Student Outcomes in a Local Systemic Change Project

Kalyani Raghavan, Shira Cohen-Regev, and Shelley A. Strobel
University of Pittsburgh

The Local Systemic Change initiative of the National Science Foundation supports projects focusing primarily on teacher enhancement through extensive professional development and the use of standards-based curriculum materials. The underlying rationale is that the effective use of such materials will ultimately result in enhanced student learning. However, the research base regarding the impact of these efforts on student learning is rather lean. This paper describes the results of a curriculum-aligned assessment comprising selected items from the Third International Mathematics and Science Study. The assessment was administered to fifth graders involved in a Local Systemic Change project to address the following questions: How does the performance of students involved in the project compare nationally and internationally? and Does length of involvement in the project make a difference in student performance? Additional evidence to relate student outcomes to the project's systemic change efforts are provided.

Considerable effort has been expended nationwide in the past decade to reform science education in schools. *The National Science Education Standards* (National Research Council, 1996) and *Benchmarks for Science Literacy* (Project 2061, 1993) have provided guidelines for aligning teaching and learning that could potentially improve students' understanding of science concepts and processes. Such an alignment of curriculum, instruction, and professional development is at the heart of the Local Systemic Change (LSC) initiative of the National Science Foundation (NSF). The primary focus of projects funded under this initiative is teacher enhancement through extensive professional development and the use of standards-based curriculum materials. The underlying rationale is that increased professional development in the effective use of standards-based curriculum materials will lead to improved instruction and ultimately result in enhanced student learning. As pointed out by Loucks-Horsley and Matsumoto (1999), however, the research base connecting teacher professional development with student learning is rather lean.

In order to garner continued support from various stakeholders, evidence of impact at the student level is needed. Until recently, the evaluation of LSC projects has focused on the quality of professional development and its effects on teaching. As a result, there is little concrete evidence that LSC efforts have improved student learning. This is particularly true in elementary science, because the emphasis at this level is on language arts and mathematics, and many schools opt out of the science component of achievement tests. In a recent report prepared for NSF by Horizon Research Inc., Banilower (2000) cited four elementary science projects that examined student performance on components of achievement tests, such as the SAT-9 open-ended assessment, to assess the impact of LSC-training. Although these studies showed increases in student performance, the report concluded that it is "impossible to judge with any certainty whether the results from these studies are real or spurious" (p. 4).

Allegheny Schools Science Education and Technology (ASSET), Inc., is an LSC project that received a 5-year grant from NSF in the fall of 1995 to improve science teaching in 16 partner districts (Cohort 1). The goal of the project is to promote hands-on, inquiry-based science instruction for kindergarten through sixth grade. Toward this end, ASSET provides professional development and distributes hands-on materials needed to implement nationally established curriculum modules. In the fall of 1998, ASSET received a 3-year supplemental grant from NSF to enlist 14 additional districts (Cohort 2), increasing the total number of participating school districts to 30.

During the past 5 years, ASSET has established a collaborative network facilitating the adoption of a fairly uniform elementary science curriculum across the region. Districts have flexibility in the choice of modules and the grade levels for which they will be used – generally no more than one grade higher or lower than the recommended level. The instructional modules are built around the Focus, Explore, Reflect, and Apply learning cycle.

By the end of the school year 1999-2000, teachers in Cohort 1 districts had received an average of 70 hours of ASSET professional development and were implementing three to four science modules per year. Cohort 2 teachers, on the other hand, had received an average of 30 hours and were implementing only one or two modules per year. The existence of the two cohorts provides a natural framework not only to assess the overall impact of the project on student learning, but also to examine differences in student outcomes between cohorts. Accordingly, a study was designed to assess ASSET fifth graders. One reason for selecting fifth grade was the availability of state mathematics and reading assessment results. Further, Cohort 1 fifth graders have been taught by ASSET-trained teachers using the curriculum modules for 3 to 5 years, whereas Cohort 2 fifth graders have had the experience for 1 to 2 years. This paper describes the results of this study, conducted at the end of the 1999-2000 school year to address the following questions:

1. How does the performance of ASSET students compare nationally and internationally?

2. Does length of involvement in ASSET make a difference in student performance?

This paper concludes by examining additional evidence to relate student outcomes to ASSET's systemic change efforts.

