Relationships Between Inquiry-Based Teaching and Physical Science Standardized Test Scores

Thomas R. Tretter
University of North Carolina at Chapel Hill

M. Gail Jones
North Carolina State University

This exploratory case study investigates relationships between use of an inquiry-based instructional style and student scores on standardized multiple-choice tests. The study takes the form of a case study of physical science classes taught by one of the authors over a span of four school years. The first 2 years were taught using traditional instruction with low levels of inquiry (non-inquiry group), and the last 2 years of classes were taught by inquiry methods. Students' physical science test scores, achievement data, and attendance data were examined and compared across both instructional styles. Results suggest that for this teacher the use of an inquiry-based teaching style did not dramatically alter students' overall achievement, as measured by North Carolina's standardized test in physical science. However, inquiry-based instruction had other positive effects, such as a dramatic improvement in student participation and higher classroom grades earned by students. In additional inquiry-based instruction resulted in more uniform achievement than did traditional instruction, both in classroom measures and in more objective standardized test measures.

Issues of educational accountability have had a high profile in recent years in the United States. Emerging from this attention is a renewed emphasis on standardized testing as a means to improve public schooling and to hold accountable the various parties involved. Consequently, many states have adopted a variety of testing and accountability programs; 50 states have state-wide accountability testing of some sort (Council of Chief State School Officers, 2002). North Carolina's testing program has been highly visible, as evidenced by its mention in President Clinton's (1999) State of the Union address. This testing program includes a standard course of study for physical science and other secondary school subjects. Students in each of these core secondary courses are required to take the end-of-course (EOC) test for that subject; this test is a standardized multiple-choice exam developed by the State of North Carolina for use state wide.

The North Carolina standard course of study for physical science (North Carolina Department of Public Instruction, 1994), as is true for many other subjects, covers a wide range of topics students are expected to have mastered. For many teachers, it seems feasible to present all of the required material for a high-stakes, standardized multiple-choice test only by a preponderance of direct instruction (Wideen, O'Shea, Pye, & Ivany, 1997), and many teachers are hesitant to implement instructional practices that do not help cover the required content (Flick, Keys, Westbrook, Crawford, & Carnes, 1997). Costenson and Lawson (1986) interviewed teachers and reported their top 10 reasons for not using inquiry instructional techniques. "Too slow content coverage" came in second, right behind "too much time and energy are required." In a study of the influence of Canadian 12th-grade high-stakes standardized testing, Wideen et al. found that instructional practices were most strongly affected among the 12th-grade teachers whose students were to be tested by the national test. These teachers felt pressured to lecture more in order to cover the extensive content and to help students memorize a string of facts for the final examination.

In North Carolina, chemistry teachers surveyed to assess the impact of EOC testing on their classrooms indicated that the chemistry curriculum was becoming more uniform throughout the state and that teachers were feeling pressure from the test (Smith, Hounshell, Copolo, & Wilkerson, 1992). Jones et al. (1999) also found that the implementation of high-stakes standardized testing in North Carolina significantly impacted classroom instructional approaches. Among other effects, they found that the curriculum was narrowed to the tested material, and more class time was spent practicing for the tests.

The heavy emphasis on direct instructional strategies that teachers report in response to standardized testing is in opposition to the current philosophy of creating student-centered classrooms and student-centered activities (National Research Council, 1996; North Carolina Department of Public Instruction, 1999). The apparent conflict between these two major trends in science education is not new. A 1983 meta-analysis of the effects of "new" science curricula on student performance (curricula stressing science process objectives during the 1960s and 1970s, as opposed to stressing facts, laws, and theories), during a then-current public sentiment for "back to the basics," found that students exposed to the "new" science curricula exceeded the performance of students in traditional courses (Shymansky, Kyle, & Alport, 1983). In spite of many teachers' potential reluctance to use inquiry teaching techniques in a high-stakes standardized testing environment, possibly because of less content coverage, some studies report that the advantages of inquiry learning may offset the disadvantages of slow content coverage (Henson, 1986).

