A Meta-Analysis of National Research: Effects of Teaching Strategies on Student Achievement in Science in the United States

Carolyn M. Schroeder,[1] Timothy P. Scott,[1] Homer Tolson,[2]
Tse-Yang Huang,[3] Yi-Hsuan Lee[4]

[1]*Center for Mathematics & Science Education, Texas A&M University, College of Science, MS 3257, College Station, Texas 77843*

[2]*Department of Educational Administration and Human Resources, Texas A&M University, College Station, Texas*

[3]*National Hsinchu University of Education, Hsinchu City, Taiwan*

[4]*Department of Business Administration, National Central University, Jung-li City, Taiwan*

Received 5 April 2006; Accepted 5 March 2007

Abstract: This project consisted of a meta-analysis of U.S. research published from 1980 to 2004 on the effect of specific science teaching strategies on student achievement. The six phases of the project included study acquisition, study coding, determination of intercoder objectivity, establishing criteria for inclusion of studies, computation of effect sizes for statistical analysis, and conducting the analyses. Studies were required to have been carried out in the United States, been experimental or quasi-experimental, and must have included effect size or the statistics necessary to calculate effect size. Sixty-one studies met the criteria for inclusion in the meta-analysis. The following eight categories of teaching strategies were revealed during analysis of the studies (effect sizes in parentheses): Questioning Strategies (0.74); Manipulation Strategies (0.57); Enhanced Material Strategies (0.29); Assessment Strategies (0.51); Inquiry Strategies (0.65); Enhanced Context Strategies (1.48); Instructional Technology (IT) Strategies (0.48); and Collaborative Learning Strategies (0.95). All these effect sizes were judged to be significant. Regression analysis revealed that internal validity was influenced by Publication Type, Type of Study, and Test Type. External validity was not influenced by Publication Year, Grade Level, Test Content, or Treatment Categories. The major implication of this research is that we have generated empirical evidence supporting the effectiveness of alternative teaching strategies in science. © 2007 Wiley Periodicals, Inc. J Res Sci Teach 44: 1436–1460, 2007

Keywords: general science; achievement; quantitative

Contract grant sponsor: Texas Education Agency (part of the Texas State-Wide Science Initiative).
Correspondence to: C.M. Schroeder; E-mail: cschroeder@science.tamu.edu
DOI 10.1002/tea.20212
Published online 30 October 2007 in Wiley InterScience (www.interscience.wiley.com).

The No Child Left Behind (NCLB) Act of 2001 calls for the reform of education—including science education—using research-based methods to produce excellence. NCLB establishes math and science partnerships, requires teachers to be ''highly qualified'' in the areas they teach, and requires that only evidence-based programs receive federal funding. States are required to establish academic standards in science and administer standards-based assessments to determine students' science achievement.

Reform documents such as *Science for All Americans* (Rutherford & Ahlgren, 1990), *Benchmarks for Science Literacy* (American Association for the Advancement of Science, 1993), and the *National Science Education Standards* (National Research Council, 1996) establish *what* science should be taught to accomplish the goal of science literacy for all students (Shymansky, Yore, & Anderson, 2004). Researchers, using a Delphi study of experts (Osborne, Collins, Ratcliffe, Millar, & Duschl, 2003), have identified essential elements of the nature of science that should be taught in schools. In their report on the results of a series of seminars addressing the form of science education required for the new millennium, Millar and Osborne (1998) suggested the ''use of a wide variety of teaching methods and approaches'' (p. 23), but did not address the details of these methods and approaches. Duschl, Schweingruber, and Shouse (2005) discussed the importance of teachers' knowledge of instructional strategies for teaching science (or pedagogical content knowledge, or PCK) and pointed to the dearth of research linking PCK to student achievement in science. The questions of pedagogy—of *how* science should be taught and specific strategies for teaching science effectively—have not been definitively addressed in recent years.

In an effort to meet the requirements of NCLB and promote science achievement, the state of Texas established the Texas Science Initiative program to promote rigorous, effective science teaching and to provide highly qualified teachers for every science classroom. In an effort to provide research-based information on effective science teaching, the initiative contracted with the Center for Mathematics and Science Education at a large state university to conduct a meta-analysis of current U.S. research to determine what science teaching strategies have been shown to be effective in improving student achievement in science. This article reports on the procedures and results of that meta-analysis.

Background

Studies included in this meta-analysis were limited to those conducted in the USA because the directive from the Science Initiative was for ''national research.'' These results, nevertheless, should be of interest to international scholars of science education. Recognition of a crisis in Western (and particularly the USA) science education in the early 1980s triggered a spate of reports decrying the ''abysmal state of scientific and mathematical knowledge of high school graduates'' (Matthews, 1994, p. 29). As a result, a goal of science literacy for all was adopted for the USA and most other countries. In the USA, the National Science Foundation and other government and private agencies have poured many millions of dollars into research on improving science education. Almost a quarter of a century later, we need to take a close look at what that research has revealed about factors influencing science achievement.

Student Achievement in Science

The National Assessment of Educational Progress (NAEP), also known as the Nation's Report Card, is administered periodically to a representative sample of students to measure student achievement in science, mathematics, reading, and several other subjects. The NAEP framework emphasizes science concepts and application of knowledge and skills rather than recall of facts. The 2005 report revealed that younger students are making the most progress in science as compared with middle and high school students, and that most states showed no improvement at grades 4 and 8. At grade 12, the mean score has declined since 1996 (National Center for Education Statistics, 2005).

The USA participates in two international assessments, the Trends in International Mathematics and Science Study (TIMSS) and the Program for International Student Assessment (PISA), to benchmark its performance versus other countries. The TIMSS is administered to fourth and eighth graders in math and science, whereas PISA is given to 15-year-olds to measure reading, mathematical, and scientific literacy. TIMSS scores are compared with NAEP scores at early and middle grades, whereas PISA is the U.S. source for comparative information at upper grades. TIMSS 2003 data reveal that U.S. fourth- and eighth-grade students performed above the average of all countries tested, but when compared with other industrialized countries the U.S. students lagged behind students from Belgium, Hungary, Japan, Korea, and The Netherlands. PISA 2003 data indicate that U.S. students do not perform as well as other students tested on literacy measures. In making comparisons it is nevertheless important to remember that comparing age- and grade-based samples, differently sampled groups, and assessments with differing goals and content preclude direct comparisons. Both TIMSS and NAEP data show no change in U.S. fourth-graders' average scores in science from the mid-1990s to 2003. Eighth graders' TIMSS scores show an increase over the same time period (with most of the improvement occurring between 1999 and 2003), whereas the NAEP scores show no change (National Center for Education Statistics, 2004).

Examining NAEP scores in light of the results of TIMSS and PISA enables us to compare the performance of U.S. students as a whole to the performance of students from other nations, adding an important international perspective to our understanding of how our students are doing. Scores reveal little or no change in U.S. students' science achievement over the last decade as well as a distressing position relative to other industrialized nations. As the USA moves into the 21st century with its increased globalization and high-tech economy, the importance of a work force with a high-quality science education and skills becomes more evident. To achieve the goal of science literacy for all, there is little question that professional development for teachers of science is necessary to improve the effectiveness of instruction.

Teacher Effectiveness

Multiple studies have shown that teachers have a profound effect on student learning (Nye, Konstantopoulos, & Hedges, 2004; Rivkin, Hanushek, & Kain, 2005; Wright, Horn, & Sanders, 1997). After multivariate, longitudinal analyses of schools, class sizes, teachers, and other effects, Wright et al. (1997) concluded "Differences in teacher effectiveness were found to be the dominant factor affecting student academic gain" (p. 66). Identifying and isolating the specific characteristics that influence teacher effectiveness and thereby student achievement is problematic (Nye et al., 2004; Wayne & Youngs, 2003; Wright et al., 1997). If strategies employed by effective teachers can be isolated, however, and then taught to less effective teachers, student learning should improve.

Teaching Strategies

Effective teaching is a complex activity requiring, among other types of knowledge, pedagogical content knowledge (PCK), and curricular knowledge. Shulman defined PCK as the knowledge of subject matter *for teaching* and includes within it:

> . . .the most useful forms of representation of. . .ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations—in a word, the ways of representing and formulating the subject that make it comprehensible to others. Since there are no single most powerful forms of representation, the teacher must have at hand a veritable armamentarium of alternative forms of representation, some of which derive from research whereas others originate in the wisdom of practice (p. 9).