The Assessment

Released items from the Third International Mathematics and Science Study (TIMSS) were used to create a reliable and objective assessment instrument for fifth-grade students that would provide national and international comparison data. Questions were selected to address science concepts and process skills emphasized in the modules used in ASSET districts in grades 1 through 5. The resulting test, called the ASSET TIMSS Assessment (ATA), consisted of 10 multiple-choice items and 10 free-response questions. Three of the latter were two-part questions, which were scored as

separate items, in accordance with TIMSS scoring procedures. TIMSS population-1 items were administered to students in grades 3 and 4, and international difficulty indexes and average scores are available for both grades. In this report, fourth-grade indexes and scores will be cited for population-1 items, which will be referred to as fourth-grade items. Similarly, TIMSS population-2 items, which were administered to seventh and eighth graders, will be called seventh-grade items, and seventh-grade difficulty indexes and scores will be cited. The international difficulty index for fourth-grade items ranged from 537 to 678, with a mean of 630.9 ($SD = 45.76$) and, for seventh-grade items, from 429 to 807, with a mean of 623.92 ($SD = 99.1$). TIMSS item formats were maintained, and questions were divided into two sections of 10 questions each to facilitate administration over two 35- to 40-minute class periods. Teachers were instructed not to exceed 90 minutes overall for the testing.

There were 72 schools with fifth graders in the 29 participating ASSET school districts. (One district declined.) To generate a representative sample containing at least 20% of all fifth graders in the project, one fifth-grade science teacher was randomly selected from each school and asked to administer ATA to one heterogeneous class. Assessment booklets were sent to each school, along with a letter to the school principal and instructions for the selected teacher. Student responses were received from 65 schools (90.3% response rate), representing all districts.

The tests were administered between March 6 and May 15, 2000. Teachers were asked to administer ATA at their convenience, either in two successive class periods or in one period on each of two consecutive days. Teachers were also asked not to assist during the test. Because many of the tests were administered over two days, there were slight differences in the number of student responses for each section. Table 1 depicts the breakdown in response numbers among different groups.

After initial training, members of the evaluation team scored the tests in accordance with TIMSS

Table 1. *Number of Students Responding to Each Section of ATA.*

| | Section 1 | | | | Section 2 | | | |
	Boys	**Girls**	**Not Identified**	**Total**	**Boys**	**Girls**	**Not Identified**	**Total**
Cohort 1 (15 districts/ 33 schools/ 3,805 fifth graders)	422	407	3	832	415	412	1	828
Cohort 2 (14 districts/ 32 schools/ 3,040 fifth graders)	375	373	1	749	353	354	7	714
Total Number of ATA Responses	797	780	4	1,581	768	766	8	1,542

published scoring rubrics with interrater agreement of 92%. Disagreements were resolved after discussions.

Results

This section is divided into three segments. The first outlines overall ASSET student performance in comparison with performances at the international and national levels. Specific test items are described to provide a more complete picture of student performance. The second segment compares the performance of students in the two ASSET cohorts. The third segment presents additional project evaluation results, such as teacher questionnaire responses, professional development, and classroom observations that provide insight into and may help explain differences in student performance.

Performance of ASSET Students

Table 2 provides an overview of ATA, including a brief description of each item, the grade level, international difficulty index, the percent of ASSET students who provided correct answers to each item, and the corresponding international and national averages provided by TIMSS. Following TIMSS scoring rubrics, two-part questions were scored as separate items, resulting in 23 scored items, including 9 fourth-grade items and 14 seventh-grade items.

ASSET students' mean overall score on the ATA was 55.1%, with an average of 64% on the fourth-grade items and 51.4% on the seventh-grade items. ASSET students' scores were significantly higher than both international and national scores ($p < 0.05$) at both grade levels. Out of nine fourth-grade items, ASSET students' scores were better than the national and international scores on eight items. Of the remaining 14 seventh-grade items, ASSET fifth graders' scores were better than the international scores on all but four items.