More recent studies also have indicated that inquiry or hands-on science instruction better prepares students, as measured by standardized tests (Chang & Mao, 1998; Kaiser, 1996; Stohr-Hunt, 1996). In each case, however, the authors stressed that the tests used to measure achievement were designed to place greater emphasis on the processes of science as opposed to factual information. These studies notwithstanding, the 1996 National Assessment of Educational Progress (NAEP) science report for North Carolina stated, "Research on the relationship between exposure to hands-on science tasks and overall science performance is sparse and inconclusive" (O'Sullivan, Jerry, Ballator, & Herr, 1997, p. 66). The topic of hands-on science activities was addressed by the NAEP in this report because when the 1996 science framework for the NAEP was developed it took into account the current reforms in science education by including hands-on tasks for each student who participated. Such tasks were not included in the prior 1990 NAEP science assessment.

Considering the impact of both the movement to increase the number of standardized tests and the renewed emphasis on teaching science through inquiry rather than as a set of facts, discerning how these two phenomena interact with each other to affect student learning in a typical classroom becomes vital. The two agendas may seem incompatible, which may generate confusion for the classroom teacher who is trying faithfully to implement a curriculum within the parameters imposed. Are teachers who emphasize inquiry or hands-on instructional strategies having a negative impact on their students' standardized test scores?

Purpose

The purpose of this article is to examine the relationship of an increase in inquiry-based instructional approaches on students' physical science standardized test scores and other classroom measures. The data for this article are taken from an analysis of students' test scores, classroom grades, and attendance records. These data were generated in a natural classroom environment, as opposed to a carefully controlled experimental environment. The goal was to assess the impact of one teacher's use of inquiry-based teaching approaches as implemented in practice over a period of several years.

Participants

The data for this article come from a case study of one of the authors' physical science classes encompassing 4 years. The 1,300-student high school from which this data was gathered is part of an urban school district in a midsized city in North Carolina. Parents of the students were employed mainly in technical, clerical, and blue-collar occupations, with a small percentage employed in professional occupations. Approximately 90% of the students in this school who pursue postsecondary education are first generation college students, and about 30% of the student body participates in the free or reduced price lunch program. The teacher of the classes was certified in both secondary mathematics and physics. Prior to teaching in North Carolina, beginning with school year 1997-1998, he had been teaching secondary mathematics for 7 years in other schools. All classes for this high school are scheduled on a block schedule, so the entire course was taught in blocks of 90 minutes a day for 90 days of one semester.

Assigned to teach all new courses upon coming to North Carolina (primarily physics and physical science), the teacher focused his limited time and energy resources on developing a hands-on approach for the physics curriculum and reverted to a heavily textbook-oriented approach for his physical science classes. After 2 years, he was able to revamp his instruction to an inquiry-based approach for the physical science classes. The teacher decided to implement more of an inquiry-based approach because of dissatisfaction with low student achievement and student disinterest in the course. The teacher was enrolled in an advanced education degree program at the time, which exposed him to detailed recommendations of national science standards and encouraged reflective practice. He cites the combination of advanced coursework and personal experience as providing the impetus to significantly alter his teaching practice.

During years 1 and 2 of the study, seven different physical science classes were taught using traditional approaches with a relatively low level of inquiry-based teaching. These 161 students (52% female and 48% male; 67% African-American, 27% Caucasian, and 6% Hispanic) composed the non-inquiry group. During years 3 and 4 of the study, four physical science classes were taught using inquiry techniques to a much higher degree. These 94 students (50% female and 50% male; 65% African-American, 26% Caucasian, and 9% Hispanic) composed the inquiry group. For both groups, all students were enrolled in standard physical science and were assigned to this teacher's class by the school's computer.

Methods and Data Sources

The implementation of inquiry learning can take different forms in different classrooms. For purposes of this article, an operational definition of the level of inquiry teaching will be based on laboratory work. Along with the differing amounts of time students in the two groups spent in laboratory-based work, the assessment instruments used by the teacher were also significantly modified to reflect more emphasis on experimentation and explanation, as opposed to terminology and description.