Shulman went on to discuss curricular knowledge:

> The curriculum and its associated materials are the. . .pharmacopeia from which the teacher draws those tools of teaching that present or exemplify particular content and remediate or evaluate the adequacy of student accomplishments. . .. [W]e ought to expect that the mature teacher possesses such understandings about the curricular alternatives available for instruction. . .. How many individuals whom we prepare for teaching biology, for example, understand well the materials for that instruction, the alternative texts, software, programs, visual materials, single-concept films, laboratory demonstrations, or "invitations to enquiry?" Would we trust a physician who did not really understand the alternative ways of dealing with categories of infectious disease, but who knew only one way? (p. 10).

Marzano (1998, pp. 5–7) cited previous meta-analyses on instructional research conducted by Hattie in 1992 and Fraser et al. in 1987 to illustrate the difficulties of categorizing research into useful groups. The goal of Marzano's study of instruction in multiple disciplines was to "make sense of what research has to say about instruction" (p. 1) from the multitude of various types of studies of factors influencing student achievement. The result was a theory-based meta-analysis of studies concerning the effect of classroom instructional behaviors on student achievement by teachers at all grade levels and in all subject areas. Marzano concluded that knowledge goals, cognitive goals, and metacognitive goals are achieved through the use of different instructional strategies. A teacher must identify the instructional goal for a unit or lesson and then choose appropriate strategies to achieve the goal.

A large number of studies have been conducted in science classrooms to determine the effect of specific factors on student achievement in science. In the early 1980s, a multi-institutional effort sponsored by the National Science Foundation addressed several questions concerning science education using meta-analytic techniques to synthesize the results of multiple studies (Anderson, Kahl, Glass, & Smith, 1983). The researchers used meta-analyses to look at the effects of curricular programs, instructional systems, teaching techniques (strategies), and teacher education programs (both pre-service and in-service). They also considered the relationships between teacher and student characteristics and student outcomes. The results of several of these meta-analyses were reported in the *Journal of Research in Science Teaching*. One of these studies (Wise & Okey, 1983) specifically examined the effects of various teaching strategies (methods, techniques) on achievement in science in elementary through post-secondary settings. The investigators grouped the studies into 12 categories of teaching techniques: "audio-visual, focusing, grading, inquiry, manipulative, modified, presentation approach, questioning, teacher direction, testing, wait-time, and miscellaneous" (p. 420). They found that the group of 12 experimental strategies resulted in a mean effect size (ES) of 0.34, indicating an improvement of one-third standard deviation over traditional teaching strategies. The range of effect sizes reported was from 0.90 for wait-time strategies to −0.15 for grading strategies.

Wise (1996) revisited the question and reported the results of a secondary meta-analysis of teaching strategies research conducted to determine if there is a relationship between alternative or experimental teaching strategies and student achievement at the middle school and high school levels. The majority of studies included in this second meta-analysis had been published in science education journals, doctoral dissertations, and ERIC documents between 1965 and 1985. They compared alternative teaching strategies to traditional strategies, which Wise described as teachers dispensing ‘‘knowledge to passive student audiences, with textbooks alone constituting the science curricula; students are rarely involved in direct experiences with scientific phenomena’’ (p. 1). In an attempt to make the categories more usable, he placed the studies into eight categories of strategies: Questioning; Focusing; Manipulation; Enhanced Materials; Testing; Inquiry; Enhanced Context; and Instructional Media. A mean effect size of 0.32 was found for the alternative strategies examined in the study, with effect sizes ranging from 0.58 for Questioning Strategies to 0.18 for Instructional Media Strategies. Wise concluded that, at the secondary levels, the alternative teaching strategies were more effective than the traditional strategies at improving student achievement in science. He also asserted that the difference that defines and distinguishes the alternative strategies from the traditional science teaching strategies is the use of inquiry-oriented instruction, with inquiry strategies permeating all eight categories.

Meta-Analysis

Meta-analysis is a quantitative technique that integrates the results of a number of different primary studies to analyze and synthesize them into a coherent product. It provides a rigorous method for building on the findings of previous studies and aggregating the results to advance scientific knowledge and to guide policy development during educational reform (McNamara, Morales, Kim, & McNamara, 1998). In a meta-analysis, the results of each study are dealt with as autonomous observations that are then combined to compute an overall effect size.

Eugene V. Glass introduced meta-analytic procedures to the social sciences in his presidential address to the American Educational Research Association in 1976 (Hunt, 1997). The phases of the process he delineated include formulation of the problem, data collection (searching for relevant studies), data evaluation (eliminating studies that do not meet the criteria), data synthesis (using statistical methods such as combining effect sizes to aggregate studies), and presenting findings to the research community. In the years since 1976, the meta-analytic technique has grown steadily and has been adopted by disciplines outside the social sciences; its practitioners ‘‘nearly all share a desire to discover patterns in the seemingly hopeless jumble of dissimilar findings’’ (p. 13).

Effect size is reported in quantitative studies to enable the reader to understand the importance of the results of the study. An effect size indicates the ‘‘degree to which sample results diverge from the null hypothesis. . .. As the sample results increasingly diverge from whatever is specified by the null hypothesis, the effect size will increasingly diverge from zero’’ (Thompson, 2002). The APA’s *Publication Manual* (American Psychological Association, 2001) has recommended the inclusion of effect size in the Results section of studies. Thompson also encouraged inclusion of confidence intervals for effect sizes, defining a confidence interval as ‘‘one interval from among an infinite or at least large sample of CIs for a given parameter in which $1 - \alpha\%$ of the intervals would capture the population parameter’’ (p. 26). He noted, ‘‘Confidence intervals for effect sizes are especially valuable because they facilitate *meta-analytic thinking* and the interpretation of intervals via comparison with the effect intervals from related prior studies’’ (p. 25).

Many more science education studies have been conducted in the decade since publication of the last meta-analysis concerning science teaching strategies. As a result of the great push for improving science instruction, the time was right for another aggregation of results of more recent studies. The current meta-analysis addressed the question: *What teaching methodologies have been shown to improve student achievement in science in the USA?*

Methods

Acquisition of Studies

A broad search resulted in the collection of over 390 studies obtained from journal articles, conference papers, books, dissertations, government reports, and unpublished papers. The majority of the reports were identified through searches of electronic databases, including ISI Web of Science, ERIC, ERIC EBSCO, ERIC FirstSearch, ERIC CSA (Cambridge Scientific Abstracts), Academic Search Premier, PsycINFO, and ProQuest dissertations and theses. Government websites, such as the Berkeley National Laboratory and the Department of Education, and other science education sites, including the National Academies, National Science Resources, American Association for the Advancement of Science (AAAS), Education Development Center, and National Science Teachers Association, provided reports, links to sources, and names of science programs. In addition, general web searches using standard search engines, such as Google and Google Scholar, were conducted. Initial search terms included ''science education'' and ''student achievement.'' Subsequent searches were expanded by using various combinations of alternatives for achievement, such as ''performance,'' ''success,'' and ''outcomes,'' and by combining them with individual science disciplines, such as ''biology,'' ''chemistry,'' ''physics,'' ''physical science,'' ''earth science,'' ''ecology,'' and ''environmental science.'' Other terms, such as ''science teaching'' and ''professional development,'' were also used in search strings.

A request was sent by e-mail to the National Association for Research in Science Teaching (NARST) listserve soliciting suggestions of published articles or reports that might contain useful research or for copies of unpublished research. Experts in the field of science education research were asked to review the bibliography and make suggestions. In addition, direct requests for information were sent individually to project directors of ''exemplary and promising science programs'' identified by the Department of Education Expert Panel on Mathematics and Science Education. The programs contacted included the Biological Sciences Curriculum Studies (BSCS), Project ARIES, Project InSIGHT, Event-Based Science, Foundational Approaches in Science Teaching (FAST), Full Option Science System (FOSS), Great Explorations in Math and Science (GEMS), Modeling Instruction in High School Physics, Phenomena and Representations for the Instruction of Science in Middle Schools (PRISMS), and Science 2000+. Finally, reference lists from books, dissertations, studies, and other meta-analyses were examined for potential inclusion.