ASSET students' performance on fourth-grade items. More than half the students correctly answered seven of the nine fourth-grade items. On four of these items, the difference was 15% or higher than national and international scores. In general, performance was

---

Table 2. Overview of ATA Items.

Item Description	Grade Level	Difficulty Index	Average Scores ASSET	Intl	National
Buoyancy and density	Grade 4	665	40	34	31
Separation of mixtures using filtration	Grade 7	605	34	44	39
Solubility rates of loose sugar vs sugar cube *	Grade 4	641	53	37	43
Arm and elbow as a lever system	Grade 7	561	56	51	50
Interpretation of experimental results about seed germination*	Grade 7	678	54	30	46
Electric circuits and proportional reasoning *	Grade 7	571	25	47	28
Science process - observation of a chemical process	Grade 7	594	69	43	57
Explaining why a plain near a river is good for farming *	Grade 7	537	93	76	83
Explaining why a plain near a river is not good for farming *	Grade 7	632	49	38	28
Science process - using the senses for observation	Grade 4	625	83	43	58
Role of plants in an aquarium ecosystem *	Grade 7	474	77	58	61
Role of light in an aquarium ecosystem *	Grade 7	685	30	26	21
Provide evidence for the existence of air *	Grade 7	676	53	34	44
Identification of materials that conduct electricity	Grade 7	429	76	69	75
Design of experiments with plants - controlling variables	Grade 7	624	34	40	41
Chemical/physical changes	Grade 7	670	36	32	30
Stability of ecosystem *	Grade 7	807	9	9	14
Conservation of mass *	Grade 4	672	29	34	21
Experimental design - seed growth	Grade 4	661	76	36	61
Mass comparison and logical reasoning *	Grade 4	640	61	37	42
Identification of animal groups using a chart	Grade 4	534	66	59	59
Suggest one way to reduce air pollution *	Grade 4	580	83	48	59
Suggest two ways to reduce air pollution *	Grade 4	659	56	34	48

Note: Scores indicate the percent of students offering a correct response to each question. Free-response questions are marked with an asterisk. Comparison data reflects international and national performance at the indicated level.

better on items involving environmental science concepts than on those addressing life or physical science. One of the environmental items required students to suggest two different actions that people could take to reduce air pollution. Students were able to suggest several ideas, such as limiting the use of automobiles, making manufacturing changes to reduce emissions from cars and trucks, reducing industrial pollution, and increasing personal efforts such as not smoking or not using aerosol cans. In another item, students were to select the most appropriate experimental design to test whether seeds grow better in the light or dark. More than 75% of ASSET students selected the one choice that would allow a controlled comparison to test the hypothesis. A majority also responded well to a third item that displayed a sequence of paired comparisons involving three objects on a balance. Students were to draw logical inferences from the depicted results, concluding which of the three objects was heaviest and explaining their reasoning.

ASSET students did not do well on two physical science items. The first, involving conservation of mass, depicted the scale reading when a beaker of water and an object are placed side-by-side on the scale's platform, as shown in the left in Figure 1. Students were to mark the scale reading in a second diagram as displayed on the right in Figure 1, in which the beaker is in the center of the platform and the object is inside the beaker of water.

Student responses were equally divided, with roughly a third indicating that the scale reading would be higher, a third indicating it would be lower, and a third indicating it would be unchanged. This pattern of response suggests that ASSET students, who were accustomed to equal arm balances to compare masses and spring scales to measure weight, were probably unfamiliar with the type of scale displayed. It may also be that students who had been exposed to the "Floating and Sinking" module confused apparent and actual weight when an object is immersed in water.