Whereas the assessments for the non-inquiry group tended to focus on terminology, definition of formulas, and the use of formulas in well-defined problem solving situations, the assessments for the inquiry group asked students to draw conclusions from demonstrations presented, to develop experiments that could test proposed hypotheses, and to describe everyday situations that could be explained by the science content of the current unit. Table 1 summarizes some of the instructional differences employed between these two groups of students. These data come from an analysis of the lesson plans maintained by the teacher over the course of those years.

Table 1
Group Instructional Differences

Group	No. of Labs	Time in Lab	Days in Lab
Non-Inquiry	14	770 min.	16
Inquiry	31	2,015 min.	39

Note: All values are for the complete course over a 90-day semester.

Although the number and duration of labs are indications of the level of inquiry teaching being used, these data alone do not suffice as proof of an inquiry-oriented teaching style. The nature of the lab work must also be taken into account. Most of the lab work done by the inquiry group was of an exploratory nature wherein students sought to understand and explain underlying physical concepts, and test questions and other assessments were designed to measure students' knowledge accordingly. The lab work done by the non-inquiry group tended to be mostly of a recipe-style format from the textbook publisher, in which students completed a set of instructions on worksheets without necessarily demonstrating an understanding of the underlying concepts. For example, the non-inquiry group was given a set of objects and explicit directions how to go about determining their densities. The inquiry group was asked to devise their own investigations for determining densities of objects and was given the freedom to choose which measuring tools to use and to develop their own procedures, along with modifying those procedures as experience dictated.

Three different aspects of student involvement and achievement were investigated and compared between these two groups: student participation (including attendance); student EOC scores on the physical science standardized test; and classroom grades earned by the students. Student participation was measured by three different factors: (a) the percentage of students who took the EOC test from the total enrollment in the classes at the time of the test; (b) the mean absence rate throughout the course; and (c) the percentage of students who "gave up" on the course as evidenced by an extended time period (2 weeks or more) of not completing or making up coursework and/or classroom tests.

By not showing up to take their final exam (25% of their course grade), students indicated that they did not intend to pass the class and, hence, generally no longer significantly participated in class activities. Absence rate is another indication of the degree to which students withdrew from active class participation. Third, the failure to complete significant amounts of required work over an extended period of time (2 weeks or more) indicated that students had given up on passing the course. The data for these measures were taken from the teacher's attendance records and grade book records.

The EOC score reports for physical science classes contain mean scale scores by class. The results from the non-inquiry classes were compared to those of the inquiry classes. The overall EOC scores can be compared between the two groups by mean scale score and by the percentage attaining the various achievement levels assigned by North Carolina's testing program (Levels 1, 2, 3, and 4). Within North Carolina's system of assessments, levels 3 and 4 indicate proficiency in the subject, whereas levels 1 and 2 indicate nonproficiency.

Results

Participation

Table 2 summarizes the three measures of participation in physical science class between the two groups of students. The percentage of students within each group is given in parentheses for each measure.

Standardized Test Scores

A second aspect investigated in this paper is the relationship between using an inquiry instructional approach and EOC scores. EOC results are reported as a scale score. This scale score is then converted into one of four levels: Levels 3 and 4 are considered proficient; Level 2 indicates students are minimally prepared for subsequent courses; and Level 1 indicates lack of knowledge and skills to succeed in subsequent courses. Both levels 1 and 2 are considered not proficient. Table 3 summarizes these results, with the percentage of students within each group in parentheses.

The residuals computed by the chi-square test suggest that the influential cells accounting for the statistically significant difference in EOC levels between the two groups are the cells at Level 2 (non-inquiry 38%; inquiry 58%).

Table 2
Student Participation in Physical Science Class

Group	n	Taking EOC	Giving Up[a]	Absences[b] (Days/ Student)
Non-inquiry	161	145 (90%)	47 (29%)	10.8
Inquiry	94	90 (96%)	10 (11%)	4.9

[a]"Giving up" is defined as not completing two or more weeks of work.
[b]Mean absence rate in days per student for the 90-day semester.