Coding of Studies

Characteristics of the studies (moderator variables) were coded to investigate the possible influence of some of these variables on effect size. A coding document (Appendix A) was developed to include the following attributes:

- Study number and citation.
- Publication type: (1) refereed journal article, (2) dissertation, or (3) unpublished report.
- Study type: (1) experimental—complete randomization, (2) quasi-experimental—randomization used, (3) quasi-experimental—no randomization, and (4) correlational.

- Dependent variable: (1) type of test, (2) test name, (3) number of items, and (4) content area's.
- Independent variable: (1) treatment name and description, and (2) control and/or alternate treatment.
- Length of treatment/study.
- Setting and characteristics, including:
  - —schools: (1) number of schools, (2) if selected at random, (3) if unit of analysis, (4) public/private, (5) urban/rural/suburban, (6) size, and (7) percent free lunch.
  - —students: (1) number of students, (2) if selected at random, (3) if assigned at random, (4) if unit of analysis, (5) gender, (6) grade, (7) ethnicity, and (9) socioeconomic status; and
  - —teacher(s): (1) number of teachers, (2) if volunteer or selected, (3) age/experience, (4) gender, and (5) certification.
- Study results: (1) effect size(s), (2) *p*(s), (3) *t*(s), (4) *F*(s), (5) eta square(s), and (6) omega square(s)

Due to time constraints, studies coded as correlational were not included in the meta-analysis. We hope to include them in a future study.

Intercoder Objectivity

To establish intercoder reliability, three journal articles were selected at random and coded independently by the senior analyst and two members of the research team. For two of the articles, the degree of objectivity was 90%. Due to the non-inclusion of correlational studies in this meta-analysis, only two items were coded for the third article as the coders stopped coding after identifying the study as a correlational study. The remaining articles for this project were divided between the two members of the research team who coded their respective articles and then submitted them to the senior analyst. The senior analyst then read and coded all of the articles and any differences in coding values were resolved at the discretion of the senior analyst. All of the dissertation material was evaluated and coded by the senior analyst and two recent PhD recipients who possess strong statistical backgrounds.

Criteria for Selection of Studies

In the initial search for studies we employed liberal criteria, and many reports, studies, and articles were collected that proved to be unusable for this meta-analysis. For final inclusion in the meta-analysis, rigorous criteria were employed, in part due to the time constraints imposed by contract stipulations. Studies had to have:

- been published between January 1, 1980, and December 31, 2004
- been concerned with K–12 science education in the United States
- used student achievement (or success, performance, etc.) as the dependent variable
- used science education teaching strategies as independent variables
- been experimental or quasi-experimental
- reported effect size (ES) or the statistics necessary to calculate ES (means and standard deviations, *p*-values, ANOVA tables, etc.
- not been totally correlational
- been for general education students/classes (not deal exclusively with a special population); and
- not been included more than once (e.g., the same study reported in a conference paper and a journal article

Table 1
Reasons and frequencies for non-inclusion of studies

Reasons	Number of Studies
Did not fit time frame[a]	
Did not concern K-12 science education in the USA	58
Did not use student achievement as dependent variable (DV) and/or science education methodologies as independent variable (IV)	97
Was not experimental or quasi-experimental	41
Did not have effect size or necessary data to calculate effect size	62
Was not for general education students (could not be concerned with special populations)	25
Duplicated another study	6
Other (a meta-analysis report, bibliography, descriptive case study, position paper, literature review, etc.)	48
Total	337

[a]Large number of studies not collected because it was immediately obvious they did not fall within the time frame.

The number of studies that were not included and the reasons for non-inclusion are presented in Table 1. Many studies were not included because their publication dates fell outside the parameters established in the selection criteria. A large number of other studies were eliminated because student achievement was not the dependent variable or science teaching strategies were not used as the independent variables.

Computation of Effect Size

For each achievement measure reported in the included studies, an ES was calculated. Basic ES calculations are based on Glass's Δ using the formula $ES = (\bar{X}_T - \bar{X}_C)/S_C$, where Δ is replaced with ES and $\bar{X}_T$ is the mean for the experimental or treatment group, $\bar{X}_C$ is the mean for the control group, and S_C is the standard deviation of the control group. To meet the meta-analytical assumption of independence of effect sizes, only one effect size indicator per study should be represented in an analysis. If a study used several different outcome measures (e.g., some of the dissertations), then there would be more than one effect size for that study. In the present meta-analysis, multiple effect sizes per study were averaged in two ways. If a study used the same subjects with more than one dependent variable (student achievement tests):

- a weighted mean effect size was calculated whenever the sample sizes were very divergent; and
- an unweighted mean effect size was calculated whenever the sample sizes were equal or approximately equal.

These averaging procedures resulted in one effect size for each study.

Results

Section 1: Description of Studies

Characteristics and frequencies of the selected studies. A disaggregation of the characteristics of the studies selected for the meta-analysis is presented in Table 2. These characteristics have the potential to influence the effect sizes obtained from the studies. A majority

Table 2
Frequencies of variable characteristics for included studies

Independent Variable	Number of Cases	Percent (%)
Publication Year		
1980–1984	6	9.7
1985–1989	7	11.3
1990–1994	4	6.5
1995–1999	15	24.2
2000–2004	30	48.4
Publication Type		
Refereed journal article	40	64.5
Dissertation	18	29.0
Unpublished report	4	6.5
Type of Study		
Experimental (complete randomization)	3	4.8
Quasi-experimental (randomization used)	33	53.2
Quasi-experimental (no randomization)	26	41.9
Test Content Area		
Biology	17	27.4
Chemistry	12	19.4
Physics	5	8.1
Earth science	7	11.3
Science (general)	21	33.9
Length of Treatment (months)		
0.0–1.0	20	32.3
1.1–7.5	22	35.5
8.0–12.0	10	16.1
Missing information	10	16.1
Totals (for each variable)	62	100.0

of the studies included were published between 1995 and 2004 in refereed journals, possibly because of the increase in both the number of science education researchers and the number of journals publishing science education research. Far more studies were quasi-experimental rather than experimental, most probably due to the difficulty of conducting randomized experimental studies with students and teachers in a school setting. Instructional methods used for the control groups were generally not well-defined, but "traditional" typically meant more teacher-dominated instruction with passive student participation. Tests used to determine achievement focused primarily on general science, biology, or chemistry. Approximately one-third of the studies were 1 month or less in duration, and another third lasted between 1 and 7.5 months.

Dependent variable. The dependent variable (DV) of this meta-analysis, student achievement, can be referred to by many names (student performance, success, outcomes) and may be evaluated with a multitude of assessment tools. In any attempt to synthesize results from different studies, consideration must be given to the validity and reliability of the different forms of the DV that will be used. In this meta-analysis of studies of science education, five different forms of achievement assessment were scrutinized. The types of assessments used in the studies and the frequencies with which they occurred are presented in Table 3.

The form of achievement assessment most often encountered (a total of 47 times) was local teacher- or researcher-developed tests. It is assumed that the instruments constructed by these individuals would possess acceptable levels of logical relevancy because tests were formulated to

Table 3
Test type

Test Type	Number of Cases	Percent (%)
National standardized—Multiple science content	3	4.8
National standardized—Single science content	6	9.7
Local standardized—Multiple science content	2	3.2
Local standardized—Single science content	4	6.5
Other type test	47	75.8
Totals	62	100.0

match the instructional units that were being investigated. The degree of reliability of the teacher-made instruments is unknown as most of the authors did not provide information regarding this characteristic. Therefore, it was necessary to assume that this form of assessment (local type test) was conducted with adequate levels of reliability.

For the 15 studies that used standardized national or local assessments of either multiple or single content areas, an indication of validity was usually not reported but the authors of those studies referred to the original developers of the instruments when discussing validity. Most authors gave indications of strong content and construct validity for these forms of assessment. In terms of reliability for these tests, the range of values reported was 0.70 to 0.94. Because these assessment tools have been widely used and evaluated, the validities and reliabilities are well within recommended tests and measurement guidelines.

