Journal of Science Education and Technology, Vol. 14, Nos. 5/6, December 2005 (© 2005)
DOI: 10.1007/s10956-005-0222-2

The Effects of a Kit-Based Science Curriculum and Intensive Science Professional Development on Elementary Student Science Achievement

Betty J. Young[1,3] and Sharon K. Lee[2]

The science achievement of 226 5th graders from districts that have a kit-based inquiry science curriculum supported by intensive professional development (PD) is compared with data from a group of 173 5th graders from other districts that use nonkit science materials and do not have systematic science PD for teachers. Within the kit-based project, the sample of project teachers is stratified to select teachers with a high number of science PD hours versus those with few hours. While there were no significant differences in the mean total scores for kit-based students with low PD versus high PD teachers, the kit-based classrooms scored significantly higher than students in nonkit classrooms on both the pretest and posttest, though there were significantly more minutes of science instruction in the nonkit classrooms. Finally, nonkit teachers taught more units of shorter length and reported lower levels of preparedness to use reform pedagogical approaches.

KEY WORDS: elementary science achievement; kit-based science; professional development.

INTRODUCTION

The call for improved science literacy in the early 1990s and the science standards that followed inspired widespread activity to develop better materials and prepare teachers to teach science using methods guided by advances in cognitive science. Curriculum development occurred as groups such as Lawrence Hall of Science (i.e., *Full Option Science Systems*—FOSS) and the National Science Resources Center (i.e., *Science and Technology for Children*—STC) developed new elementary science kits. Though kit-based science has a long history, earlier generations of kit curricula such as the *Elementary Science Study* (ESS), *Science—A Process Approach* (SAPAS), and *Science Curriculum Improvement Study* (SCSS) were difficult to sustain once initial federal support ended (Shymansky, 1989). Later meta-analyses of the programs did reveal that many of the programs successfully improved student achievement and in some cases had a positive impact on student attitudes (Bredderman, 1985; Shymansky *et al.*, 1982a,b; Stohr-Hunt, 1996). Further studies uncovered a relationship between the amount of teacher training and the level of student achievement (Shymansky *et al.*, 1990).

More recent research results support the development and implementation of laboratory inquiry, critical thinking, and decreased teacher-centered instruction, major objectives of most current kit-based programs (von Secker and Lissitz, 1999). The latest generation of kits has been extensively evaluated for teacher professional development and implementation of kit curricula (Lawrenz *et al.*, 2001). However, information regarding student outcomes remains limited. This study provides valuable information about the impact of a kit-based program, supported by intensive teacher development, on student achievement.

[1]School of Education, Chafee 616, University of Rhode Island, Kingston, Rhode Island.
[2]Rhode Island Department of Education, Providence, Rhode Island.
[3]To whom correspondence should be addressed; e-mail: byoung@uri.edu

The purpose of this study is to examine the science achievement of 5th graders participating in a kit-based science curriculum that is supported by intensive science professional development (PD). The study reports the science achievement of 226 5th graders from seven districts that participated in a 5-year professional development initiative supporting the implementation of a kit-based exploratory science curriculum for kindergarten to grade eight. The seven New England districts were in the fifth year of a Local Systemic Change (LSC) initiative funded by the National Science Foundation. The data are compared with the science achievement of a group of 173 5th graders from three other districts that are demographically matched, have used other types of science materials, and do not have a comprehensive science professional development program for teachers. Additionally, the achievement data within the kit-based project are analyzed to examine any differences between student performance in classrooms with teachers who have a high number of professional development hours versus students taught by teachers with relatively few hours (e.g., teachers new to the district). Thus, this study examines the effects of a particular treatment, at two different intensities, with a comparison group that has not used the same materials or provided the same science professional development for teachers.

Research Questions

This study is guided by five main questions about student achievement related to curriculum materials, teachers' sense of preparation to teach science, and different pedagogical approaches to science instruction. The research questions are:

- Are there differences in science achievement for 5th graders who participate in a kit-based curriculum versus students in nonkit based instruction?
- Are there differences in science achievement for 5th graders who participate in a kit-based curriculum with teachers who have a high number of science professional development hours versus teachers who have a low number of science professional development hours?
- What are the differences in teacher characteristics between kit and nonkit teachers?
- What are the differences in instructional approaches between kit and nonkit teachers and between teachers with high and low levels of professional development within the kit-based classrooms?
- What teacher and instructional characteristics are associated with higher levels of science achievement for the total sample and for kit and nonkit groups?

