -7.72 | -5.15 | -10.240 | 29 | 0.000 |

**Figure 3 T-test results for the pre-/ post-test scores. These pre-/post-tests scores are for the combined data from the four individual pre-/post-tests that had low reliability.**

on a final number of students lower than the total number available.

When examining the data available, the key statistic is the Kuder-Richardson or KR20 reliability coefficient (Grunland, 1998). The KR20 coefficient calculates the average split half reliability for all possible split half combinations for a given test. A KR 20 value of 0.7 would mean the tests were reliable enough to draw conclusions on students' learning using the modules in question. For our study, the university computer grading system that we used calculates the KR 20 value. Unfortunately, three of the tests had coefficients less than 0.5. Using the Spearman-Brown formula to correct the coefficients for the individual tests to as high as .64, which did not change our conclusion that the tests were individually unreliable. The scores on the fourth test for the Detecting the Layered Earth module were strongly skewed to the right indicating that nearly all students scored very high on the test, and the KR20 coefficient could not be calculated for this test. The test consisted of graph interpretation questions and was relatively easy for high school seniors. Therefore, because none of the tests we gave had the required reliability to be useful, we cannot answer the question about student learning using these individual tests. The low reliability coefficients were likely due to too few questions (8-10) per test and only an average of 12 students completing both the pre- and post-test for each individual module. Also because these students were seniors at the end of a year of physics instruction, they were already well versed in physics concepts and tended to do very well on all tests. Reliability coefficients assume a normal distribution and this was not the case with all of our tests.

For these reasons, it was decided to combine all the questions into one test and to see if the test was any more reliable. In order to combine the four tests into one larger test, all the tests had to be sorted and those students who had taken all four post-tests had to have their tests recopied onto another scan sheet and re-scored. The number of students dropped to 30 because only this number took all four post-tests. Students were previously assigned a number in order to track them from pre-test to post-test, and maintain anonymity for all participants. By combining the data from all four tests, we created one test of 34 questions that had a much more normal distribution. This resulted in the reliability of the combined test being a very acceptable KR20 value of 0.80.

## CONCLUSIONS

Combining the data and creating the larger test produced a very acceptable reliability coefficient. The question that the study attempted to answer has now changed slightly, however. We cannot determine the individual learning related to any of the modules individually, only as a group.

Figure 1 is a histogram of the difference scores for the large combined data set. It allows us to look at how the modules did as a group. We begin by calculating a paired sample t-test to see if our pre-test and post-test difference scores differ significantly from zero.

Figure 2 shows the mean and standard deviation for the combined data set. Figure 3 shows the results of the paired sample t-test. The result was $t = -10.24$, significant with $\alpha = 0.05$ for a 95% confidence interval. This test shows that on average student scores on the post-test will be between 5 and 7 points higher on a test of 33 questions over the course of all four modules.

It should be noted that the student scores increased 5 to 7 points out of 33 (15% to 20%) after the students had completed nearly a year of instruction in related material. It might have been expected that we would have seen a "ceiling effect" (Carpenter, 1983) in the pre-test scores such that gain scores would show no change but this was not the case. Other factors that potentially could have negatively influenced this outcome were the large drop in the number of students due to end-of-year activities, not to mention a serious drop in the motivation of seniors already accepted into the college of their choice. This is an important point. According to the cooperating teacher, the senior class was pretty much "going through the motions" without any consistent effort on their part. The fact that a significant difference between pre- and post- tests scores occurred despite these difficulties makes this result even more meaningful. Therefore, we believe that the gain in scores can be taken as evidence that students can learn science material using the SCEPP modules in general.

It is possible that other factors influenced the student gain scores. We will likely try to determine if affective factors such as the novelty of using the new software and seismic data might have been an influence. One new study will attempt to determine if student interest might have been enhanced by the use of real scientific equipment and data, and if so, if the enhanced interest might have had an effect on subject matter gains. We will attempt to use a questionnaire consisting of several Likert scale questions to assess student attitudes. Also trying the modules in a 9th, grade physical science class may yield very different results with students who have not received such a large amount of instruction in physical science. Finally, going to one post-test and one pre-test test when testing the modules will help prevent students from becoming bored with taking multiple tests. The project goal that this study could not meet was to determine the individual module effectiveness as a part of regular classroom instruction. If we are able to obtain a large enough group of participating teachers, we will have the luxury of testing just one or two modules at a time and still test all the modules more than once. Unfortunately, without the use of a control group, it will be difficult to answer how well the modules perform when integrated as a part of instruction and not done after instruction has taken place. Separating the effect of instruction from that of the module itself will be impossible without the use of a control group. Again, requesting a few teachers to test only one or two modules in order to reduce the time needed for a control group could enable a new experimental design to be used. All further testing will be done with approximately twice the number of questions for each module pre/post-tests in order to increase reliability for the tests.

## REFERENCES


Campbell, D.T., and J.C. Stanley., 1963, Experimental and Quasi-Experimental Designs for Research, Chicago, Rand McNally College Publishing Co.

Carpenter, J.R., 1983, A Comparison of Student Attitudes, Values and Interests Toward Two Earth Science Courses, Journal of College Science Teachers, v.12, p.441-445.

Clouser, R.H., Nolet, G., Phinney, R.A., 1994, Current status of the Princeton Earth Physics Project (PEPP) Anonymous, AGU 1994 spring meeting, Eos, Transactions, American Geophysical Union, 75 (16 Suppl.), 238 p.

Gronlund, N.E., 1998, Assessment of Student Achievement, 6th edition, Needham Heights, MA: Viacom Co.

Hamburger, M., Pavils, G. L., Phinney, R. A., Stienberg, D., Owens, T. J., Hall-Wallace, M., 2001, New science education initiative brings seismology into the classroom, Eos, Transactions, American Geophysical Union, v. 82, p. 266-267.

Saunders, W.L., 1992, The Constructivist Perspective, Implications and Teaching Strategies for Science, School Science and Mathematics, v.92, p.136-141.