Independent variables. To explore some of the heterogeneity of effect sizes among the studies, the various treatment conditions (teaching strategies) were placed into treatment categories. Rather than operationally define treatment description categories based on the studies used in this meta-analysis, an established set of teaching strategy categories (Wise, 1996) was modified and employed. *Direct instruction* was added to the original set due to the recent emphasis on it as an effective means of instruction, although it is not a well-defined term (Li & Klahr, 2006). *Collaborative learning strategies* was added due to the presence of several studies wherein group learning was the main emphasis. *Instructional media* was renamed *instructional technology* to broaden the category to include the use of computers in classroom teaching. *Assessment strategies* replaced *testing strategies* to update terminology and to include strategies such as feedback and informal evaluation. The modified list of treatment categories with the description used for each includes the following strategies:

- *Questioning strategies*. Teachers vary timing, positioning, or cognitive levels of questions (e.g., increasing wait time, adding pauses at key student-response points, including more high-cognitive-level questions, stopping visual media at key points and asking questions, posing comprehension questions to students at the start of a lesson or assignment)
- *Focusing strategies*. Teachers alert students to the intent of the lesson or capture their attention (e.g., providing objectives or reinforcing objectives at the middle or closing of lesson, using advance organizers)
- *Manipulation strategies*. Teachers provide students with opportunities to work or practice with physical objects (e.g., developing skills using manipulatives or apparatus, drawing or constructing something)
- *Enhanced materials strategies*. Teachers modify instructional materials (e.g., rewriting or annotating text materials, tape recording directions, simplifying laboratory apparatus)

- *Assessment strategies*. Teachers change the frequency, purpose, or cognitive levels of testing/evaluation (e.g., providing immediate or explanatory feedback, using diagnostic testing, formative testing, retesting, testing for mastery)
- *Inquiry strategies*. Teachers use student-centered instruction that is less step-by-step and teacher-directed than traditional instruction; students answer scientific research questions by analyzing data (e.g., using guided or facilitated inquiry activities, laboratory inquiries)
- *Enhanced context strategies*. Teachers relate learning to students' previous experiences or knowledge or engage students' interest through relating learning to the students'/ school's environment or setting (e.g., using problem-based learning, taking field trips, using the schoolyard for lessons, encouraging reflection)
- *Instructional technology strategies*. Teachers use technology to enhance instruction (e.g., using computers, etc., for simulations; modeling abstract concepts and collecting data; showing videos to emphasize a concept; using pictures, photographs, or diagrams
- *Direct instruction*. Teachers deliver information verbally or explicitly guide students through a sequence of tasks (e.g., learning by listening, designing experiments, using a microscope, making measurements)
- *Collaborative learning strategies.* Teachers arrange students in flexible groups to work on various tasks (e.g., conducting lab exercises, inquiry projects, discussions)

A rating sheet (Appendix B), consisting of the treatment categories and a five-point Likert-type rating scale, was developed to sort each study into an appropriate category. Three science educators, each possessing 20–30 years of public school science instruction experience, read a description of the treatment condition of each study. They then ranked each of the treatment strategies from 1 to 5 for the study. The strategy category receiving the highest ranking response for a given study was chosen as the predominant treatment category for that study. It should be noted that, although ten strategies are listed, only eight were included in the analysis because no studies meeting the criteria were found for *focusing strategies* or *direct instruction*. In fact, strategies used in teaching control groups for the studies often met the definition of direct instruction in being teacher-centered rather than student-centered. The studies included in the meta-analysis are listed by category in Appendix C. The analysis of the effect sizes for the different treatment categories is presented in Section 4 concerning the outcomes of the meta-analysis.

Section 2: Identification and Analysis of Validity Issues

In this meta-analysis, internal validity is concerned with study quality issues that might influence the effect size obtained for science instruction strategies. The three variables related to internal validity examined in this study are shown in Table 4: (a) Publication Type; (b) Type of Study; and (c) Test Type. Because the numbers of cases associated with the levels of the Test Type are too diverse to warrant analyses across all levels, the studies using standardized assessments were aggregated for comparison to other type tests. The mean effect sizes for these potential internal validity variables ranged from 0.28 to 1.03.

External validity deals with generalization issues, such as whether the effect of science instruction strategies could be generalized to other populations or situations. In this study, four variables that might influence external validity were examined: (a) Publication Year; (b) Test Content; (c) Grade Level; and (d) Treatment Categories. The number of studies within these categories varied from as few as 2 to as many as 45. The effect-size means associated with the levels of the external validity variables (Table 5) ranged from 0.33 to 1.43.

The potential moderator variables were placed in a weighted least-squares multiple-regression analysis to determine the influence of all the coding components on the effect-size data.

Table 4
Internal validity influences on the effect size of science instruction strategies

Internal Validity Influence	Mean	*SD*	*N*
Publication Type			
Refereed journal article	0.91	0.68	40
Dissertation	0.28	0.61	18
Unpublished report	1.03	1.19	3
Type of Study			
Experimental (complete randomization)	0.83	0.38	3
Quasi-experimental (randomization used)	0.53	0.68	33
Quasi-experimental (no randomization)	1.00	0.78	25
Test Type			
Standardized	0.73	0.48	15
Other type test (teacher- or researcher-constructed)	0.75	0.79	46
Total studies (for each influence)			61

The total *N* for each influence is 1 less than previously presented due to exclusion of an outlier study.

From the results of the regression analysis, the *F* ratio associated with these data indicates that there was a significant relationship between the moderator variables and the dependent variable of effect size ($R^2 = 0.80$). The regression beta weights for this analysis are presented in Table 6. The beta weights associated with the internal validity variables of Publication Type, Type of Study, and Test Type were judged to be significant. This means that the effect sizes obtained in this meta-analysis are influenced by the fact that the studies have different levels of these potentially

Table 5
External validity influences on the effect size of science instruction strategies

External Validity Influence	Mean ES	Mean *SD*	*N*
Publication year			
1980–1984	0.61	0.47	6
1985–1989	0.33	0.16	7
1990–1994	1.15	0.54	4
1995–1999	0.70	0.88	14
2000–2004	0.81	0.79	30
Test content area			
Biology	0.48	0.76	17
Chemistry	1.01	0.88	11
Physics	0.54	0.26	5
Earth science	0.37	0.50	7
Science (general)	0.96	0.72	21
Grade level			
Elementary and middle school (K–8)	0.66	0.68	16
High school (9–12)	0.76	0.76	45
Treatment strategy categories			
Questioning Strategies	0.74	0.53	3
Manipulation Strategies	0.86	0.83	8
Enhanced Materials Strategies	0.41	0.48	12
Assessment Strategies	0.49	1.00	2
Inquiry Strategies	0.63	0.71	12
Enhanced Context Strategies	1.43	1.05	6
Instructional Media Strategies	0.79	0.77	15
Collaborative Learning Strategies	0.59	0.59	3
Total studies (for each influence)			61

Table 6
Regression analysis of dependent variable: effect size (N = 61)

Moderator Variable	Beta	t	Significance
Publication year	0.12	0.91	0.37
Grade level	−0.23	−1.21	0.23
Test content	0.32	1.06	0.30
Treatment categories	0.01	0.06	0.95
Type of study	−0.42	−2.11	0.04[a]
Publication type	−1.03	−6.07	0.00[b]
Test type	−0.59	−4.82	0.00[b]
Length of treatment (months)	−0.34	−1.02	0.31

$R^2 = 0.801$. The effect size outlier had been excluded in this analysis.
[a]$p < 0.05$.
[b]$p < 0.01$.

contaminating attributes. In terms of external validity, the regression results show that there is no significant relationship between Publication Year, Grade Level, Test Content, or Treatment Categories and effect size. In other words, the effect sizes are not influenced by differing levels of these external variables.

Section 3: Meta-Analysis for All Studies

To explore the 62 effect sizes obtained in this study, box plots obtained using SPSS were constructed. Box plots are used to graphically portray observations that would be judged to deviate substantially from typical values. The first box plot is shown in Figure 1.