Students who did not take the EOC test were generally those who were not doing well in the course and, therefore, would likely not do well on the test. Because a larger percentage of students in the non-inquiry group did not take the EOC test as compared to the inquiry group, including only test-takers in Table 3 likely skews the results. While it is impractical to attempt to assign a scale score to those who did not take the test, it is reasonable to assume that students refusing to take the test would have scored nonproficient (level 1 or 2). Based on this assumption, Table 4 shows the number of students at the levels of proficiency or nonproficiency.

Classroom Grades

Another measure of student achievement is the grade for the course for the two groups (see Table 5). This course grade incorporates the EOC exam score as 25% of the total grade, as required by school district policy. Other components of the classroom grade are approximately equally distributed between tests, homework, and laboratory work.

The residuals computed by the chi-square test suggest that the influential cells accounting for the statistically significant difference in grades between the two groups are the ones at Level C (non-inquiry 20%; inquiry 34%) and at Level D/F (non-inquiry 58%; inquiry 37%).

Discussion of Results

The inquiry group students had higher attendance, were less likely to give up, and were more likely to show up to take the standardized test at the end of the course. The inquiry instructional technique seemed to engage students more in the course throughout its duration and, by implication, their attitudes toward the study of physical science were improved. Although difficult to conclude from the participation numbers alone, the teacher commented that the classroom atmosphere was also much improved over that of the non-inquiry group. According to the teacher, the students in the inquiry group posed fewer disruptions and exhibited less antagonistic behavior that diminished the learning environment.

The EOC scores (Table 3) showed a lower overall mean for the inquiry group, although this difference was not statistically significant. The inquiry group may have scored lower because of the nature of the questions asked on the state's multiple-choice standardized test. The skills and concepts learned during inquiry instruction are not as easily testable in a multiple-choice

Table 3
Physical Science EOC Results for Those Taking the Exam

Group	EOC Scale Score[a]			Level[b]			
	Mean	*SD*	*n*	4	3	2	1
Non-Inquiry	52.2	7.68	145	8 (5%)	59 (41%)	55 (38%)	23 (16%)
Inquiry	50.9	6.09	89	3 (4%)	26 (29%)	52 (58%)	8 (9%)

[a] $t(df=232)=-1.328, p<.185$.
[b] $\chi^2(3, N=234)=9.57, p<.05$.
Note: The mean scale score for all physical science students in North Carolina in 1997 was 53.7 with a standard deviation of 9.4.

Table 4
EOC Results by Proficiency

Group	*n*	Proficient	Nonproficient
Non-Inquiry	161	67 (42%)	94 (58%)
Inquiry	94	29 (31%)	65 (69%)

$\chi^2(1, N=255)=2.929, p<.087$.
Note: Assuming nonproficiency for nontesttakers.

format. This postulate is supported by the data in Table 5, which indicate that the inquiry group earned higher grades for the course. The teacher-designed tests were constructed to closely match the inquiry instructional style of content delivery, and students in the inquiry group outperformed the non-inquiry group as measured by classroom assessments.

The spread of the achievement levels in Table 3 indicates that the non-inquiry group experienced a more variable range of success, whereas the more concentrated grouping of student achievement at Level 2 for the inquiry group suggests a more uniform achievement for a broad range of students. This conclusion is also supported by the data in Table 5, which shows that not only were the course grades higher for the inquiry group, but also that they were more tightly clustered than those of the non-inquiry group. One possible explanation for the clustering of inquiry group scores is that an inquiry instructional technique may be more effective for reaching a wider range of students, thus not creating a large gap between those experiencing success and those not.

Limitations

These results should be interpreted cautiously because one of the authors was also the teacher under study. The data were examined for sequential years, and the impact of experience on the instructional quality must be considered. Although the differences in the inquiry and non-inquiry groups are thought to be primarily due to differences in the instructional approach, the potential for researcher bias exists. Researcher bias is controlled for in part by the use of standardized test scores and school absentee rates. It is also possible that variables other than instructional approach may have influenced student outcomes.

Table 5
Course Grades

Group	Percentage Grade[a]			Grade Level[b]			
	Mean	*SD*	*n*	A	B	C	D/F
Non-Inquiry	67.3	20.4	161	3 (2%)	33 (20%)	32 (20%)	93 (58%)
Inquiry	74.7	13.9	94	4 (4%)	23 (25%)	32 (34%)	35 (37%)

[a] $t(df=253)=3.153, p<.01$.
[b] $\chi^2(3, N=255)=11.392, p<.01$.

