

*Prentice Hall  
Science Explorer*

# Research Summary

Foundational Research  

---

Program Efficacy Studies



# *Prentice Hall Science Explorer*

## *Research Summary*

### *Contents*

---

<b>I.</b>	<b>Prentice Hall's Research Methodology</b>	<b>2</b>
	Exploratory	
	Formative	
	Summative	
<b>II.</b>	<b>Putting Research Into Practice</b>	<b>4</b>
	Inquiry	
	Differentiated Instruction	
	Reading	
	Assessment	
<b>III.</b>	<b>Results</b>	<b>9</b>
	Program Efficacy Study	
	National Effect-Size Study	
	Japanese Lesson Study	
<b>IV.</b>	<b>Beyond the Data: Practitioners Tell Us Why It Works</b>	<b>20</b>
<b>V.</b>	<b>NCLB Implications for Research</b>	<b>22</b>
<b>VI.</b>	<b>Evaluator's Checklist</b>	<b>24</b>
<b>VII.</b>	<b>Independent, Longitudinal Study Plans in Progress</b>	<b>27</b>

# Prentice Hall's Research Methodology

The stakes for science educators are high. The plain fact is, you are expected to raise student achievement using research-based methods with measurable results. Prentice Hall understands your dedicated efforts and gives you the confidence to meet this challenge. In developing Prentice Hall programs, the use of research studies is a central, guiding construct. Research on *Prentice Hall Science Explorer* indicated key elements of a science program that ensure student success: considerate text, consistent skills instruction, and an ongoing assessment strand. This research was conducted in three phases:

**Phase 1:** Exploratory Needs Assessment

**Phase 2:** Formative, Prototype Development and Field Testing

**Phase 3:** Summative, Validation Research

## 1 Exploratory Needs Assessment

In conjunction with Prentice Hall authors, research was done to explore educational research about science. This research was incorporated into our instructional strategy and pedagogy to make a more effective science program.

Along with periodic surveys concerning curriculum issues and challenges, we conducted specific product development research, which included discussions with teachers and advisory panels, focus groups, and quantitative surveys. We explored the specific needs of teachers, students, and other educators regarding each book we developed in *Prentice Hall Science Explorer*.

## 2 Formative, Prototype Development and Field Testing

During this phase of research, we worked to develop prototype materials for each feature in *Prentice Hall Science Explorer*. Then we tested the materials, including field testing with students and teachers, and qualitative and quantitative evaluations of different kinds. We received solid feedback in our early prototype testing. Results were channeled back into the program development for improvement. For example, teachers commented positively on the easy navigation of instructional pages.

## 3 Summative, Validation Research

Finally, we are conducting independent, longitudinal research using quasi-experimental and experimental research designs. A national, quasi-experimental effect size study is underway to compare efficacy of *Prentice Hall Science Explorer* users versus matched non-users. A longitudinal, experimental study will begin in the fall of 2004 to study specific program features in addition to efficacy and fidelity of implementation. *Prentice Hall Science Explorer* users will be compared to matched non-users and tested using national standardized examinations. This information will inform the need for revision, monitor student success, and identify how well our program works.

# Prentice Hall Research Time Line

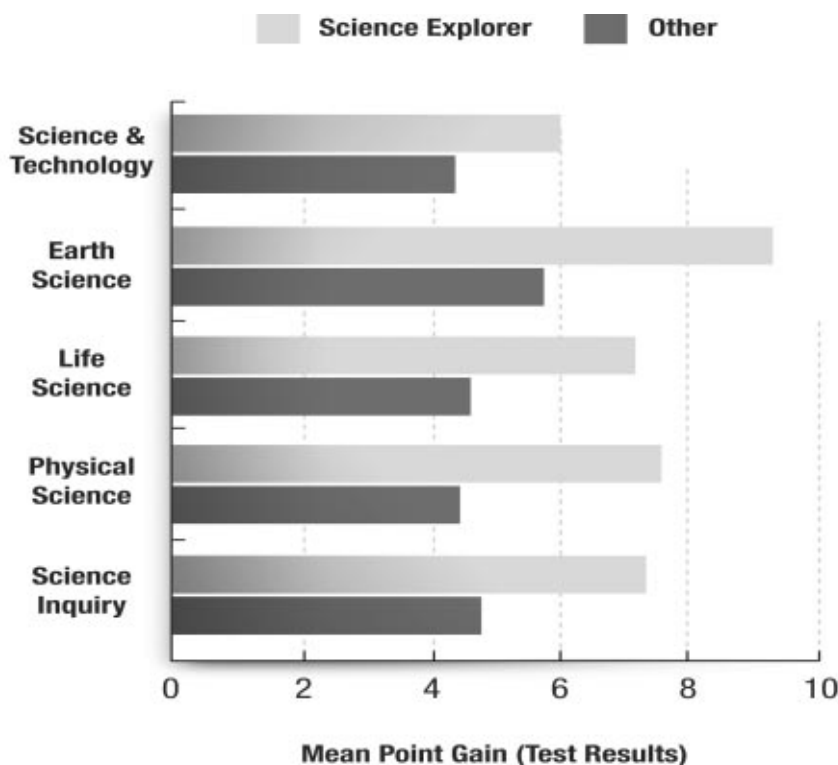


## Summative Research Summary

In a year-long study\* in 2000–2001, students in six states using *Prentice Hall Science Explorer* outscored students using other science programs on a nationally normed standardized test.

The study investigated the effects of science textbook programs at the eighth-grade level. Twelve eighth-grade science classes with a total of 223 students participated in the study. The selected classes were of similar student ability levels.

**Each class was tested\*\* at the beginning of the school year using the TerraNova™ CTBS Basic Battery Plus, and then retested at the end of the school year. The final results, shown in the graph, show a significant improvement in test scores from the pre-test to the post-test evaluation.**



\*For a complete report, see page 9.

\*\*All tests were scored by CTB/McGraw-Hill, the publisher of the TerraNova exam. Statistical analyses and conclusions were performed by an independent firm, Pulse Analytics, Inc.

# Putting Research into Practice

## Inquiry

### *What Research Indicates*

The National Research Council in *Inquiry and the National Science Education Standards* (2000), identifies several “essential features” of classroom inquiry, including the following:

*Who asks the question?* As an educator, you need to look for labs that, at least on a periodic basis, allow students to pursue their own questions.

*Who designs the procedures?* In order to gain experience with the logic underlying experimentation, students need continuous practice with designing procedures.

*Who decides what data to collect?* Students need practice in determining the critical data to collect.

*Who formulates explanations based upon the data?* Activities should make students analyze and draw conclusions based on their data.

*Who communicates and justifies the results?* Activities should be thoughtfully designed and interesting so that students want to share their results and justify their conclusions.

*Michael Padilla*  
*Prentice Hall Author*

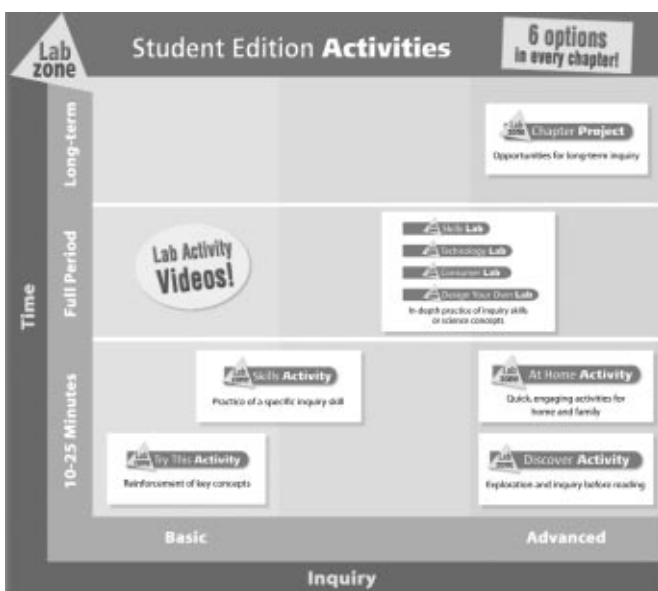
### *Prentice Hall's Response*

In *Prentice Hall Science Explorer*, each activity is designed to stretch and extend students' inquiry abilities.