Perspectives

The outcomes related to various instructional approaches and materials in science have been a topic of educational research since the flurry of science reform activities in the post-Sputnik1960s. In the newest generation of science materials and implementation approaches, changes in the materials themselves as well as in the support for teachers who use the materials have become the focus of contemporary studies.

Changes in the science instructional materials have emanated from advances in our understanding of the ways children learn. A fundamental notion in science education is that children construct knowledge based on personal and social interactions in relation to physical events in their lives (Driver *et al.*, 1994; Piaget and Garcia, 1989). Much of the literature which has guided science instruction in the past, particularly related to what is developmentally appropriate for elementary children, has been revisited in recent years with the conclusion that children are far more capable of scientific inquiry than previously thought (Metz, 1995).

The greater the opportunities to explore questions, the more competent children become in making data-based inferences (Schauble *et al.*, 1991), improving the accuracy of evidence–inference connections, and using better inference strategies (Kuhn *et al.*, 1992). The enriched experiences help students develop conceptual "scaffolds" or frameworks on which they arrange new knowledge.

For example, the STC and FOSS kits provide materials and teacher guidance to allow children to construct science knowledge and develop science process skills by working on particular science topics in depth (e.g., for periods of 6–8 weeks). The lessons are designed to provide exploration of materials and hands-on investigations that lead learners to key science concepts and help them develop their skills and a scientific disposition in conducting investigations. Zacharia (2003) identifies the need for good quality tools to promote learning through exploration. Njoo and de Jong (1993) found that exploratory experiences help children develop deeper understanding of concepts by allowing children to incorporate new learning into a more elaborate cognitive structure.

A key factor in the success of any educational reform is the knowledge and skill of the classroom teacher who is called upon to implement a new curriculum. Fensham (1992) and others have suggested that earlier educational innovations in science have failed, in part, because the intentions of the curriculum developers were not reflected in teachers' actions at the classroom level. The importance of teacher professional development is seen in Bredderman's 1984 study that showed the level of questioning and student activity increases while the amount of "teacher talk" decreases in teachers trained in the use of kit-based science instruction. Teachers must be provided with rich environments in which they can increase their expectations of what children are capable of knowing and doing in science. Further, given the higher expectations, there is a need to enrich teachers' knowledge of science content and processes, as well as to learn new pedagogy related to inquiry-based science instruction. Inadequate teacher preparation and support have been cited as contributing factors in the failure of the earlier generation of science kits to lead to significant advances in student achievement (Spickler and McCreary, 1999; Walberg, 1991). The traditional model of providing teachers with better materials accompanied by hasty staff development sessions is inadequate to the task of fostering the fundamental changes required.

There is an increasing volume of evidence associating hands-on, inquiry-based science instruction with an increase in student achievement over students participating in more traditional programs (Escalada and Zollman, 1997; Freedman, 1997; Morrell and Lederman, 1998; Okebukola, 1987; Parker, 2000; Shymansky, 1989; Shymansky *et al.*, 1982a,b, 1990). Most recently, Schneider *et al.* (2002) found that students in project-based science outperformed traditional students on the National Assessment of Educational Progress (NAEP).

In spite of the studies that show the advantages of kit-based inquiry science, there is a need for more detail about the science achievement patterns and teacher characteristics associated with achievement gains. Many factors influence the science achievement of elementary children. Zacharia (2003) identifies the format of the learning experience in combination with the education of teachers in using more exploratory methods as important ways to increase student science achievement. Simpson and Oliver (1990) stress the importance of a focus on factors that can be manipulated, namely instructional materials and teaching practice. The current study examines achievement in relation to these two malleable factors.

METHODOLOGY

Design

This investigation gathers evidence to reveal the relationship between the independent variables (i.e., level of kit use in the science curriculum and amount of teacher training) and two dependent variables (i.e., pretest and posttest science achievement scores). This study is a posttest comparison design that involves random selection from two clusters (teachers with few PD hours and teachers with high PD hours) within the LSC district classrooms and an opportunity sample (Champion, 2002) of teachers from nonkit/no science PD districts that are demographically matched with the LSC students. Students in selected classrooms took the science pretest in October 2002 and were posttested in June 2003. Teacher questionnaires were administered each time to collect information about preparedness, experience, and practices that could be associated with student science achievement.

Sample

The LSC project classes included six classes ($N = 116$) with teachers who had the most professional development hours within the project and six classrooms ($N = 110$) that were taught by teachers with relatively few hours of LSC professional development. The teachers from the LSC project were selected from a randomized list provided by Horizon Research, Inc. Based on the requirements from the Institutional Review Board at the researchers' institution, teachers were given the right to refuse to participate. Two teachers on the low PD hour list declined to participate and were replaced by the next names on the list. Each of the first six teachers on the high PD list agreed to participate.