Conclusions

The use of an inquiry-based teaching style does not seem to be an effective instructional technique if the goal is limited to increasing proficiency achievement as measured by North Carolina's EOC exam in physical science, which focuses primarily on objective knowledge that can be measured with multiple-choice assessment items. However, inquiry-based teaching seemed to have many positive effects, such as dramatic improvement in student participation, higher course grades earned by students, and higher universal achievement by more students instead of a large gap between those experiencing success and those not. If the goals of education go beyond mere test scores and include developing positive attitudes toward the subject matter, then these results suggest that an inquiry-based teaching style may be effective. These results also suggest that a closer examination of the nature and format of the questions on North Carolina's EOC standardized test in physical science may be necessary. Given that the North Carolina standard course of study (North Carolina Department of Public Instruction, 1994) upon which the EOC is based emphasizes science as inquiry as one of the four strands that should permeate all of science education, the EOC may not be appropriate for measuring those inquiry characteristics that North Carolina lists as important.

References

Chang, C.-Y., & Mao, S.-L. (1998, April). *The effects of an inquiry-based instructional method on earth science students' achievement.* Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, San Diego, CA. (ERIC Document Reproduction Service No. ED 418 858)

Clinton, W. J. (1999, January 19). State of the Union address [Online]. Available: http://clinton3.nara.gov/WH/New/html/19990119-2656.html

Costenson, K., & Lawson, A. E. (1986). Why isn't inquiry used in more classrooms? *The American Biology Teacher, 48*(3), 150-158.

Council of Chief State School Officers. (2002). *State education accountability reports and indicator reports: Status of reports across the States - 2002.* Washington, DC: Office of Educational Research and Improvement. Also available online: http://www.publications.ccsso.org/ccsso/publication_list.cfm

Flick, L. B., Keys, C. W., Westbrook, S. L., Crawford, B. A., & Carnes, N. G. (1997, March). *Perspectives on inquiry-oriented teaching practice: Conflict and clarification.* Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, Oak Brook, IL. (ERIC Document Reproduction Service No. ED 407 255)

Henson, K. (1986). Inquiry learning: A new look. *Contemporary Education, 57*(4), 181-183.

Jones, M. G., Jones, B., Gardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan, 81*(3), 199-203.

Kaiser, B. (1996, December). *Teachers, research, and reform: Improving teaching and learning in high school science courses.* Paper presented at the Global Summit on Science and Science Education, San Francisco, CA. (ERIC Document Reproduction Service No. ED 400 191)

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

North Carolina Department of Public Instruction. (1999). *Standard course of study, revised 1999.* Raleigh, NC: Author.

North Carolina Department of Public Instruction. (1994). *Standard course of study, revised 1994.* Raleigh, NC: Author.

O'Sullivan, C., Jerry, L., Ballator, N., & Herr, F., (1997). *NAEP 1996 science state report for North Carolina. Findings from the National Assessment of Educational Progress.* (ERIC Document Reproduction Service No. ED 413 226)

Shymansky, J., Kyle, W. Jr., & Alport, J. (1983). The effects of new science curricula on student performance. *Journal of Research in Science Teaching, 20*(5), 387-404.

Smith, P., Hounshell, P., Copolo, C., & Wilkerson, S. (1992). The impact of end-of-course testing in chemistry on curriculum and instruction. *Science Education, 76*(5), 523-530.

Stohr-Hunt, P. (1996). An analysis of frequency of hands-on experience and science achievement. *Journal of Research in Science Teaching, 33*(1), 101-109.

Wideen, M., O'Shea, T., Pye, I., & Ivany, G. (1997). High-stakes testing and the teaching of science. *Canadian Journal of Education, 22*(4), 428-444.

Editors' Note: Correspondence concerning this article should be addressed to Thomas R. Tretter, University of North Carolina at Chapel Hill, 2701 Homestead Rd. Apt. 1407, Chapel Hill, NC 27516.

Electronic mail may be sent via Internet to tretter@email.unc.edu