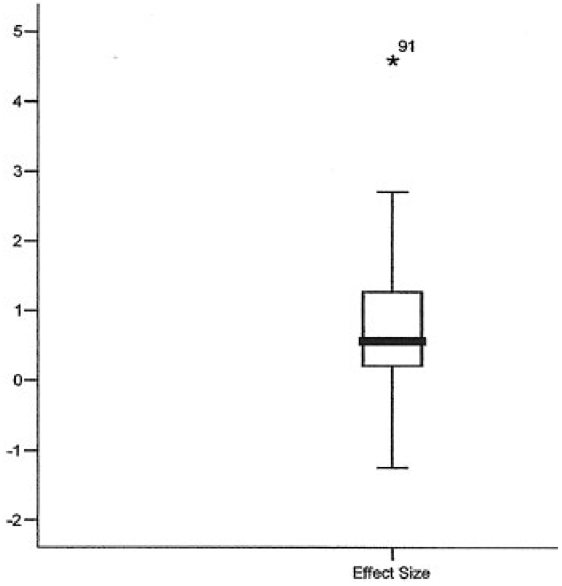

Figure 1. Box plot of ES for total data ($N = 62$).

Based on this analysis, Study #91 was identified as an extreme outlier, meaning that its value was substantially different from the median value. The data for mean achievement that were presented in Study #91 were judged to be suspect. Therefore, this particular report was excluded from the analysis and a new box plot was constructed. Figure 2 illustrates that, after removal of the extreme outlier study, two studies were identified as mild outliers. These studies were not excluded from the meta-analysis because they were within the ± 3 effect-size range.

Based on this analysis, one study was identified as an extreme outlier and excluded from the analysis. A summary of the analysis of effect sizes for the total data set of 61 included studies is presented in Table 7.

The results across approximately 160,000 student achievement scores indicate an effect of science instruction methodology on student achievement. The effect size of 0.67 translates to a treatment performance level that would be located at the 75th percentile of the control group. This effect size was judged to be statistically different from 0.00 as indicated by the probability value associated with the *t* statistic.

Section 4: Meta-analysis for Studies Classified by Treatment Categories

Comprehensive meta-analysis was used to analyze the effect sizes for the treatment categories. A summary of the results of these analyses is presented in Table 8. For comparison purposes, the results for the total group are also included in the table.

Several interesting and informative results are presented in Table 8. First, the effect sizes for all of the treatment categories exceed the lower effect-size benchmark value of 0.20. Second, two treatment categories, Enhanced Context Strategies and Collaborative Learning Strategies, exceed the upper effect size benchmark value of 0.80. Third, all of these effect sizes would be judged to be significantly different from zero. The treatment category of Enhanced Material Strategies exhibited an effect size that would be classified as small. The 95% confidence intervals for the effect sizes of the treatment categories clearly show the treatment category of Enhanced

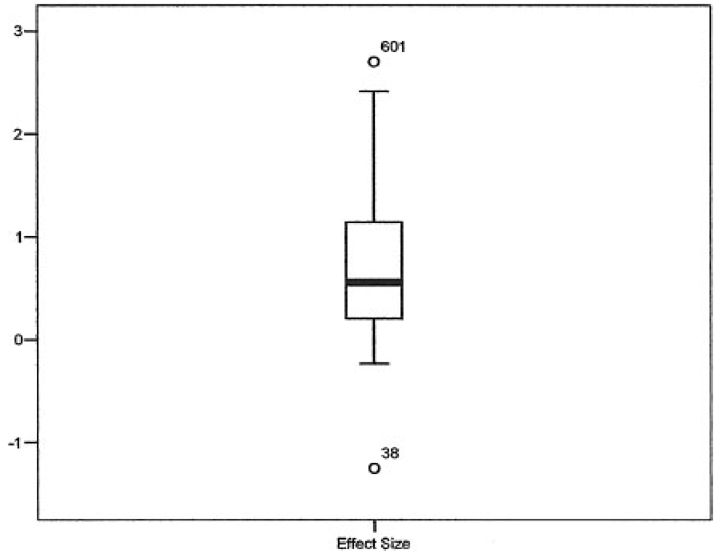

Figure 2. Box plot for data with extreme outlier removed ($N = 61$).

Table 7
Meta-analysis result for all studies

Total	*N* Total	Effect Size	Lower Limit	Upper Limit	*t*	*p*
61 studies	159,695	0.67	0.66	0.68	128.96	0.00

Context Strategies to be more effective than the other strategies. Many of the confidence intervals for the other treatment strategies contain overlapping ranges and would therefore be judged not to differ from each other.

When making decisions based on the results of meta-analysis, the researcher is always concerned with what is known as the ‘‘file drawer problem.’’ This problem refers to the idea that only studies that result in an effect are included in a meta-analysis. Studies that did not uncover an effect are usually filed away by the researchers and are not available for inclusion. A calculation that attempts to address this concern is the *Failsafe N* (N_{fs}). The numerical value obtained from this statistic is an estimate of the number of non-significant ‘‘file drawer’’ studies that would need to be obtained and included in a meta-analysis before a statement of no effect would be given to the completed meta-analysis. *Failsafe N*s for the total data set and the studies in the treatment categories are shown in Table 9.

Based on the estimates in Table 9, the only questionable treatment category effect size in terms of a file drawer problem is for Assessment Strategies. If 19 nonsignificant results were found in file drawers, a decision of no effect for this category of strategies would be indicated. The resultant N_{fs} for the overall data would lead to the conclusion that a decision to deny an effect is not probable given that approximately 756 nonsignificant studies would be needed to reverse the decision about the overall effect.

The funnel plot (Figure 3) graphically displays the effect sizes for the 61 studies along with their treatment group sample sizes. Generally, a study with small sample size can produce some surprisingly good or poor results by chance. For studies with larger sample sizes, the outcome is less likely to be a result of chance (Torgerson, 2003). The smaller and less precise studies will be scattered along the x-axis, whereas the larger and more precise studies will be clustered together. The inverted funnel shape as presented for this study suggests that there is little or no evidence of publication bias.

Table 8
CMA results for total data and treatment categories

			95% Confidence Interval			
Total	*N* Total	Effect Size	Lower Limit	Upper Limit	*t*-value	*Q*-value
61 Studies	159,695	0.67	0.66	0.68	128.96	158.09
Treatment categories						
Questioning Strategies	279	0.74	0.49	0.99	5.90	11.86
Manipulation Strategies	1240	0.57	0.46	0.69	9.65	104.88
Enhanced Material Strategies	2450	0.29	0.21	0.37	7.05	61.55
Assessment Strategies	166	0.51	0.19	0.82	3.20	0.19
Inquiry Strategies	145,722	0.65	0.64	0.67	121.64	364.49
Enhanced Content Strategies	7235	1.48	1.42	1.54	47.06	147.59
Instructional Technology Strategies	1962	0.48	0.39	0.58	10.27	80.20
Collaborative Learning Strategies	641	0.95	0.78	1.14	10.41	26.87

Table 9
Failsafe N for total data and treatment description categories

Data	ES	*N*	N_{fs}
Overall	0.67	61	756
Questioning Strategies	0.74	3	42
Manipulation Strategies	0.57	8	84
Enhanced Materials Strategies	0.29	12	58
Assessment Strategies	0.51	2	19
Inquiry Strategies	0.65	12	145
Enhanced Context Strategies	1.48	6	172
Instructional Technology Strategies	0.48	15	130
Collaborative Learning Strategies	0.96	3	55

N_{fs} = failsafe N or number of studies needed to decrease effect size below .05.

Summary and Discussion

The purpose of this study was to examine the extant body of recent studies in science teaching to provide research-based evidence of effective teaching strategies or "tools of teaching." We collected a large number of studies dealing with teaching strategies and student achievement in science and, of these, 61 met our criteria for inclusion in the meta-analysis. Studies were classified into eight types of strategies and analyzed to determine the mean effect size for each strategy; they are shown ranked by effect size in Table 10.