The comparison group was composed of nine nonkit classes ($N = 173$) that did not use kit-based science and were taught by teachers who had not experienced ongoing, intensive science professional development. This opportunity sample (Champion, 2002) was obtained by contacting the curriculum directors in three districts that had not participated in the LSC program and did not use kit-based science or have systematic professional development. While one district provided a randomized list of 5th-grade teachers that could be contacted, the remaining districts solicited volunteers that allowed their classes to be tested and agreed to complete the teacher questionnaires. Given the lack of random selection of the comparison group, the generalizability of the results may be questionable. However, it can be reasoned that teachers volunteering from these districts would be the most confident in their science teaching. Thus, if the kit-based classrooms were to outperform children in the nonkit classrooms, it would be a comparison between a representative sample of the LSC students and teachers versus, arguably, the best of instructional settings from the other districts.

The total sample of 5th graders was 92% Caucasian (not Hispanic) with about 2% in each of the remaining categories (Black, Hispanic, Asian, American Indian). The sample was fairly evenly divided between males (54%) and female (46%) students. There were no differences in the numbers of children with special education or limited English proficiency status or who received assistance on either the pre- or the postassessment; however, there was a somewhat higher number of students on free/reduced lunch in the nonkit group (nonkit = 25; kit = 16), which was statistically significant.

Nineteen of the 21 teachers were female, and all teachers in the study were Caucasian. Ten of the 21 teachers had between 6 and 10 years of teaching experience, and five of sampled teachers had 5 years or less in the teaching profession. There was no significant difference between the experience of the kit and nonkit teachers. The average class size for the sample was 21 students (SD = 4) with no significant difference between the kit and nonkit class sizes.

Setting

All students in the LSC districts experience a core curriculum of 19–21 science kits sequenced at two to three kits a year from kindergarten to grade six. The science kits used in the core curriculum include topics on life, earth, and physical science developed by Science and Technology for children (STC) and Full-Option Science Systems (FOSS). This instruction is supported by professional development that relates directly to the content and processes in each science kit as well as training on increasing the critical thinking through inquiry pedagogy. Fifth graders in the LSC districts had experienced 12–14 kits taught by teachers who have received specific training in the use of the science kits as well as exploratory methods, inquiry pedagogy, and science content knowledge. Thus, the LSC students enter 5th grade with a wealth of science background. In the 5th-grade curriculum, students studied *Floating and Sinking* (STC), *Microworlds* (STC), and *Levers and Pulleys* (FOSS). Appendix A contains the curriculum matrix that indicates the total science curriculum for the grades leading up to grade 5. For each of the three kits used at a grade level, LSC teachers are required to participate in a 6-h basic training. After teachers have used the kit for 1 or 2 years, they participate in another 6-h follow-up session, cotaught by teacher leaders and university scientist/mentors who emphasize science content related to a kit and develop skills in opening up lessons to include more inquiry and exploration during kit lessons. Additionally, many LSC project teachers have taken extra courses and summer institutes on various topics related to the science curriculum.

The students in non-LSC classrooms received science instruction in a variety of ways using texts and teacher-developed units. Teachers in the sample reported using *Holt Science Series*, *Discovery Works* (Houghton Mifflin Science), *Bill Nye Videos*, *Tom Snyder Videos*, as well as teacher-developed units supported by various Internet sources. Generally, the teachers in the nonkit classrooms covered many more topics in shorter duration than the kit-based settings. Professional development in science for the non-LSC teachers was dependent on teachers choosing from outside courses or workshops as their districts did not provide specific science training in the delivery of the science curriculum.

DATA SOURCE

Instrumentation

Data for this study is derived from three instruments, an elementary science achievement test and two teacher questionnaires, all developed as program evaluation instruments by Horizon Research, Inc. (HRI) with funding from the National Science Foundation. These instruments were used for all LSC initiatives funded through this Teacher Enhancement program at NSF. The science achievement test was composed of released items from NAEP and *Third International Mathematics and Science Study*

Table I. Subscales Characteristics for HRI Science Achievement Test

Scale	Number of items	p-value range	Reliability
Life science	10	0.39–0.79	0.67
Earth science	12	0.78–0.97	0.65
Physical science	10	0.39–0.81	0.66
Electricity and magnetism	10	0.55–0.90	0.63
Nature of science	9	0.41–0.80	0.66

(TIMSS) tests that were "above the recall level." LSC project directors and project evaluators reviewed these items for their match to the objectives of the *National Science Education Standards* related to content, process, and pedagogy. Additionally, HRI developed items that were underrepresented by the selected NAEP and TIMSS tests. In 2001, HRI piloted 78 items with approximately 3,000 students in 12 LSC projects across the country. Overall p-values and differential item functioning for item performance for limited English proficient (LEP) versus English fluent students and for white students versus students of color were examined to allow removal of biased items. Analysis of the piloted items resulted in a 52-item assessment with five subscales: life science, earth/space science, physical science, electricity/magnetism, and nature of science (Table I).