**Discover** focuses on introducing the content that follows in the book with insightful and open-ended questions to stimulate student thinking.

**Sharpen Your Skills** emphasizes the components of inquiry, such as observing, inferring and graphing.

The **Lab Activities** vary considerably, with a few specifying both procedures and data to collect, and many others asking students to formulate a question, design their own experiment, and decide what data is important.



# Differentiated Instruction

## What Research Indicates

All students are unique and their differences do not necessarily remain constant. Their strengths and needs change as their literacy develops and as they grow in their knowledge of science. Instruction that attempts to meet the needs of all students must be flexible and adaptable for each individual, and assessment of students' strengths and needs must be continuous to ensure that each student learns all he or she can.

Instruction must be multi-sensory with learning opportunities that rely on all of the senses, and it must be scaffolded for each student's learning level. Further, instruction with multimedia enhances the probability of each student's learning.

James Flood  
Diane Lapp  
San Diego State University

## Prentice Hall's Response

The "Differentiated Instruction" feature provides techniques for adapting the resources for a variety of students. Inside the *All in One Teaching Resources*, review and reinforce worksheets and guided reading and study worksheets provide practice for students who are struggling. Through a blend of print and technology resources, all labs/activities are leveled L1-L3, from basic to advanced.

Chapter 4 E • 141

### Differentiated Instruction

**Special Needs**

**Writing Captions** Pair students with more able students. Distribute photocopies of Figure 18. Suggest that students write a caption for each source of indoor pollution shown. Then have students add other sources of air pollution to the drawing (indoors or from outdoors) and write captions for those. **learning modality: visual**

**Less Proficient Readers**

**Analyzing Word Parts** Write carbon dioxide and carbon monoxide on the board, and ask students to identify the difference between the words. Explain that *di* means "two," and *mono* means "one." Carbon dioxide has two atoms of oxygen, and carbon monoxide has one atom. Draw a diagram of each molecule. **learning modality: visual**

**Monitor Progress**

**Oral Presentation** Call on students to name one source of air pollution.

**Answers**

**Figure 17** Acid rain has reacted with the statues, eating away some of the stone.

**Figure 18** Cigarette smoke, cat and dog dander, glue, nail polish, paint, cleansers

**Reading Checkpoint** Acid rain reacts with stone statues and metal in buildings and damages them.

**Reading Checkpoint** A colorless, odorless gas that forms when wood, coal, oil, or gas is incompletely burned

E • 141

## Reading

### What Research Indicates

Modern understanding of comprehension emphasizes the importance of in-depth prior knowledge and how that knowledge is organized. Studies comparing novices and experts show that the conceptual organization of experts' knowledge is very different from that of novices. For example, experts emphasize core concepts when organizing knowledge while novices focus on superficial details. To facilitate comprehension, effective teaching strategies should support and scaffold students as they build an understanding of the key concepts and concept relationships within a text unit.

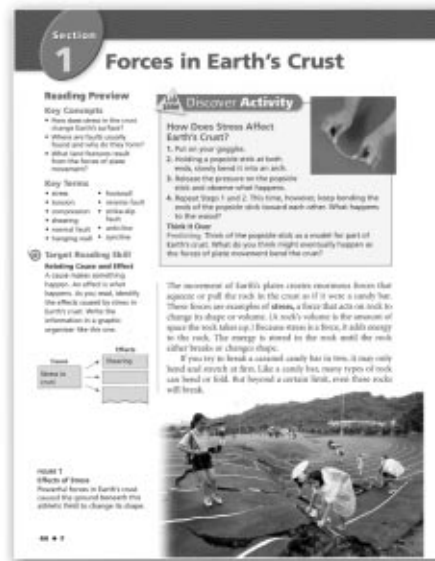
Three complementary strategies are very important in facilitating student comprehension of science texts. First, guide student interaction with the text using the built-in strategies. Second, organize the curriculum in terms of core concepts (e.g., Key Concepts in each Section). Third, develop visual representations of the relationships among the key concepts and vocabulary that can be referred to during instruction.

Dr. Nancy Romance  
Florida Atlantic University

## Prentice Hall's Response

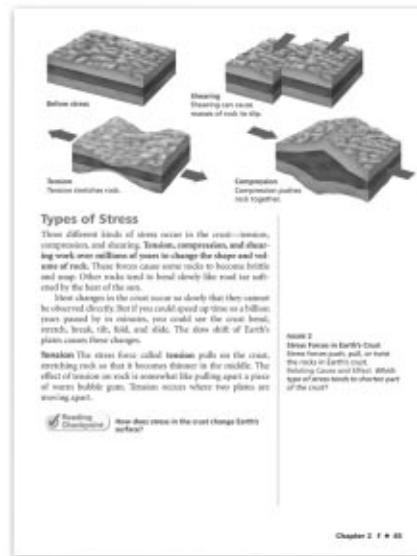
The concept of 'considerate' text has been built into the Student Edition through the careful design of text features that provide instructional support for comprehension. Each of these instructional features provides a strong literacy framework for science teachers as they guide student interaction with the text. Each, in turn, also serves as a scaffold for students as they become more independent learners.

Before students read—



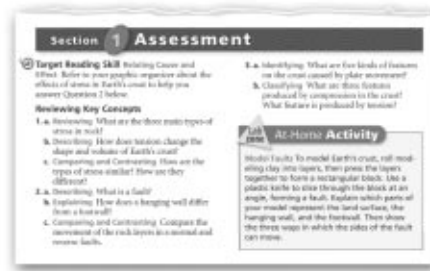
Reading Preview introduces students to the key concepts and key terms they'll find in each section. The Target Reading Skill is identified and applied with a graphic organizer.

During the section—



Boldface Sentences identify each key concept and encourage students to focus on the big ideas of science. Reading Checkpoints reinforce students' understanding by slowing them down to review after every concept is discussed. Caption Questions draw students into the art and photos, helping them connect the content to the images.

After students read—



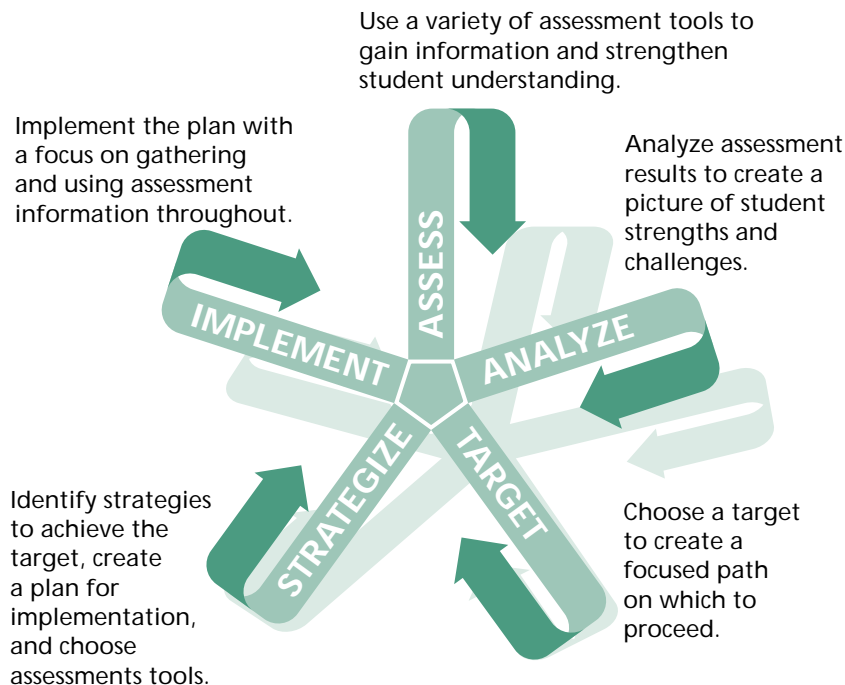
Section Assessment revisits the Target Reading Skill and encourages students to use the graphic organizer. Each review question is scaffolded and models the way students think, first by easing them into a review, and then challenging them with increasingly more difficult questions.