The largest effect size, 1.48, was for Enhanced Context Strategies such as relating topics to previous experiences or learning and engaging students' interest. Teachers can make learning

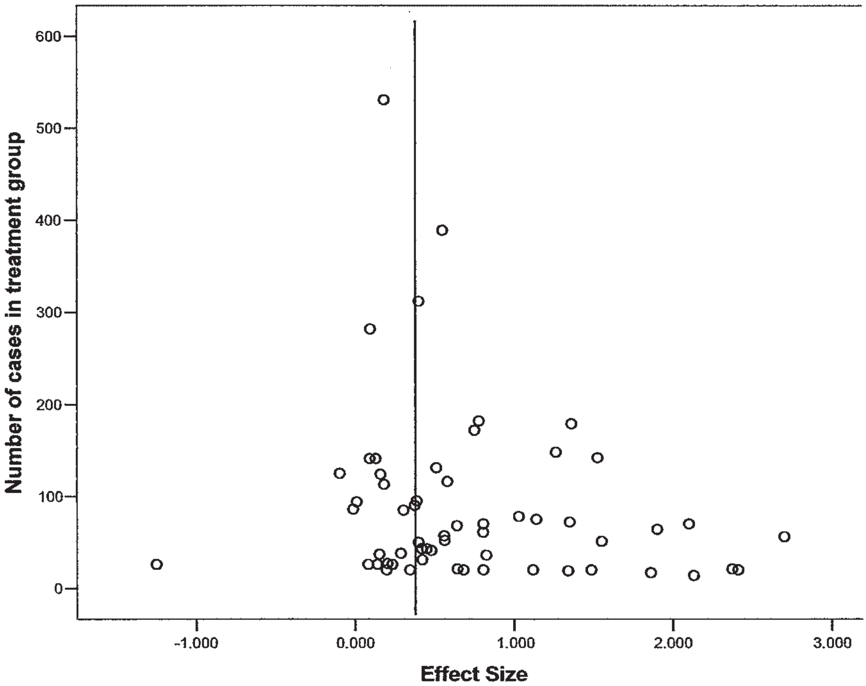

Figure 3. Funnel plot of 61 studies.

Table 10
Ranking of teaching strategies

Strategies	Effect Size	Rank
Enhanced Context Strategies	1.48	1
Collaborative Learning Strategies	0.96	2
Questioning Strategies	0.74	3
Inquiry Strategies	0.65	4
Manipulation Strategies	0.57	5
Assessment Strategies	0.51	6
Instructional Technology Strategies	0.48	7
Enhanced Material Strategies	0.29	8

relevant to students by presenting material in the context of real-world examples and problems. The real world can be brought to students through technology and students may be taken out of the classroom into the real world through field experiences. Collaborative Learning Strategies such as flexible heterogeneous groupings and group inquiry projects also displayed a strong effect. It must be remembered that these categories of strategies are not discrete and often overlap. The importance of this study is that the alternative teaching strategies exhibited a positive influence on student achievement when compared with the traditional teaching methods used in instruction of the control groups. Wise (1996) concluded that innovative science instruction is a mixture of teaching strategies and no one strategy is as powerful as utilizing a combined strategies approach. If students are placed in an environment in which they can actively connect the instruction to their interests and present understandings and have an opportunity to experience collaborative scientific inquiry under the guidance of an effective teacher, achievement will be accelerated.

As we educators strive to achieve science literacy for all and to improve the standing of U.S. students in comparison with students from around the world, we must employ those strategies that have been shown to be effective in improving student achievement. In addition to the types of teacher knowledge discussed earlier in this study (PCK, curricular knowledge), Shulman (1986) described forms in which teacher knowledge is represented. He suggested that teaching strategies are propositional knowledge, and that when we examine the research on teaching strategies and look at its implications for classroom teaching, we are examining propositions. Shulman further classifies propositional knowledge as principles, maxims, and norms. Principles derive from empirical research, maxims ‘‘represent the accumulated wisdom of practice’’ (p. 11), and norms are principles that guide the work of teachers because they are morally or ethically correct. I would propose that the meta-analysis of existing research strengthens existing principles by adding a preponderance of evidence and may transform widely accepted maxims into principles. The results of this meta-analysis have provided us with evidence that the eight types of science teaching strategies listed may be considered *principles* for effective science teaching. Teachers' pedagogical content knowledge must enable them to purposefully select from among the principles to accomplish the particular goal for a lesson (Bransford, Brown, & Cocking, 2000, p. 22; Shulman, 1986).

Research Limitations and Further Study

The majority of studies included in this meta-analysis were classified as quasi-experimental. Although some would advocate the use of only the results obtained from randomized, controlled trials (RCTs) as indicators of sound, rigorous research, RCT studies may have ethical and legal issues when applied to school and classroom settings and researchers often do not have the kind of control available in laboratory settings (Shavelson & Towne, 2002). A major point of interest from this meta-analysis is that the effect sizes associated with the predominant methodology in science instruction studies indicate separation of science teaching strategies.

Although we were advised to include direct instruction as one of the categories of strategies, we did not find any studies specifically in science education that met the criteria established for the meta-analysis. We examined studies by Chen and Klahr (1999), Klahr and Nigam (2004), and Li and Klahr (2006) that included direct instruction as a strategy in teaching science. However, in those studies the researchers examined the teaching of a process skill, looking at the ability of young students to design unconfounded experiments and make inferences from the results (1) after instruction in experimental design plus probes versus (2) after only probe questions versus (3) after no instruction in experimental design. The direct instruction in those studies was used to teach a process skill without content. There was no science achievement test given at the conclusion of the studies, and therefore they did not meet the criteria for inclusion in this meta-analysis. Direct instruction is a term sometimes used to describe a highly scripted instructional model used primarily in early reading and mathematics instruction. It may, however, also refer to "a wide range of teacher-controlled talking, showing, questioning, and demonstrating. . . . Such a lack of precise operational definitions has allowed earnest, passionate, but empirically ungrounded debates to flourish in the area of science instruction policy" (Li & Klahr, 2006, p. 19). The lack of precise definitions of critical teaching strategies highlights challenges for future researchers posed by Shavelson and Towne (2002)—both to operationally define critical science teaching strategies and to perform empirical studies that meet the criteria for scientifically based research.

In the 2 years since the articles were collected for this meta-analysis, a number of additional studies have been published in the *Journal of Research in Science Teaching* (*JRST*) that would possibly have met the criteria for inclusion. One study (Tal, Krajcik, & Blumenfeld, 2006) described the positive effect of inquiry-oriented projects used in urban classrooms. They emphasized the importance of teacher PCK in successful implementation of inquiry learning curriculum materials and promoting student achievement. A second study (Lee, Deaktor, Hart, Cuevas, & Enders, 2005) focused on the results of interventions, including teacher training and inquiry-based curricular materials designed for a diverse student population. The interventions resulted in significant gains and large effect sizes on science achievement tests. A quasi-experimental study of a diverse population taught using a highly rated guided inquiry unit (Lynch, Kuipers, Pyke, & Szesze, 2005) reported a significant difference in posttest results of the study group and the comparison group. Pine et al. (2006) used performance assessments of inquiry ability to compare the performance of students taught using hands-on curricula with that of students taught using textbook curricula. The researchers found no significant difference in the effects of the two types of curricula. The effect sizes associated with the selected *JRST* articles from the last 2 years are of such magnitude that they would not substantially alter the average effect sizes obtained in the present meta-analysis.

This initial report is limited in scope due to the rigorous inclusion criteria necessitated by the time limits for completion. It is highly recommended that the following types of studies be included in future meta-analyses: (a) international studies; (b) correlational studies; (c) studies dealing with attitudinal and motivational changes in science students and teachers; (d) studies dealing with teaching science to special populations (e.g., English-language learners, students with special needs, underrepresented populations); and (e) studies dealing with science teacher professional development.

The authors acknowledge the contributions of Gina Day, Chris Comer, and Irene Pickhardt at TEA, as well as the project advisory board members, Carol L. Fletcher, Ginny Heilman, Anna McClane, Sandra West, and Jo Ann Wheeler, from the original project. We also thank the project review team members, Katherine Price Blount, Patti Castellano, Diane Jurica, Judy Kelley, Mayra Martinez, Sharon Kyles Ross, and Fernando Ruiz, for their comments on an earlier draft of the meta-analysis. In addition, we are grateful for the valuable feedback of Cathleen C. Loving in helping to improve the manuscript.