Two additional items not associated with these subscales were added at the beginning of the test to build confidence. In the spring of 2002, an additional 12 items along with the previous 52 were piloted yielding a final assessment with 53 items (Overstreet, 2002). The achievement test did not directly relate to the particular units of the LSC districts in this study. The test was designed with a range of difficulty and topics that could be used in grades 4–6. LSC projects across the country select from a menu of "NSF-approved" materials, but may use different kits, thus different science topics, at different grade levels. Thus, the test was not designed to assess, in particular, the concepts that either group studied in the 5th grade.

The teacher questionnaires were also developed by HRI for use of LSCs around the country. The teacher questionnaires collected demographic data (gender, race/ethnicity, science coursework, years of teaching experience), information regarding the setting in which they teach science (grade level, class size, duration of typical science lesson, number and types of units taught, and instructional approaches), amount and type of professional development in science, and opinions on science teaching and their perception of preparedness for teaching science. Because the questionnaires were designed for LSCs, the instruments used with the non-LSC teachers were modified slightly. For example, "Approximately how many hours have you spent on LSC professional development in science or science education?" was changed to "Approximately how many hours have you spent on formal professional development in science or science education?" Thus, the questions yielded information related to amount of science professional development without limiting it to sessions provided by the LSC project.

Data Collection

The science test and teacher questionnaires were administered by the researchers in the fall of 2002 and the spring of 2003. In advance of the administration students took home letters that gave their parents the right to decline to have their child tested. Researchers provided alternate science-related quiet activities for any students who were not participating. In all, there were very few students who did not have permission to take the science test. The same instructions, provided by HRI, were read to all students who were then given 50 min to complete the science achievement test. Any students who finished the test early were encouraged to look over their work and then given a quiet science activity. During the administration, the classroom teacher completed the teacher questionnaire. The classroom teachers were not allowed to view the contents of the science achievement test.

RESULTS

Science Achievement

The first question explores differences in science achievement for 5th graders who participated in a kit-based curriculum versus students in nonkit based instruction. The pretest and posttest results have been compared, using independent sample t-tests, to