## Assessment

### What Research Indicates

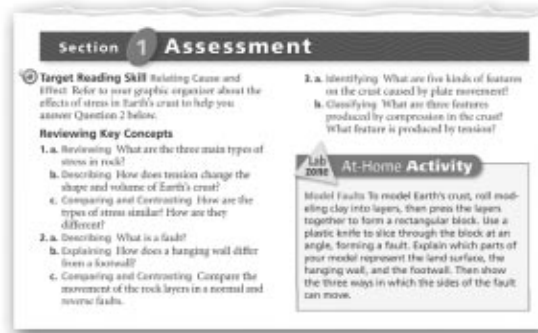
An important assessment strategy is to ensure that students have ample opportunities to check their understanding of skills and concepts before moving on to the next topic. Checking for understanding also includes asking appropriate, probing questions with each example presented. This enables students and teachers to know whether the skills or concepts being introduced are actually understood.

*Eileen Depka*  
*Supervisor of Standards and Assessment*  
*Waukesha, Wisconsin*



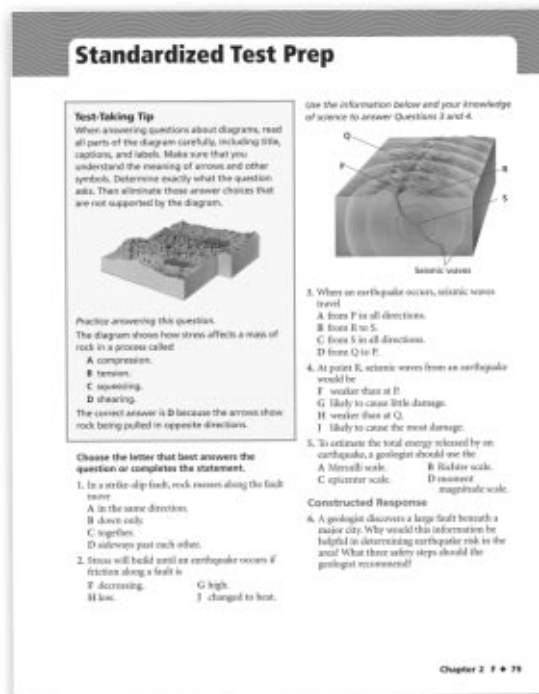
## Prentice Hall's Response

The assessment strategies in *Prentice Hall Science Explorer* will help both students and teachers alike ensure student success in content mastery as well as high-stakes test performance. A wealth of opportunities built into the Student Edition help students monitor their own progress. Teachers are supported with ongoing assessment opportunities in the Teacher's Edition and an easy-to-use, editable test generator linked to content objectives. These integrated, ongoing assessment tools assure success.



Especially to support state and national testing objectives, Prentice Hall has developed test preparation materials that model the NCLB approach.

- **Diagnostic Assessment** tools provide in-depth analysis of strengths and weaknesses, areas of difficulty, and probable underlying causes that can help teachers make instructional decisions and plan intervention strategies.
- **Progress Monitoring** tools aligned with content objectives and state tests provide ongoing, longitudinal records of student achievement, detailing individual student progress toward meeting end-of-year and end-of-schooling grade level, district, or state standards.
- **Outcomes** tools that mimic state and national tests show whether individual students have met the expected standards and can help a school system judge whether it has made adequate progress in improving its performance year by year.



# Results

## Program Efficacy Study

### ***I. Abstract***

This study investigated the effects of science textbook programs at the eighth grade level. Eighth grade science students were assigned to either a study group (Prentice Hall text) or control group (other texts). Students were tested at the start of the academic year with a nationally normed standardized test (the TerraNova™ CTBS Complete Battery Plus). At the end of the full year treatment period, students were re-tested with the same standardized test.

The eighth grade study classes utilizing the *Prentice Hall Science Explorer* program showed significant learning improvement over the course of a full school year; the control group (using other textbook programs) did not show significant learning improvement over this same period. In addition, students using *Prentice Hall Science Explorer* showed directionally higher gains in key science instructional areas (earth science, physical science, life science and science inquiry) than students using other textbook programs.

### ***II. Background***

The growing national trend toward greater school/teacher accountability has precipitated an increase in demand for validation of textbook programs from educational publishers. With this in mind, a full-year learner verification study for the *Prentice Hall Science Explorer* program ©2000 (1/e) was initiated for the 2000-2001 school year in order to measure student learning improvement using this Prentice Hall program versus competitive products in the marketplace. Results are intended to benefit both marketing efforts and product development needs based upon student performance.

### ***III. Objectives***

The objectives of this research were to:

1. Provide evidence of program effectiveness by demonstrating learner improvement over the course of a school year.
2. Determine whether the improvement of students who are enrolled in classes using *Prentice Hall Science Explorer* is above the level of performance of those not using the Prentice Hall program.
3. Obtain input from teachers and students after in-class usage for a full year to guide program development and marketing/sales communications.

## IV. Methodology

The study followed an experimental design, with a study group to whom the Prentice Hall program was administered and a control group to whom the program was not administered. The control group used the incumbent science program that had previously been adopted for use in the school. Pre- and post-tests were administered to both groups in order to compare and verify learning over the course of the school year.

### Program Efficacy Study: Experimental Design

	Study Group	Control Group
Pre-Test (administered during first week of school)	6 classes 108 students	6 classes 115 students
Post-Test (administered at end of school year)	6 classes 108 students	6 classes 115 students

Eighth grade science teachers in four states—Colorado, New Jersey, Washington, and Wisconsin—were recruited from an MDR list for participation in this study. Although only six schools participated in this study, the schools selected reflected a mix of enrollment sizes and metropolitan areas. The teachers selected for participation also reflect a mix of experience levels and backgrounds.

In order to qualify for the study, teachers were subject to the following requirements:

- Teachers could not currently be serving on a textbook adoption committee.
- They could not be affiliated in any way with a major publisher (past or present).
- They were required to have at least three years of teaching experience.
- For the 2000-2001 school year, they were required to have at least two classes of eighth grade science that were similar in student profile, either average students or students of mixed ability levels.
- In addition, these two classes were required to have approximately the same number of students.
- The teacher had to maintain the same instructional pace and cover the same topics in these two classes, and
- Participating teachers could not currently be using a Prentice Hall or Scott Foresman-Addison Wesley program.

As an incentive for participating in this research, teachers were given the *Science Explorer* program materials, including a class set of Student Editions, Teacher's Editions, and most program ancillaries/technology components. These materials were theirs to keep at the completion of the study.

## IV. Methodology (continued)

For the purposes of this research, teachers were asked to designate at random one of their eighth grade science classes as the study class, and another of their eighth grade science classes (one of similar size and ability level) as the control class. The study class was taught throughout the school year using the *Science Explorer* textbook program; the control class was taught (by the same teacher<sup>1</sup>) using the school's incumbent science program.

For the Pre-Test phase, both study class students and control class students were tested during the first week of school (August/September, 2000) using the TerraNova™ CTBS Complete Battery Plus nationally normed standardized test. For the Post-Test phase at the end of the school year (May/June, 2001), both study and control classes were tested again using the same test. Only students who completed both the pre- and post-tests were included in this analysis.