Appendix A: Coding Document

CODING OF LITERATURE FOR SCIENCE INITIATIVE META-ANALYSIS

1-3 Study number & citation

4. Pub type: 1 Refereed Jour. Art. 2 Dissertation 3 Unpub. Report 4 Conference paper

5. Type of study: 1 Experimental (complete randomization)
   2 Quasi-experimental (randomization used)
   3 Quasi-experimental (no randomization)
   4 Correlational

6. Test Type		Test Name	#of items	Content Area(s)
	1 National Standardized Multiple Science content	________	____	______
	2 National Standardized Single Science content	________	____	______
	3 Local Standardized Multiple Science Content	________	____	______
	4 Local Standardized Single Science Content	________	____	______
	5Other type test	________	____	______

7. Ind. Var.: TRT. Name ______________________________
   TRT. Description______________________________
   TRT vs ______________________________
   Course ______________
8. Length of TRT: _______Months________Years

9. Setting & Characteristics:
   A. School (s): # of schools___, Selected at random? 1 Yes 2 No, Unit of analysis? 1 Yes 2 No
   1 Public 2 Private, 1 Urban 2 Rural
   1 Small (n=______), 2 Med (n=____), 3 Lrg (n=____)
   % Free Lunch__________

   B. Students: # of students ___, Selected at random? 1 Yes 2 No
   Assigned at random? 1 Yes 2 No, Unit of analysis? 1 Yes 2 No
   1 Mixed Gender 2 all Female 3 all Male
   1 Same Grade 2 Multiple Grades
   Grade ____ n=___, Grade ___ n=__, Grade ___ n=___, Grade___ n=__
   1 Mixed Ethnicity 2 Single ethnic grp
   1 Mixed Socio 2 Single Socio, Primary Socio Status__________________

   C. Teacher(s): # of teachers _________, 1 volunteer 2 selected
   Teacher(s) description: Age/Exp________Gender_______ Certification___________

10. Study results: Effect size (s) ____, p(s) __, t(s) __ , F(s) ___, Eta Square(s) ___ Omega Square(s) ___

11. Study rating: 1 True random assignment of schools/students to TRT and Control
   2 Quasi-experimental with match of schools/students to achievement and demographics of comparison school
   3 Quasi-experimental with covariate adjustment for prior achievement
   4 Quasi-experimental comparison of schools/subjects based on a claim of similar
   5 Quasi-experimental comparison of schools/subjects to region, state, or national data
   6 Quasi-experimental single-group pre-post comparison
   7 Quasi-experimental treatment vs. control pre-posttest
   8 Quasi-experimental multiple group ANOVA
12. Notes/comments:

Appendix B: Study Rating Sheet

Study Treatment Category Rating

Study Number: __________ Judge: 1, 2, 3

1: unknown or not mentioned
2: mentioned
3: a small portion of the treatment
4: moderate portion of the treatment
5: largest portion of the treatment

Criteria	1	2	3	4	5
1. Questioning strategies					
2. Focusing strategies					
3. Manipulation strategies					
4. Enhanced materials strategies					
5. Assessment strategies					
6. Inquiry strategies					
7. Enhanced context strategies					
8. Instructional technology strategies					
9. Direct instruction					
10. Collaborative learning strategies					

Notes/comments:

Appendix C: Included Studies Listed by Category

The following is a list of studies included according to category (number in parentheses at end of citation is the study number).

Questioning Strategies: Three Studies

Osman, M.E., & Hannafin, M.J. (1994). Effects of advance questioning and prior knowledge on science learning. Journal of Educational Research, 88, 5–13. (70)

Riley, J.P. (1986). The effects of teachers' wait-time and knowledge comprehension questioning on science achievement. Journal of Research in Science Teaching, 23, 335–342. (281, 2810)

Manipulation Strategies: Eight Studies

Bluette, C.L., Jr. (1999). An evaluation of the effects of staff development for teachers who utilize multiple strategies with middle school students in mathematics and science. Unpublished dissertation, St. Louis University, St. Louis, Missouri. (32)

Clark, D.R. (2000). Effects of teaching high school chemistry with dynamic particle models on student achievement and conceptual understanding. Unpublished doctoral dissertation, Catholic University of America, Washington, DC. (30, 300)

Frederick, L.R., & Shaw, E.L., Jr. (1999, November 17–19). Effects of science manipulatives on achievement, attitudes, and journal writing of elementary science students revisited. Paper presented at the annual meeting of the Mid-South Educational Research Association, Point Clear, AL (ERIC Document Reproduction Service No. ED 436410). (468)

Glasson, G.E. (1989). The effects of hands-on and teacher demonstration laboratory methods on science achievement in relation to reasoning ability and prior knowledge. Journal of Research in Science Teaching, 26, 121–131. (276)

Matthews, D.R., & McLaughlin, T.F. (1994). Effects of learner-centered laboratory activities on achievement and students preferences in 2 high-school biology courses. Perceptual and Motor Skills, 78, 285–286. (247)

Purser, R.K., & Renner, J.W. (1983). Results of two tenth-grade biology teaching procedures. Science Education, 67, 85–98. (433)

Saunders, W.L., & Shepardson, D. (1987). A comparison of concrete and formal science instruction upon science achievement and reasoning ability of 6th grade students. Journal of Research in Science Teaching, 24, 39–51. (90)

Enhanced Material Strategies: 12 Studies

Hamrick, L., & Harty, H. (1987). Influence of resequencing general science content on the science achievement, attitudes toward science, and interest in science of 6th grade students. Journal of Research in Science Teaching, 24, 15–25. (278)

Hand, B., Wallace, C.W., & Yang, E.M. (2004). Using a science writing heuristic to enhance learning outcomes from laboratory activities in seventh-grade science: Quantitative and qualitative aspects. International Journal of Science Education, 26, 131–149. (50, 500)

Hocutt, M.M. (2003). Comparing instructional methodologies in sixth-grade science: Traditional textbook, integrated science, and integrated science with technology enhancement. Unpublished doctoral dissertation, University of Alabama, Tuscaloosa, Alabama. (6, 609)

Kremer, P.L. (1983). Effects of individualized assignments on biology achievement. Journal of Research in Science Teaching, 20, 105–115. (170)

Lavoie, D.R. (1999). Effects of emphasizing hypothetico-predictive reasoning within the science learning cycle on high school student's process skills and conceptual understandings in biology. Journal of Research in Science Teaching, 36, 1127–1147. (64)

McManus, D.O., Dunn, R., & Denig, S. (2003). Effects of traditional lecture versus teacher-constructed & student-constructed self-teaching instructional resources on short-term science achievement & attitudes. American Biology Teacher, 65, 93–102. (229)

Roberts, P.H. (1999). Effects of multisensory resources on the achievement and science attitudes of seventh-grade suburban students taught science concepts on and above grade level. Unpublished doctoral dissertation, St. John's University, New York, New York. (33)

Romance, N.R., & Vitale, M.R. (1992). A curriculum strategy that expands time for in-depth elementary science instruction by using science-based reading strategies: Effects of a year-long study in grade four. Journal of Research in Science Teaching, 29, 545–554. (439)

Sherris, J.D., & Kahle, J.B. (1984). The effects of instructional organization and locus of control orientation on meaningful learning in high school biology students. Journal of Research in Science Teaching, 21, 83–94. (451)

Turpin, T.J. (2000). A study of the effects of an integrated, activity-based science curriculum on student achievement, science process skills, and science attitudes. Unpublished dissertation, University of Louisiana at Monroe, Monroe, Louisiana. (29)

Testing Strategies: Two Studies

Dillashaw, F.G., & Okey, J.R. (1983). Effects of a modified mastery learning strategy on achievement, attitudes, and on-task behavior of high school chemistry students. Journal of Research in Science Teaching, 20, 203–211. (457)

Long, J.C., Okey, J.R., & Yeany, R.H. (1981). The effects of a diagnostic-prescriptive teaching strategy on student achievement in biology. Journal of Research in Science Teaching, 18, 515–523. (456)

Inquiry Strategies: 12 Studies

Dalton, B., Morocco, C.C., Tivnan, T., & Mead, P.L.R. (1997). Supported inquiry science: Teaching for conceptual change in urban and suburban science classrooms. Journal of Learning Disabilities, 30, 670–684. (448)

Gatlin, L.S. (1998). The effect of pedagogy informed by constructivism: A comparison of student achievement across constructivist and traditional classroom environments. Unpublished doctoral dissertation, University of New Orleans, New Orleans, Louisiana. (38)

Louden, C.K. (1997). Teaching strategies and student achievement in high school block scheduled biology classes. Unpublished dissertation, University of North Carolina, Chapel Hill, North Carolina. (41)