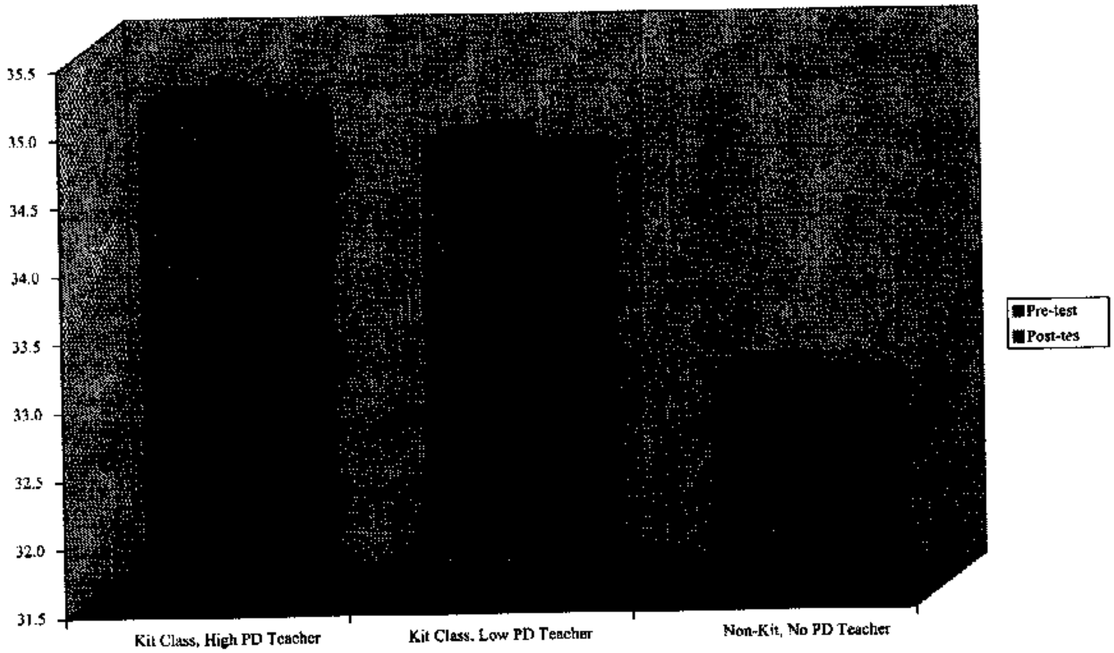

Fig. 1. Pretest and posttest average science achievement scores by group.

determine any significant differences in raw scores and scale scores between kit and nonkit 5th graders as well as comparisons between high PD/low PD kit-based classrooms.

In terms of science achievement levels for kit and nonkit 5th graders, there was a significant difference in the total pretest performance of students ($t = 2.341, p = 0.020$), with the students from the kit-based curriculum outscoring the comparison group. Among the subscales of the test, there was a significant difference in two of the five scales: earth science ($t = 2.051, p = 0.041$) and nature of science ($t = 2.287$, $p = 0.023$). In the remaining scales (i.e., life science, physical science, and electricity/magnetism), the kit-based scores were higher, but the difference between groups was not significant.

The same pattern was seen in the posttest comparisons with significant differences in total posttest performance ($t = 2.261$, $p = 0.024$), earth science ($t = 1.995$, $p = 0.047$), and nature of science ($t = 2.695$, $p = 0.007$). On the remaining scales, once again, the kit-based students' average scores were higher, but the differences were not significant.

Though the students in kit-based classrooms maintained higher scores relative to the nonkit students, neither group made a significant gain in average number of items correct between the pretest and the posttest. Figure 1 shows the pattern among classroom groupings.

A second question examines differences in science achievement for 5th graders who participate in a kit-based curriculum with teachers who have a high number of science professional development hours versus teachers who have a low number of science professional development hours. The pretest results show no significant difference between the performance of the students in the low PD treatment kit-based classrooms and the high PD treatment kit-based classrooms. Thus, based on the history of the 5th graders in district classrooms, there is no initial difference in science achievement scores as students would have had a mostly common set of kit experiences and a mix of high/low PD teachers in the previous grades. The posttest results show no significant difference between these groups. It is possible that the efforts of previous teachers as well as the entry training that the low treatment teachers receive wash out any effects of having a 5th-grade teacher who has not participated in as extensive science PD as some peers. Students in the LSC classrooms have a history of science experiences

Table II. Variables in Index of Preparedness to Use Reform-Based Practices (refindex)

- Prepared to provide concrete experiences before abstract
- Prepared to develop students' conceptual understanding
- Prepared to take students' prior knowledge into account
- Prepared to make connections between science and other disciplines
- Prepared to have students work cooperative groups
- Prepared to have students do hands-on
- Prepared to engage students in inquiry-oriented activities
- Prepared to apply science in variety of contexts
- Prepared to use portfolios
- Prepared to use informal questioning to assess.
- Prepared to lead class using investigation strategies
- Prepared to manage a class doing hands-on/project-based work
- Prepared to help students take responsibility for own learning
- Prepared to recognize/respond to student diversity
- Prepared to encourage student interest n science
- Prepared to use strategies that encourage participation of females and minorities in science
- Prepared to involve parents in science education

with teachers who have had high as well as those who have more limited PD in using kits and reform pedagogy. This study was conducted in the fifth year of the LSC effort, so there would be contributions to student achievement from previous years.

Teacher Characteristics

What are the differences in teacher characteristics between kit and nonkit teachers? One of the key areas of interest is the difference in background characteristics between kit/nonkit teachers. Of interest are the differences in years of teaching experience, number of college science courses, hours of PD in science, and sense of preparedness to teach science or to use reform pedagogy including exploratory and inquiry methods. There were no significant differences in the average years of teaching experience, the number of college courses in science, or the sense of preparedness to teach various areas of science content. Nearly half of the teachers had taught in the range of 6–10 years. Nine teachers had only one or two college science courses, but seven had five or more courses in science. The distribution of responses to the question of preparedness in various areas of science (i.e., ecology, human body, rocks, astronomy, change over time, electricity, mixtures, sound, force/motion, machines, and design) was roughly equal between groups. Here the average for the kit teachers was 21.3 compared with 19.4 for the nonkit group out of a possible score of 44 for teachers who felt very well prepared.