Tests were scored independently by the publisher of the TerraNova,™ CTB/McGraw Hill. The test scores were then compared in order to measure student learning improvement using the *Science Explorer* textbook program versus competitive science textbook programs. Statistical analysis of these scores was conducted by an independent statistician at Pulse Analytics, Inc.

The measure used in this analysis was the NCE (Normal Curve Equivalent) of the TerraNova™ CTBS Complete Battery Plus. In addition, the Objectives Performance Index (OPI) developed by CTB/McGraw Hill was used in this analysis to measure student performance in various specific skill and content areas before and after using the textbook programs.

Prentice Hall National Sales Consultants were enlisted to provide in-service training to participating teachers. Consultants were available to each teacher on an individual basis to answer questions about the *Science Explorer* program and questions pertaining to the study. The Consultants also encouraged teachers to provide feedback on the program at various points throughout the year.

In addition, at the end of the school year, participating teachers and students who used *Science Explorer* completed evaluation questionnaires detailing their overall reactions to the program. (Note: All of the teachers who participated in this study have given their permission to include their names and comments in this report. However, student names have not been revealed).

## V. Standardized Test Results

Statistical tests were used to examine two issues with regard to program effectiveness: 1) that science knowledge and skills increased among students using *Prentice Hall Science Explorer* from the pre-test to the post-test, and 2) the gains for the study group (*Prentice Hall Science Explorer*) were above that of the control group (other texts).

**Step 1.** This step consisted of an analysis of whether or not there was a statistically significant difference between the study group and the control group on the pre-test score. For this, a *t-test* on the difference between the study and control mean pre-test scores was used. The measure used in this analysis was the NCE (Normal Curve Equivalent) of the TerraNova™ CTBS Complete Battery Plus. (See Technical Post Script on page 17 for a complete explanation of the NCE measurement.)

The hypothesis of the *t-tests* shown below is that the pre-test means of the study and control groups are equal; the alternative hypothesis is that they are not equal, for which a two-tail test is appropriate. If the significance of the *t-value* is less than or equal to .05, then the hypothesis is rejected at the 95% (5% significance) level and the alternate hypothesis is accepted.

---

<sup>1</sup>In one case only, where the same teacher could not teach both study and control classes, another teacher from the same school with similar teaching style, background, and tenure was selected to teach the control class.

Since the significance of the *t-value* is greater than .05 (that is, 0.926, as shown below), we accept the hypothesis; that is, that the pre-test means of study and control are equal at the 95% level of confidence, showing no significant differences at the starting point of the study.

Subject, Grade	Study Pre-NCE (base)	Control Pre-NCE (base)	Absolute Difference	<i>t-value</i>	Sig ( <i>t-value</i> )
Science, 8th	54.15 (108)	54.34 (115)	0.19	0.092	<b>0.926</b>

## V. Standardized Test Results (continued)

**Step 2.** This step consisted of an analysis of whether or not there was a statistically significant difference between the pre-test and post-test scores within the study group and within the control group. For each group, a *t-test* on the difference between pre-test and post-test scores was used. The NCE was the measure used in this analysis.

The hypothesis of the *t-tests* shown below is that the pre-test and post-test means are equal; the alternative hypothesis is that they are not equal. If the significance of the *t-value* is less than or equal to .05 then the hypothesis is rejected at the 95% (5% significance) level and the alternate hypothesis is accepted.

For the study group, the significance of the *t-value* is less than .05 (that is, 0.001, as shown below). Therefore we accept the alternative hypothesis: the pre-test and post-test means of the study group are not equal at the 95% level of confidence.

- **Thus, students using the *Prentice Hall Science Explorer* program showed significant improvement in test scores from the pre-test to the post-test.**

In contrast, for the control group, the significance of the *t-value* is greater than .05 (that is, 0.174, as shown below). Therefore we accept the original hypothesis: the pre-test and post-test means of the control group are equal at the 95% level of confidence.

- **Thus, students using other textbook programs did not show significant improvement in test scores from the pre-test to the post-test.**

Subject, Grade	Study Pre-test NCE (base)	Study Post-test NCE (base)	Absolute Difference	<i>t-value</i>	Sig ( <i>t-value</i> )
Science, 8th	54.15 (108)	58.38 (108)	4.23	3.53	<b>0.001</b>

Subject, Grade	Control Pre-test NCE (base)	Control Post-test NCE (base)	Absolute Difference	<i>t-value</i>	Sig ( <i>t-value</i> )
Science, 8th	54.34 (115)	55.95 (115)	1.61	1.37	<b>0.174</b>

## V. Standardized Test Results (continued)

**Step 3.** In addition to NCE scores, performance on science instructional objectives for the study and control groups was also analyzed. The measure used in this analysis was the OPI (Objectives Performance Index) of the TerraNova™ CTBS Complete Battery Plus. (see Technical Post Script on page 17 for a complete explanation of the OPI measurement.)

Both students using the *Prentice Hall Science Explorer* program (the study group) and students using other textbook programs (the control group) showed significant gains in key science instructional areas over the course of a full school year. However, students using the *Science Explorer* program appear to show gains that are directionally higher than students using other textbook programs—especially in the areas of Earth Science and Physical Science:

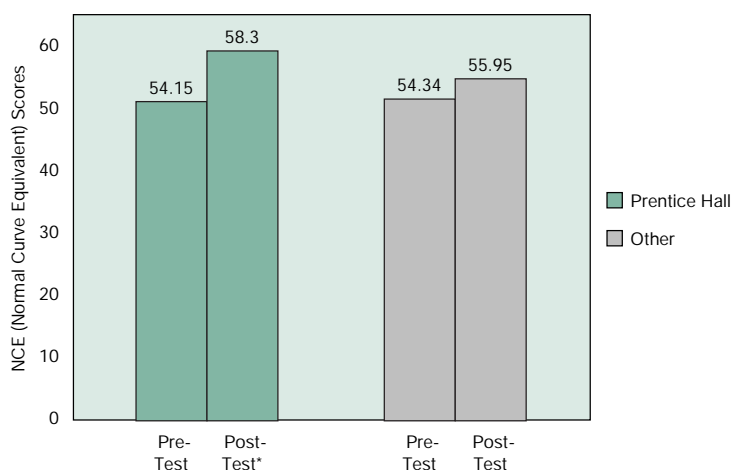
8 <sup>th</sup> Grade Science Instructional Objectives:	Study Mean Point Gain (pre to post)	Control Mean Point Gain (pre to post)
Science Inquiry	+7.30	+4.73
Physical Science	+7.56	+4.42
Life Science	+7.17	+4.61
Earth Science	+9.34	+5.74
Science and Technology	+6.00	+4.35
Personal and Social Perspectives in Science	+3.91	+3.05
<b>TOTAL GAIN</b>	<b>+41.28</b>	<b>+26.90</b>

In using *Science Explorer* for a full school year, students showed the most marked improvement in Earth Science, Physical Science, Life Science, and Science Inquiry.

## VI. Conclusions

The study group (using the *Prentice Hall Science Explorer* program) showed significant learning improvement over the course of a full school year; the control group (using other textbook programs) did not show significant learning improvement over this same period. In addition, students using *Science Explorer* showed directionally higher gains on key science instructional objectives compared to students using other textbook programs.

The analysis of results also showed that teachers' interactions with the *Science Explorer* program played a significant role in the outcome of student performance. Additional study of this phenomenon is suggested.



\*Significantly higher than the pre-test score at the 95% level of confidence.

## ***Technical Post Script***

For the purpose of completeness, we include here a brief discussion of the statistical tests and measures that were used. All of the statistical tests are in the classical statistical domain, and are broadly used across all disciplines including psychometrics. The programs that were used in the computational steps of this project were SPSS versions 6/9.