Marx, R.W., Blumenfeld, P.C., Krajcik, J.S., Fishman, B., Soloway, E., Geier, R., et al. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. Journal of Research in Science Teaching, 41, 1063–1080. (78, 780, 781, 782)

Orehowsky, W. (1999). The effect of laboratory-based instruction and assessment on student attitudes toward the laboratory experience and achievement in chemistry at the high school level. Unpublished doctoral dissertation, Temple University, Philadelphia, Pennsylvania. (36)

Rose-Baele, J.S. (2003). Report of fifth grade outcome study, Science for All Students, 2001–2002. Retrieved June 6, 2005, from http://www.ccsdschools.com/administration/assessment/PIpage.html (150)

Schneider, L.S., & Renner, J.W. (1980). Concrete and formal teaching. Journal of Research in Science Teaching, 17, 503–517. (434)

Songer, N.B., Lee, H.S., & McDonald, S. (2003). Research towards an expanded understanding of inquiry science beyond one idealized standard. Science Education, 87, 490–516. (142, 1420)

Enhanced Context Strategies: Six Studies

Berube, C.T. (2001). A study of the effects of constructivist-based vs. traditional direct instruction on 8th grade science comprehension. Unpublished doctoral dissertation, Old Dominion University, Norfolk, Virginia. (21)

Fortus, D., Dershimer, R.C., Krajcik, J., Marx, R.W., & Mamlok-Naaman, R. (2004). Design-based science and student learning. Journal of Research in Science Teaching, 41, 1081–1110. (60, 600, 601)

Yager, R.E. (1989). Comparison of standard student performance when science study is organized around typical concepts. Bulletin of Science, Technology, & Society, 9, 171–181. (98)

Yager, R.E., & Weld, J.D. (1999). Scope, sequence and coordination: The Iowa Project, a national reform effort in the USA. International Journal of Science Education, 21(2), 169–194. (369)

Instructional Technology Strategies: 15 Studies

Akpan, J.P., & Andre, T. (2000). Using a computer simulation before dissection to help students learn anatomy. Journal of Computers in Mathematics and Science Teaching, 19, 297–313. (475)

Baker, T.R., & White, S.H. (2003). The effects of GIS on students' attitudes, self-efficacy, and achievement in middle school science classrooms. Journal of Geography, 102, 243–254. (48)

Dean, D.M. (2004). An evaluation of the use of Web-enhanced homework assignments in high school biology classes. Unpublished doctoral dissertation, University of Alabama, Tuscaloosa, Alabama. (1)

Demirci, N. (2001). The effects of a Web-based physics software program on students' achievement and misconceptions in force and motion concepts. Unpublished doctoral dissertation, Florida Institute of Technology, Melbourne, Florida. (25)

Faro, S.T. (2003). An investigation of components of the studio model and supplemental online materials, on student achievement and attitudes in science at the high school level. Unpublished doctoral dissertation, State University of New York at Albany, Albany, New York. (4)

Harwood, W.S., & McMahon, M.M. (1997). Effects of integrated video media on student achievement and attitudes in high school chemistry. Journal of Research in Science Teaching, 34, 617–631. (57)

Huffman, D., Goldberg, F., & Michlin, M. (2003). Using computers to create constructivist learning environments: Impact on pedagogy and achievement. Journal of Computers in Mathematics and Science Teaching, 22, 151–168. (452, 4520)

Jafer, Y.J. (2003). The effects of computer-assisted instruction on fourth-grade students' achievement and attitudes toward desert issues. Unpublished doctoral dissertation, Utah State University, Logan, Utah. (7)

Kariuki, P., & Paulson, R. (2001, November 14–17). The effects of computer animated dissection versus preserved animal dissection on the student achievement in a high school biology class. Paper presented at the annual conference of the Mid-South Educational Research Association, Little Rock, AK (ERIC Document Reproduction Service No. ED460018). (467)

Kim, N.B. (1997). A comparison of the effects of computer-enhanced with traditional instruction on the learning outcomes of high-school students in anatomy classes. Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh, Pennsylvania. (39)

Marszalek, C.S. (1998). Effects on seventh-grade students' achievement and science anxiety of alternatives to conventional frog dissection. Unpublished doctoral dissertation, Northern Illinois University, DeKalb, Illinois. (47, 470)

Pallant, A., & Tinker, R. (2004). Reasoning with atomic-scale molecular dynamic models. Journal of Science Education and Technology, 13, 51. (75, 750)

Collaborative Learning Strategies: Three Studies

Chang, H.P., & Lederman, N.G. (1994). The effect of levels of cooperation within physical science laboratory groups on physical science achievement. Journal of Research in Science Teaching, 31, 167–181. (166)

Houtz, L.K.E. (1995). Instructional strategy change and the attitude and achievement of 7th-grade and 8th-grade science students. Journal of Research in Science Teaching, 32, 629–648. (245)

Johnson, R.T., Johnson, D.W., Scott, L.E., & Ramolae, B.A. (1985). Effects of single-sex and mixed-sex cooperative interaction on science achievement and attitudes and cross-handicap and cross-sex relationships. Journal of Research in Science Teaching, 22, 207–220. (285)

References

American Association for the Advancement of Science. (1993). Benchmarks for science literacy. New York: Oxford University Press.

American Psychological Association. (2001). Publication manual of the American Psychological Association (5th ed.). Washington, DC: American Psychological Association.

Anderson, R.D., Kahl, S.R., Glass, G.V., & Smith, M.L. (1983). Science education: A meta-analysis of major questions. Journal of Research In Science Teaching, 20, 379–385.

Bransford, J.D., Brown, A.L., & Cocking, R.R. (Eds.). (2000). How people learn: Brain, mind, experience, and school (expanded edition). Washington, DC: National Academy Press.

Hunt, M. (1997). How science takes stock: The story of meta-analysis. New York: Russell Sage Foundation.

Li, J., & Klahr, D. (2006). The psychology of scientific thinking: Implications for science teaching and learning. In J. Rhoton & P. Shane (Eds.), Teaching science in the 21st century. Arlington, VA: NSTA Press.

Marzano, R.J. (1998). A theory-based meta-analysis of research on instruction: Mid-continent Regional Educational Laboratory.

Matthews, M.R. (1994). Science teaching: The role of history and philosophy of science. New York: Routledge.

McNamara, J.F., Morales, P., Kim, Y., & McNamara, M. (1998). Conducting your first meta-analysis: An illustrated guide. International Journal of Educational Reform, 7, 380–397.

National Center for Education Statistics. (2004). Comparing NAEP, TIMSS, and PISA in mathematics and science. Retrieved September 1, 2006, from http://nces.ed.gov/TIMSS/pdf/naep_timss_pisa_comp.pdf#search=%22comparing%20NAEP%2C%20timss%2C%20and%20pisa%20in%20mathematics%20and%20science%22

National Center for Education Statistics. (2005). Science: The nation's report card. Retrieved May 2, 2006, from http://nces.ed.gov/nationsreportcard/science/

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

Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? Educational Evaluation and Policy Analysis, 26, 237–257.

Rivkin, S.G., Hanushek, E.A., & Kain, J.F. (2005). Teachers, schools, and academic achievement. Econometrica, 73, 417–458.

Rutherford, F.J., & Ahlgren, A. (1990). Science for all Americans. New York: Oxford University Press.

Shavelson, R.J., & Towne, L. (Eds.). (2002). Scientific research in education. Washington, DC: National Academy Press.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 15, 4–14.

Shymansky, J.A., Yore, L.D., & Anderson, J.O. (2004). Impact of a school district's science reform effort on the achievement and attitudes of third- and fourth-grade students. Journal of Research in Science Teaching, 41, 771–790.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. Educational Researcher, 31, 25–32.

Wayne, A.J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. Review of Educational Research, 73, 89–122.

Wise, K.C. (1996). Strategies for teaching science: What works? Clearing House, 69, 337–338.

Wise, K.C., & Okey, J.R. (1983). A meta-analysis of the effects of various science teaching strategies on achievement. Journal of Research in Science Teaching, 20, 419–435.

Wright, S.P., Horn, S.P., & Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. Journal of Personnel Evaluation in Education, 11, 57–67.