There were three variables on which these groups differed significantly: professional development hours (both at the beginning and the end of the year) and their sense of preparedness to use reform pedagogy approaches. The kit-based teachers reported, on average, between 40 and 59 science-related PD hours compared to the nonkit teachers who reported, on average between 10 and 19 hours ($t = 0.627, p = 0.017$).

Table II shows the variables from the teacher questionnaires that were combined to create the index of preparedness to use reform strategies. The possible range was a score of 77 for teachers reporting the highest sense of preparation to 17 for those with the least confidence. Teachers in the kit-based program showed a significantly higher level of confidence ($\text{mean}_{\text{kit}} = 59.0$) than the nonkit group ($\text{mean}_{\text{nonkit}} = 50.0$) ($t = 2.472, p = 0.023$).

Instructional Approaches

It is also important to examine the differences in instructional approaches between kit and nonkit teachers. How frequently do teachers use exploratory methods, traditional approaches, or assessment techniques? How many minutes of science instruction did these teachers engage in for each of these groups? Kit-based teachers reported a significantly greater frequency of utilizing exploratory activities ($\text{mean}_{\text{kit}} = 42.3$, $\text{mean}_{\text{nonkit}} = 40.33$, $t = 0.767$, $p = 0.050$). This approach included using open-ended questions, requiring students to supply

Table III. Means, Standard Deviation, and Intercorrelations for Science Achievement and Teacher and Instructional Predictor Variables

Variable	*M*	SD	1	2	3	4	5	6	7	8
Science achievement score	34.17	9.26	0.01	0.13*	0.05	0.20**	0.06	−0.01	0.11*	0.11*
Predictor variable										
1. PD hours	4.19	2.56		—	0.14**	0.27*	−0.26**	−0.10*	0.03	−0.05
2. Sense of preparedness in science content areas	26.10	8.27			—	0.37*	0.30**	−0.19**	0.26**	0.21**
3. Minutes of science instruction during the year	2689	2354				—	0.29**	−0.19**	0.16**	0.40**
4. Reform pedagogy index score	55.04	9.04					—	−0.41**	0.72**	0.28**
5. Number of college science courses,	2.95	1.59						—	0.08	−0.02
6. Years of teaching experience	3.39	1.38							—	−0.32**
7. Frequency of reform methods	44.18	6.44								0.42**
8. Frequency of traditional methods in teaching science	16.12	2.82								—

*$p < 0.05$. **$p < 0.01$.

evidence and explain concepts, having students design their own investigations, and having students do field work and record reflections in journals or science notebooks. However, there were no significant differences in how frequently these groups of teachers used more traditional approaches such as having students make formal presentations to the class, answer/review homework or worksheet assignments, or take short answer tests (e.g., multiple choice or true/false). Teachers who had the kit-based professional development used both sets of strategies, while teachers who had not participated in the science PD mainly used more traditional approaches. In reporting ways that teachers evaluated student progress in science there were no differences in the frequency of using new assessment ideas (e.g., requiring students to explain answers, write in notebooks, or work on portfolios).

The examination of instructional approaches of teachers with high levels of professional development compared with those with few hours of kit-based classrooms showed no systematic differences. The only significant difference in these groups was the number of science-related professional development hours, which was the selection criterion for the levels of treatment in the random sample of kit-project teachers. In all areas of instruction and sense of preparedness, these teachers reported similar approaches and background. This finding can be connected to the earlier report that there was no significant difference in student performance between these groups.

Finally, this study investigated predictor variables associated with higher levels of science achievement for the total sample and for kit and nonkit groups. A stepwise regression analysis was performed to determine what teacher and instructional characteristics predicted science achievement for the total sample as well as the kit/no-kit subgroups. In this analysis the number of correct posttest responses was examined as a function of PD hours, teacher sense of preparedness in science content areas, minutes of science instruction during the year, reform pedagogy score, number of college science courses, years of teaching experience, frequency of reform methods, and frequency of traditional methods in teaching science. The correlation matrix for these analyses appears in Table III.

For the total sample the optimal model indicated that the reform pedagogy score ($\beta = 0.20$), hours of PD ($\beta = 0.13$), and fewer minutes of science teaching ($\beta = -0.12$) explained 7% of the total variance (Table IV).