The *t*-test that was used states: *The mean of one population is equal to the mean of another population when the variance is unknown*. The hypothesis is that the means are equal; and the alternative hypothesis is that they are unequal; for which we use a two-sided test. The means and the variance are calculated and the *t*-statistic is computed. The significance of the *t*-statistic is then computed. If the significance of the *t*-statistic is less than .05 the hypothesis is rejected; otherwise, the alternative hypothesis is accepted.

Directional Differences: are cited as well as significant differences if they appear to lend understanding to the study results. Directional differences should be approached with more caution than significant differences.

Normal Curve Equivalents: "Comparison of Scores across Tests. The normal curve equivalent (NCE) scale, ranging from 1 to 99, coincides with the national percentile scale (NP) at 1, 50, and 99. NCEs have many of the same characteristics as percentile ranks, but have the additional advantage of being based on an equal-interval scale. The difference between two successive scores on the scale has the same meaning throughout the scale. This property allows you to make meaningful comparisons among different achievement test batteries and among different tests within the same battery. You can compare NCEs obtained by different groups of students on the same test or test battery by averaging the test scores for the groups."<sup>2</sup>

Objectives Performance Index: "The OPI is an estimate of the number of items that a student would be expected to answer correctly if there had been 100 similar items for that objective... The OPI scale runs from '0' for total lack of mastery to '100' for complete mastery. For CTB achievement tests, OPI scores between 0 and 49 are regarded as the Non-Mastery level. Scores between 50 and 74 are regarded as indications of Partial Mastery. Scores of 75 and above are regarded as the Mastery level."<sup>3</sup>

---

<sup>2</sup>Quoted directly from: Teacher's Guide to TerraNova, page 138, CTB/McGraw-Hill 1999.

<sup>3</sup>Quoted directly from: Beyond the Numbers, A Guide to Interpreting and Using the Results of Standardized Achievement Tests, page 11, CTB/McGraw-Hill 1997.

# ***Science Explorer*** **National Effect-Size Study Executive Summary:**

## **An Evaluation of Student Performance of Prentice Hall Users versus Non-Users**

### ***Background***

Prentice Hall retained the services of Gatti Evaluation, Inc., an independent research firm that designs and performs evaluations of the effects of educational programs on student performance, to conduct a large-scale, national study on the impact of *Prentice Hall Science Explorer* on student achievement levels.

### ***Study Design***

The quasi-experimental study design examined longitudinal test results of 337 *closely matched* user and non-user districts as a point of comparison across the same time periods and achievement tests. Six states\* were examined as part of the study. Because each state test is unique and measures different standard benchmarks, the analysis measured national percentile ranking (NPR), and percent of students meeting or exceeding state standards. *Note: 57% of districts included in this study were urban districts as defined by NCES (National Center for Education Statistics).*

### ***Matched Districts***

Districts were matched based on the following demographic characteristics:

- similar metropolitan location (e.g., urban, suburban, rural)
- enrollment, ethnicity (e.g., percent white/Caucasian school aged children)
- relative wealth (e.g., percent receiving free or reduced-priced lunch)
- time period in which the two groups were studied
- methods used to collect outcomes

### ***Goals***

The main goal of the study was to quantify the comparative impact of *Prentice Hall Science Explorer* on district Science achievement for sixth, seventh, and eighth grade students. The results may be used to answer critical questions for potential users such as:

- How likely are school districts that adopt *Prentice Hall Science Explorer* to see immediate gains in Science achievement on state-sponsored norm referenced and criterion referenced tests?
- Are districts that adopt *Prentice Hall Science Explorer* as likely or more likely to see achievement gains as similar non-users districts?

### ***Results***

Overall, **77% of districts that adopted *Prentice Hall Science Explorer* saw a significant increase from pre- to post- implementation in student achievement levels.** On average, the percentage of Prentice Hall users meeting state standards increased from 45% to 61%, whereas matched districts increased from 45% to 57%.

---

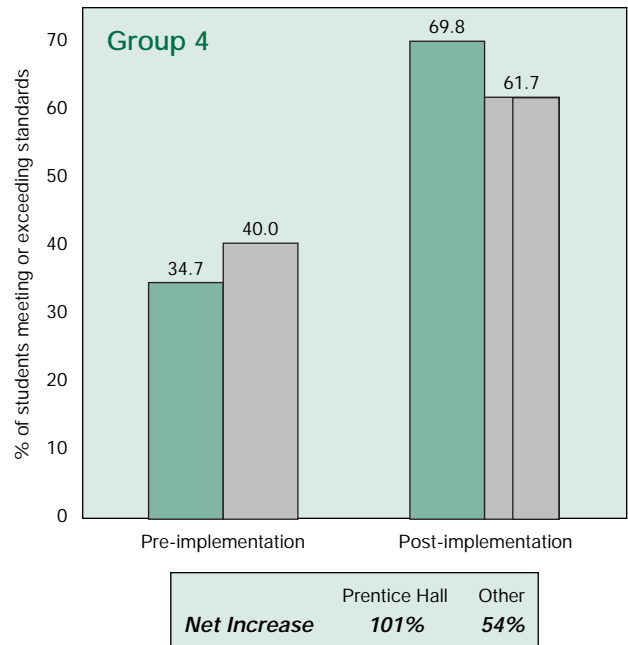
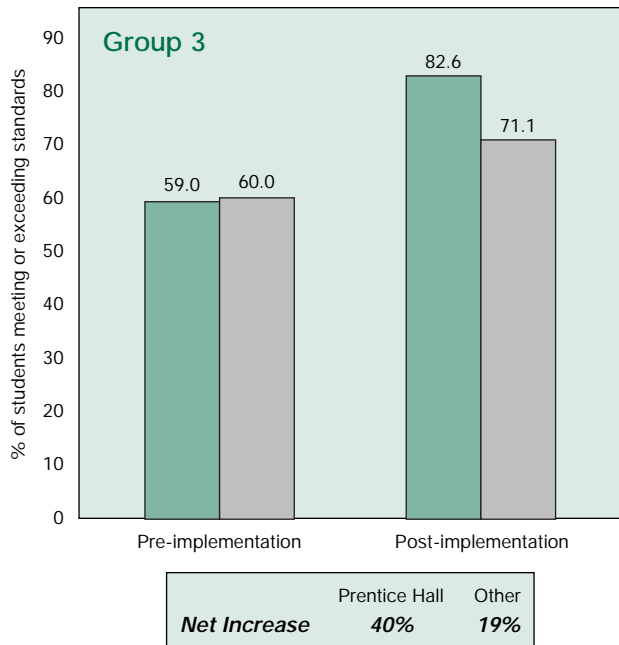
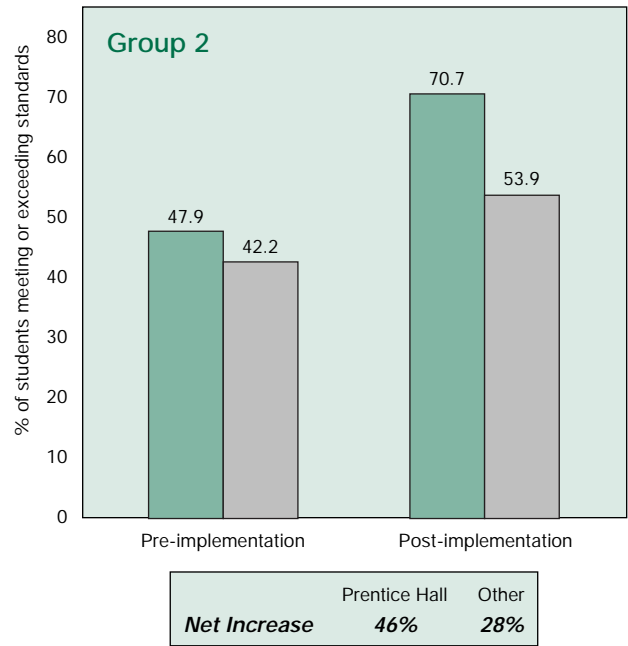
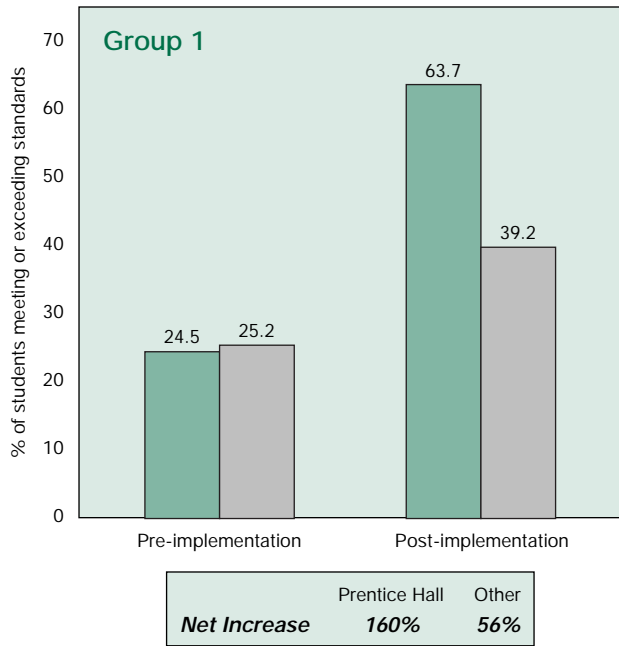
\*States included in the study: AL, IL, MI, MO, OH, WI