Table IV. Regression Analysis Summary for Teacher and Instructional Variables Predicting Science Achievement in Total Sample, Kit Sample, and Nonkit Sample

Variable	B	SE B	β
Total sample[a]			
Reform pedagogy score	0.20	0.05	0.20
Hours of PD	0.52	0.21	0.13*
Minutes of science instruction	−.001	0.00	−0.12*
Kit sample[b]			
Preparedness to teach science content	0.20	0.06	0.22
Hours of PD	0.98	0.26	0.26
Minutes of science instruction	−.003	0.00	−0.20
Nonkit sample[c]			
Frequency of traditional instruction	0.59	0.27	0.17*

[a]Variables excluded = frequency of exploratory activities, frequency of traditional activities, # of college science courses, teaching experience, sense of preparedness to teach science content, and frequency of assessment. $R^2 = 0.07$ ($N = 402$, $p < 0.01$).
[b]Variables excluded = frequency of exploratory activities, frequency of traditional activities, # of college science courses, teaching experience, reform pedagogy score, and frequency of assessment. $R^2 = 0.09$ ($N = 239, p < 0.01$).
[c]Variables excluded = frequency of exploratory activities, # of college science courses, teaching experience, reform pedagogy score, total PD hours, minutes of science instruction, frequency of assessment, and sense of preparedness to teach science. $R^2 = 0.03$ ($N = 162, p < 0.01$).
*$p < 0.05$.

For the kit sample the optimal model indicated that teacher sense of preparedness in science content ($\beta = 0.22$), PD hours ($\beta = 0.26$), and fewer minutes of science teaching ($\beta = -0.20$) explained 9.2% of the total variance (Table IV).

In the optimal model for the nonkit sample, only the frequency of traditional instruction ($\beta = 0.17$) entered the model, explaining only 3% of the total variance (Table IV).

Overall, the prediction models accounted for very limited amounts of the variability in student achievement. These findings demonstrate the multitude of factors associated with student outcomes. The study included a wealth of possible sources of variability, yet there remain many unidentified contributors, pointing to the complexity of classroom processes.

DISCUSSION

The results of this study add to the evidence that sustained science education programs that combine high-quality materials and intensive teacher professional development in science and reform pedagogy have a positive impact on children's learning of science. There are several key findings that are surprising. Teachers in nonkit classrooms reported teaching science for significantly more minutes throughout the year of this study. These nonkit teachers presented more topics, each of shorter duration than reported in the kit-based curriculum. In spite of the greater effort on the part of these teachers, their students did not perform as well on the measure of science achievement. Additionally, the achievement that was seen in these classrooms was predicted by the frequency of more traditional approaches to teaching science, primarily through textbooks and worksheets. This finding is particularly important as elementary teachers struggle to find time in the day to address the more pressing demands of No Child Left Behind that currently focuses schools on showing progress in reading/language arts and mathematics. More effective science teaching may be accomplished in less time, given high quality materials and professional development targeted specifically to the use of reform pedagogy in guiding students in lessons.

Within the kit-based classrooms there was no apparent difference in science achievement based on the science professional development experiences of the classroom teacher. As mentioned above, this absence of difference could be a result of the coordinated science curriculum program that provides children with a consistent menu of hands-on science taught by teachers who have all had the benefit of intensive professional development related to the specific curricular materials they are using, the science content of each kit, and constructivist, exploratory approaches to instruction. The teachers in the kit classrooms reported significantly greater frequencies of using inquiry and exploratory methods such as student-led investigations, open-ended questioning, and providing field experiences for their students. These teachers also used a substantial amount of more traditional strategies in combination with the more open methods. Thus, the kit teachers were teaching fewer topics in more depth and in fewer instructional minutes throughout the year, yet their students were achieving more than students spending more time in that subject area. This appears to support the popular reform notion that "less is more" if content is taught in greater depth.

Another insight from this study relates to the difficulties of assessing science achievement using an instrument that samples the broad domain of elementary science. The lack of growth seen in this study may be the result of low sensitivity of the test itself to the instruction in any of the groups. The average scores do show a clear difference between the groups at pre- and posttest administrations, but none of the groups showed growth or decline during their 5th-grade year.

Given the expense of developing, implementing, and maintaining the new generation of curricular programs in science, the findings of this study are useful for both participating and nonparticipating schools as educators and administrators seek support for initiating or maintaining kit-based science programs as well as the critical professional development component. In the financial challenges that districts are facing coupled with the national emphasis on reading and mathematics, the ability to examine the value of kit-based science combined with ongoing training support for teachers is essential to implementing and continuing science curricular reform at the elementary level.