If these results are generalizable, they indicate:

- Districts adopting *Prentice Hall Science Explorer* are likely to see immediate gains in science achievement outcomes.
- Districts adopting *Prentice Hall Science Explorer* are as likely or more likely to see statistically significant achievement gains as similar non-user districts.

**Research indicates: *Prentice Hall Science Explorer* is bringing more students up to state standards and increasing district rankings.**

## Science Explorer National Effect-Size Study Highlights



**Key:** ■ Prentice Hall  
■ Other

# Japanese Lesson Study

## ***Background and Methodology***

As part of a comprehensive plan to provide scientifically-based research for the *Prentice Hall Science Explorer* program, a modified Japanese Lesson Study approach was used to field test the ©2005 prototype under actual classroom conditions. In Japan, Lesson Study Research has been employed for a number of years as a tool for teachers to improve their curriculum. We adapted this methodology to focus on a Prentice Hall lesson, in order to test the effectiveness of and improve lesson pedagogy while in the program development stage.

This qualitative approach allows one to test, observe, and improve lessons as they are being developed. The evaluation takes place under actual classroom conditions, where the lesson is judged based on student learning and receptivity to the lesson pedagogy, and the evaluation is undertaken by the instructing teacher and two teacher observers.

Three middle grades science teachers, teaching three different regular-level classes of 8<sup>th</sup> grade students at Randall Middle School in Lithia, FL, participated in this lesson study research over a three-week period in April-May, 2003.

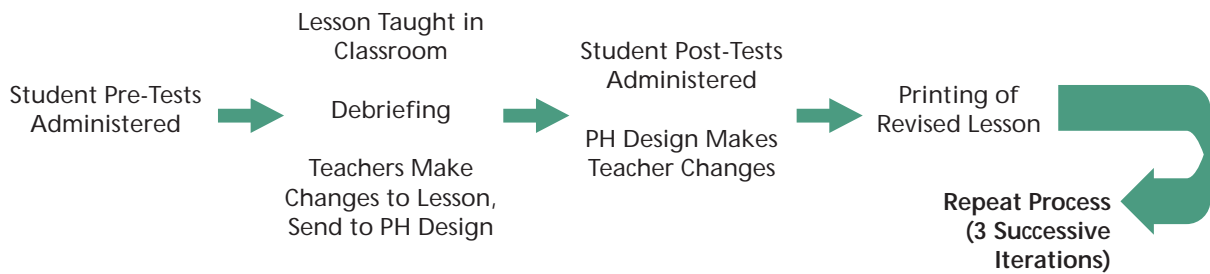
School Profile*	
FL State Grade 2003:	A
Grade Configuration:	6–8
Metro Status:	Suburban
Size:	1,315 students
Percentage of Capacity:	105%
Free and Reduced Lunch:	15.29%

\*Source: School District of Hillsborough County, [www.sdhc.k12.fl.us](http://www.sdhc.k12.fl.us)

The *Prentice Hall Science Explorer* prototype lesson on “Measuring Earthquakes” (Section 2 from the Earthquakes Chapter) was tested in the three classrooms. On each occasion, one teacher taught the lesson, while the other two teachers, plus one Prentice Hall market researcher and one Prentice Hall editor, observed. A pre- and post-assessment (created and scored by the teachers) was administered to students in each class to help determine if learning improvement had taken place. The pre-test was given the day before actual instruction with the “Measuring Earthquakes” lesson; the same post-test was given the day after instruction took place. Each test took approximately 20 minutes to administer.

After each lesson iteration, the teachers participated in a 45-minute debriefing to discuss what worked and what did not work in the lesson, and what improvements they felt should be made. Afterwards, the teachers met as a team to revise the lesson for the following week. Based on the teachers’ recommendations, a design team revised the lesson and reprinted the new version to be distributed to students for the following week’s class. A 60-minute debriefing was held after the final lesson was taught to discuss in detail the rationale behind the changes they made, as well as to get feedback from the teachers on the study process.

A flowchart of the study process appears below:



## Key Findings

- Student learning increased in all three lesson iterations. All three classes showed an average 10-point improvement or better from the pre- to the post-assessment. *(Note: only students who took both the pre-test and the post-test were evaluated. Scoring was completed by the instructing teachers).*

	Lesson Iteration 1	Lesson Iteration 2*	Lesson Iteration 3
<b>Class Mean Point Gain:</b>	14	10	18

\*Note: Lesson Iteration 2 lasted only 55 minutes (due to an impromptu pep rally organized by the school), as opposed to the regular block period of 70 minutes for Iterations 1 and 3. This may have contributed to a slightly lower point gain for this iteration.

- In addition, the teachers observed other indicators of student achievement with the *Prentice Hall Science Explorer* lesson as revisions were made:
  - The ease with which students answered correctly the questions posed during class by the instructor improved with each lesson.
  - Student success with completing labs and activities increased with successive iterations (e.g., getting the “right answer” more often and more quickly).
  - Student concentration and attention to tasks (reading, following activities/labs) was at higher levels than previously observed.
- The three instructing teachers noted the following strengths of the *Prentice Hall Science Explorer* ©2005 prototype:
  - The content is clearly stated and tells students what just what they need to know, without a lot of extras.
  - It’s easy for regular-level kids to use and understand.
  - The illustrations are outstanding.
  - Has an effective lab (despite a few problems with making the data “work”).
  - The Reading Checkpoints emphasize the main points and cue the teacher to remember to make sure that kids are understanding the material—a good indicator if the teacher can go on with the material or not.
  - The Reading Preview is helpful for a new teacher, or one without a background in science. It provides a way to make sure students are getting the key concepts.

- Following is a summary of the suggestions for improvements, changes, and additions the teachers made to the “Measuring Earthquakes” lesson during and after each iteration. (Note that the teachers felt no further changes were needed after Lesson 3).