APPENDIX A:
Lsc Science Curriculum Matrix for Schools in Current Study—Grades K-6

	K	1	2	3	4	5	6
LIFE		New Plants FOSS	Insects FOSS	Plant Growth STC		Microworlds STC	Ecosystems STC
EARTH and SPACE	Seasons and Weather (GEMS-NET)	Pebbles, Sand, and Silt (FOSS)		Water FOSS	Land and Water STC		
PHYSICAL	Balls and Ramps (INSIGHTS)	Balance and Motion FOSS	Solids and Liquids FOSS	Sound STC	Electric Circuits STC	Sinking and Floating STC	Magnets and Motors STC
TECHNOLOGY			Simple Machines Dacta and GEMS-NET		Motion and Design FOSS	Pulleys and Levers FOSS	Measuring Time STC

REFERENCES

Bredderman, T. (1984). The influence of activity-based elementary science on student outcomes: A quantitative synthesis. *Review of Educational Research* 53: 499–518.

Champion, R. (2002). Sampling can produce solid results. *Journal of Staff Development* 23: 12–16.

Driver, R., Asoko, H., Leach, J., Mortimer, E., and Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher* 23: 5–12.

Escalada, L. T., and Zollman, D. A. (1997). An investigation on the effects of using interactive digital video in a physics classroom on student learning and attitudes. *Journal of Research in Science Teaching* 34: 467–489.

Fensham, P. (1992). Science and technology. In Jackson, P. (Ed.), *Handbook of Research on Curriculum*, Macmillan, New York, pp. 789–829.

Freedman, M. P. (1997). Relationship among laboratory instruction, attitude toward science, and achievement in science knowledge. *Journal of Research in Science Teaching* 34: 342–357.

Kuhn, D., Schaubel, L., and Garcia-Mila, M. (1992). Cross-domain development of science reasoning. *Cognition and Instruction* 9: 285–327.

Lawrenz, F., Huffman, D., and Welch, W. (2001). The science achievement of various subgroups on alternative assessment. *Science Education* 85: 279–290.

Metz, K. (1995). Reassessment of developmental constraints on children's science instruction. *Review of Educational Research* 65: 93–127.

Morrell, P. D., and Lederman, N. G. (1998). Students' attitudes toward school and classroom science: Are they independent phenomena? *School Science and Mathematics* 98: 76–84.

Njoo, M., and de Jong, T. (1993). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching* 30: 821–844.

Okebukola, P. A. (1987). Students' performance in practical chemistry: A study of some related factors. *Journal of Research in Science Teaching* 24: 119–126.

Overstreet, C. (2002). Information re: HRI science assessment instrument.

Parker, V. (2000). Effects of a science intervention program on middle-grade student achievement and attitudes. *School Science and Mathematics* 100: 236–242.

Piaget, J., and Garcia, R. (1989). *Psychogenesis and the history of science*, Columbia University Press, New York.

Schauble, L., Klopfer, L., and Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching* 28: 859–882.

Schneider, R. M., Krajcik, J., Marx, R. W., and Soloway, E. (2002). Performance of students in project-based science classrooms on a national measure of science achievement. *Journal of Research in Science Teaching* 39: 410–422.

Shymansky, J. A. (1989). What research says. . .About ESS, SCIS, and SAPA. *Science and Children* 26: 33–35.

Shymansky, J. A., Kyle, W. C., and Alport, J. (1982a). Research synthesis on the science curriculum projects of the sixties. *Educational Leadership* 40: 63–66.

Shymansky, J. A., Kyle, W., and Alport, J. (1982b). How effective were the hands-on science programs of yesterday? *Science and Children* 20: 14–15.

Shymansky, J. A., Hedges, L. V., and Woodworth, G. (1990). A reassessment of the effects of inquiry-based science curricula of the 60s on student performance. *Journal of Research in Science Teaching* 27: 127–144.

Simpson, R., and Oliver, J. (1990). A summary of major influences on attitude toward achievement in science among adolescent students. *Science Education* 74: 1–18.

Spickler, T., and McCreary, C. (1999). *Making the case for teaching science using a hands-on inquiry-based approach*, Bayer Corporation, Pittsburgh, PA.

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

Von Secker, C. E., and Lissitz, R. W. (1999). Estimating the impact of instructional practices on student achievement in science. *Journal of Research in Science Teaching* 36: 1110–1126.

Walberg, H. J. (1991). Improving school science in advanced and developing countries. *Review of Educational Research* 61: 25–69.

Zacharia, Z. (2003). Beliefs, attitudes, and intentions of science teachers regarding the educational use of computer simulations and inquiry-based experiments in physics. *Journal of Research in Science Teaching* 40: 792–823.