	Additions/Changes Lesson 1	Additions/Changes Lesson 2	Additions/Changes Lesson 3
<b>Reading Preview</b>	<ul style="list-style-type: none"> <li>• Key Concepts modified to statement of objectives</li> <li>• Target Reading Skill modified to concept maps</li> </ul>	<ul style="list-style-type: none"> <li>• Key Concepts modified to statement of objectives</li> <li>• Target Reading Skill modified to concept maps</li> </ul>	<ul style="list-style-type: none"> <li>• Key Concepts modified to statement of objectives</li> <li>• Target Reading Skill modified to concept maps</li> </ul>
<b>Activities</b>	<ul style="list-style-type: none"> <li>• Added Foldable Activity</li> </ul>	<ul style="list-style-type: none"> <li>• Moved Discover Activity to bottom of page 61, after introductory paragraphs</li> <li>• Added Group Work Activities</li> </ul>	<ul style="list-style-type: none"> <li>• Moved Discover Activity to bottom of page 61, after introductory paragraphs</li> <li>• Added Foldable Activity</li> </ul>
<b>Reading Checkpoints</b>		<ul style="list-style-type: none"> <li>• Reading Checkpoints modified to concept maps</li> </ul>	<ul style="list-style-type: none"> <li>• Reading Checkpoints modified to concept maps</li> </ul>
<b>Text/Content</b>		<ul style="list-style-type: none"> <li>• Added Figure references into text</li> </ul>	<ul style="list-style-type: none"> <li>• Added Figure references into text</li> <li>• Added further detailed information about the Mercalli Scale</li> </ul>
<b>Illustrations</b>		<ul style="list-style-type: none"> <li>• Added Illustration of Surface Waves</li> </ul>	<ul style="list-style-type: none"> <li>• Added Illustration of Surface Waves</li> </ul>
<b>Lab</b>	<ul style="list-style-type: none"> <li>• Added Map without cities identified</li> </ul>	<ul style="list-style-type: none"> <li>• Added Graph, with new Step 2 (interpretation)</li> <li>• Revised Data Table with new data points (substituted Chicago for Miami)</li> <li>• Added Map with scale (1km=200cm), but without cities identified</li> </ul>	<ul style="list-style-type: none"> <li>• Added Graph, with new Step 2 (interpretation)</li> <li>• Revised Data Table again, with new data for Denver</li> <li>• Added Map with scale (1km=200cm), and with cities identified</li> </ul>
<b>Other</b>	<ul style="list-style-type: none"> <li>• Added Bellwork at beginning of class</li> </ul>	<ul style="list-style-type: none"> <li>• Added Bellwork at beginning of class</li> <li>• Added Analogies</li> </ul>	<ul style="list-style-type: none"> <li>• Added Bellwork at beginning of class</li> <li>• Added Analogies</li> </ul>

# Beyond the Data

## Practitioners Tell Us Why it Works

### ***What the teachers in the study had to say about Prentice Hall Science Explorer . . .***

“Students using the *Science Explorer* text showed only one-third of the failures of the non-user population. In fact, not only did the low ability students do better, so did the overall range of students.”

*Phil S., 8<sup>th</sup> Grade Science Teacher, Wisconsin*

“The [*Science Explorer*] program has good visuals (diagrams, charts, etc.) and support activities. The clear, concise visuals add clarity to the text. I like the enrichment reading activities to support literacy with good questions to check comprehension. The summary pages are very useful in every case to support classroom instruction. The students had a resource at hand outside of the classroom to complete activities.”

*Louise B., 8<sup>th</sup> Grade Science Teacher, Colorado*

“I especially like the Teaching Resource Book because it was chock filled with great ideas and had great materials for the overhead projector. It has review materials that are great to duplicate and use as homework. Also, the performance assessments are wonderful!”

*Joanne F., 8<sup>th</sup> Grade Science Teacher, New Jersey*

“The projects are great, the content as well as the processing skills. In addition, I used the Inquiry Skills Workbook to prepare for CSAP.”

*Debbie B., 8<sup>th</sup> Grade Science Teacher, Colorado*

“The [Human Biology] book was well organized, covered essential information, and presented it in an easy-to-use manner. Students were excited about the projects and were forced to problem solve and actively think.”

*Scott C., 8<sup>th</sup> Grade Science Teacher, Washington*

“[There is] a nice use of technology, the books are colorful and concise, there are many choices of activities, and the tests offer a nice variety of questions.”

*Dave S., 8<sup>th</sup> Grade Science Teacher, Washington*

## ***What the students in the study had to say about Prentice Hall Science Explorer . . .***

“The *Science Explorer* textbook helped me understand more about science, and made it easier for me. It has lots of information in it and prepared me for the tests because of the review at the end of every chapter.”

*8<sup>th</sup> Grade Student, Wisconsin*

“I have really enjoyed working out of this book. I feel like I learned more. The word descriptions are pretty clear and pretty easy to understand. The pictures in the book are very colorful and fun to look at. I love how very descriptive the pictures are. It’s fun to read.”

*8<sup>th</sup> Grade Student, Washington*

“I thought the Science Explorer textbooks were very informative. The books had a lot of ‘Try This’ examples that worked well and were fun when we did them. The books were easy to understand.”

*8<sup>th</sup> Grade Student, Colorado*

“The Science Explorer textbooks we used this year were great. When we used them I understood everything clearly. The book was helpful in many ways.”

*8<sup>th</sup> Grade Student, Colorado*

“These books are good for reviewing for the G.E.P.A. They have good review questions and the text is easy to read. The pictures were very self-explanatory, making it easier to grasp the concept of science.”

*8<sup>th</sup> Grade Student, New Jersey*

“The integrated science sections were very informative. The section reviews kept the information fresh in my mind.”

*8<sup>th</sup> Grade Student, Colorado*

“I liked the checkpoints. I think that they help you understand what you read, and they have plenty of pictures to help to comprehend.”

*8<sup>th</sup> Grade Student, Colorado*

# NCLB Implications for Research

As we continue to track the progress of Prentice Hall users' performance, our research design for the longitudinal, effect-size studies follows closely the criteria of the No Child Left Behind legislation. Below is a brief description of the NCLB implications for research that inform our work:

## Scientifically Based Research

One of the most prominent features of No Child Left Behind is the call for scientifically based research to determine the extent to which a program or approach to learning is effective in the classroom. The term scientifically based research, or SBR, is mentioned in the NCLB Act no fewer than 110 times. It is critical, therefore, to understand just what SBR means and what it implies in practice.

### ***The NCLB Act of 2001 defines scientifically based research as:***

Research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to educational activities and programs . . .

- It involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn.
- It relies on measurement or observational methods that provide reliable and valid data.
- It is evaluated using experimental or quasi-experimental designs in which programs or activities are assigned to different conditions, with appropriate controls to evaluate the effects of varying the condition of interest.
- It ensures that experimental studies are presented in sufficient detail and clarity to allow for replication, or, at a minimum, to offer the opportunity of building systematically on their findings.
- It has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably objective and scientific review. (*No Child Left Behind 2001*)

## Elements of Scientific Research

When a scientifically based study is conducted, it can be repeated, or replicated, with the same research design, and the outcomes will be nearly identical. Random assignment of treatment and control, or non-treatment, groups is the best way to achieve a similarity of outcomes over a series of replications.

As you review research and make decisions about its quality or rigor, think about some of the questions posed below when conducting a serious review of a research document. A comprehensive document, “Design and Implementation Assessment Device” (DIAD), is provided at the What Works Clearinghouse Web site and elaborates on the criteria to examine when reviewing a research document.

- Are the research questions clearly defined?
- Is the research design described comprehensively?
- Is the sample size large enough to convince you that the results are generally applicable?
- Does the sample group reflect the characteristics of your students?
- Is there a control group that was randomly chosen from the subject population?

Since NCLB will hold schools accountable for improving the performance of all students, it is critical that the instructional program that is selected in your district be based on sound evidence of effectiveness. In addition, knowing that a program may be generally beneficial to learning will not be enough. No Child Left Behind will require schools and districts to examine their student achievement test score data with respect to various groups of demographic factors. Studying data in this way is called disaggregating the data.

## What Works Clearinghouse (WWC)

This project of the U.S. Department of Education’s Institute of Education Sciences was established to provide educators, policy makers, and the public with a trusted source of scientific evidence on what works in education. The purpose of the WWC is to identify interventions or approaches in education that have a beneficial causal relationship with important aspects of student performance.

The WWC is proposing a set of standards of evidence that will help judge the quality of studies claiming to prove the effectiveness of an approach. The WWC will also produce summaries of research called evidence reports. These reports will provide a synthesis of research findings on specific educational interventions (for example, the effectiveness of certain school reform models) and approaches (such as class size reduction or bilingual education).

# Evaluator's Checklist

## Criteria for Evaluating Research in Support of an Educational Program

The field of education contains a vast array of educational interventions that claim to be able to improve educational outcomes and, in many cases, to be supported by evidence. However, education practitioners are faced with the challenge of deciding if the evidence is credible and whether an intervention is truly effective. The National Center for Educational Evaluation and Regional Assistance (NCEE\_ has issued a new guide on evidence-based education, titled “Identifying and Implementing Educational Practices Supported by Rigorous Evidence.” The guide offers tools to educational practitioners that help distinguish practices supported by rigorous evidence from those that are not.<sup>4</sup>

The following checklist provides a set of standards of evidence that will help you judge the quality of the studies you are reviewing and deem whether the data is valid and reliable. You may find it helpful to use the descriptors provided as a guide to evaluating the various criteria of educational research. For more information or for a copy of the complete publication visit, [www.ed.gov/about/offices/list/ies/news.html#guide](http://www.ed.gov/about/offices/list/ies/news.html#guide).


---

<sup>4</sup>Institute of Educational Sciences, New Guide on Evidence Based Education

## Evaluating Research Supporting an Educational Intervention

QUALITY OF STUDIES		
QUESTION	DESCRIPTION AND RATIONALE	RATINGS
1. Is the study described clearly?	The study should clearly describe the intervention, including who administered it, who received it, and at what cost; how it differed from what the control group received; and the logic of how the intervention is supposed to affect outcomes.	1 2 3 4
2. Are there closely matched control and treatment groups used as a point of comparison?	Control and treatment groups must be <i>very closely matched</i> by prior test scores and other measures of academic achievement; demographic characteristics such as age, ethnicity, and poverty level; time period in which two groups are studied; and methods used to collect the outcomes.	1 2 3 4
3. Is data provided that shows there were no systematic differences between the matched groups?	The study should show data that, prior to introduction of the program, there were no systematic differences in the vast majority of measured characteristics.	1 2 3 4
4. Are objective, real-world measures used to report outcomes?	The study should use outcome measures that are “valid” such as tests that accurately assess true skill levels (e.g., Stanford Achievement Test)	1 2 3 4
5. Are the long-term effects of the intervention reported?	The study should obtain data on long-term outcomes so that you can judge whether the program’s effects are sustainable over time.	1 2 3 4
6. Are the findings described in easily understandable, real-world terms?	The findings should be reported free from any technical jargon that may cause the reader to misinterpret the data.	1 2 3 4
QUANTITY OF STUDIES		
QUESTION	DESCRIPTION AND RATIONALE	RATINGS
7. Is the size of the effect documented?	To quantify outcomes, the effect-size should be reported as well as if the effects are statistically significant so that the reader can judge their educational importance. Effect size can be indicated in a number of ways, e.g., point gains, percentage increases, grade level increases, etc.	1 2 3 4
8. Is the sample size generalizable/representational?	In order to obtain statistically significant results, the study must have a relatively large sample size. 300 students (150 treatment group, 150 control group) is a good rule of thumb.	1 2 3 4
9. Was the study administered in more than one site?	The outcomes of a study in a single site may be the result of site-specific factors (i.e., intense professional development or instructional intervention) and may not be generalizable to other sites.	1 2 3 4
10. Are the sites typical school/community settings?	To build a high degree of confidence that a program will be effective in a school/district, the study must demonstrate the program’s effectiveness in a typical setting, i.e., public school.	1 2 3 4
11. Are all of the effects reported, not just the positive ones?	The study should report all of the program’s effects on all measured outcomes so that the reader can judge whether the positive effects are the exception or the pattern.	1 2 3 4

## Evaluating Research Supporting an Educational Intervention (continued)

<b>STUDY FEATURES TO APPROACH WITH CAUTION</b> 		
<b>FEATURE</b>	<b>RATIONALE</b>	<b>RATINGS</b>
1. Pre-post comparisons with groups that are not well matched (e.g., comparing a low-income control district to a high-income treatment group).	Groups that are not well matched can lead to inaccuracies large enough to indicate incorrect judgments about a program's effectiveness.	1 2 3 4
2. Studies that do not use random assignments.	Random assignment ensures that any naturally occurring differences in the treatment and control groups are accounted for in the data collection process.	1 2 3 4
3. Self-reported outcomes.	If outcomes are not properly monitored, under or over-reported, the study results will be inaccurate.	1 2 3 4
4. A study's claim that a program's effects on a sub-group (i.e., Hispanic students) are different than its overall effect.	Often these indiscrepancies are a result of eliminating important data to achieve a desirable outcome or just happen by chance.	1 2 3 4
5. Studies that are conducted by the publisher/owner of the educational materials and not supported by an independent source.	Independently conducted research is the most reliable. "In-house" study results may be biased or may not include all of the relevant information in an attempt to hide any negative results.	1 2 3 4
6. Studies that report only positive effects.	It is a rare study that only yields positive results. Typically, portions of the study have been left out of the report that will bias the results of the study and give an unrepresentative picture of the true findings.	1 2 3 4
7. Research studies that include only qualitative data analysis.	Qualitative data analyses tend to be more subjective in nature. Qualitative data is a good method to explore the "whole picture," but should not be taken as a declarative statement as to whether a product is effective. Only a randomized, controlled trial can indicate true efficacy.	1 2 3 4

<sup>1</sup>Institute of Educational Sciences, New Guide on Evidence Based Education

# Longitudinal Study Plans in Progress

As part of our ongoing commitment to provide the most valid and reliable research to support the efficacy of our programs, Prentice Hall has contracted with Claremont Graduate University, *Institute of Organizational and Program Evaluation Research*, to conduct a rigorous, independent, quantitative study on the effects of *Prentice Hall Science Explorer* beginning in the 2004-2005 school year.

This longitudinal study will consist of a randomized control trial that has been designed to fully meet the quality criteria put forth by the What Works Clearinghouse in the Study DIAD (v. 1.0)<sup>1</sup>.

Both qualitative and quantitative instruments will be developed and/or adapted for the study. The study will examine teacher's implementation of the science curriculum as well as a diverse set of student outcomes that together give a comprehensive picture of how the Prentice Hall Science curriculum affects students' achievement in science during middle school and high school.

The primary questions motivating this research include the following:

1. What is the relation between program implementation and student achievement in science?
2. What is the relation among students' higher-order thinking, student attitudes towards science, and achievement in science?
3. How do student outcomes differ for Prentice Hall Science programs compared to other similar science programs ?
4. How do students with different characteristics (e.g., English Learners, various ethnic groups and socioeconomic status) participating in the Prentice Hall Science program perform on student-related outcomes?

Results will be published as they become available. If you wish to obtain more information about the longitudinal study or any of the research that supports *Prentice Hall Science Explorer*, visit the research-dedicated Web site at [PHSchool.com/ScienceResearch](http://PHSchool.com/ScienceResearch).

---

<sup>1</sup>Valentine, J.C. & Cooper H. (2003). What Works Clearinghouse Study Design and Implementation Assessment Device (Version 0.6). Washington, DC: US Department of Education.



Visit [PHSchool.com/ScienceResearch](http://PHSchool.com/ScienceResearch)  
to obtain the complete research  
supporting *Prentice Hall Science Explorer*.



Customer Service: 1-800-848-9500 [PHSchool.com](http://PHSchool.com)

0-13-166304-6 Copyright Pearson Prentice Hall, Inc. All Rights Reserved. Printed in the United States of America. 0303311