The second item to which more than half responded incorrectly required students to apply their understanding of density and buoyancy to a floating object. Based on a picture depicting a block of wood floating in fresh water (top of Figure 2), students were to select one of four pictures (bottom of Figure 2) depicting what would happen when the block was placed in salt water.
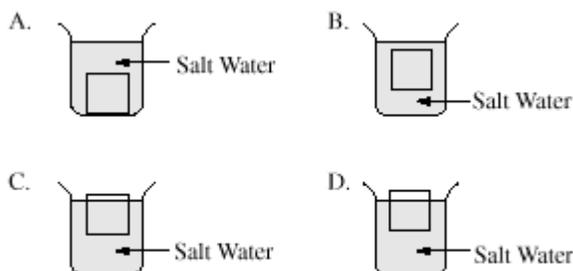
More than a third indicated that the block would sink or float below the surface, which clearly suggests that, although students knew salt water would make a difference, they did not understand why and did not know what kind of difference it would make.

ASSET students' performance on seventh grade items. More than half the students correctly answered 8 of the 14 items. In contrast, less than a third correctly answered the remaining six items. Students performed best on earth science items, and they did moderately well on environmental and life science items. They did not do well on physical science items, particularly chemistry. Students did better on questions that required direct application of known concepts or that elicited understanding of science processes, such as differentiating between observation and prediction or drawing conclusions from experimental results, than on those that required more complex reasoning involving multiple variables, consideration of negative consequences, and differentiation between experimental and control variables in multivariable situations. Overall, these fifth graders' performance on the seventh-grade items paralleled the performance of national and international seventh graders.

A majority of students were able to explain why a plain containing a river might be a good place for farming and why

Figure 1. *Drawing from ATA item involving conservation of mass.*

Figure 2. *Drawing from ATA item involving density and buoyant force.*

it might be bad for farming. Many students recognized that plants in an aquarium ecosystem provide oxygen and food for fish. On the other hand, only a third of the students recognized why light was needed in the ecosystem. Many cited anthropomorphic reasons, such as fish need light to see or for warmth, rather than referring to photosynthesis or energy. This is surprising because Plant Growth and Development, a popular module taught in many ASSET third-grade classrooms, emphasizes that water and light are needed for plant growth, and students grew their plants under artificial light.

A majority of ASSET students, like those in many of the TIMSS countries, did not respond correctly to the two chemistry-related items, even though the concepts addressed are embedded in two modules, Chemical Tests taught in third grade, and Mixtures and Solutions, taught in fifth grade in some schools. One item asked students to identify one of four specified processes resulting in a chemical change. Two thirds of the students appeared to think that change of state or shape is a chemical change. A second item sought to elicit students' understanding of filtration as a process to separate mixtures. A funnel with filter paper was displayed, and students were to select from a list of five mixtures and solutions, the one for which the depicted equipment would be appropriate for separating materials. One fourth of the students selected a solution of alcohol and water or a mixture of sand and sawdust rather than the correct response, which was a mixture of mud and water.

A majority of ASSET students were not able to identify dependent and independent variables needed to design a controlled experiment. The item displayed a plant with sand, minerals, and water in the sun. Students were to select one of five other pictures depicting similar combinations to design an experiment to determine whether plants need minerals for healthy growth. Two of the selections involved multivariable change, and two others controlled for extraneous variables such as water and light. Only about a third of the students answered correctly. More than a third selected experiments that tested the wrong variable, suggesting that although students appear to be aware of the need to control variables, they are not able to select the one that is relevant for the purpose of the investigation.

A vast majority of ASSET students, like those in many TIMSS countries, had difficulty predicting the unwanted consequences of introducing a new species into an environment. Many wrote only about extinction caused by adverse weather conditions or lack of physical strength. "The species could eat a lot of the other animals and they would become extinct," is an example of the most common response. One fifth provided an adequate explanation but did not give an example. "The new species might change the food chain. It would do that by eating all of one species, rare or common, but would change its course of life," is an example of one such response. Another fifth provided an example without a clear explanation, as this example illustrates: "Rabbits were introduced to Australia in order to eat some sort of animal, but then people discovered that rabbits were vegetarians. Now there are too many rabbits."

Another item that was problematic for ASSET and U.S. students displayed an electrical circuit and a partially completed table listing voltages and corresponding current readings from an ammeter in the circuit. Students were to fill in a missing current value, which required proportional reasoning. More than half the students used an additive rule rather than the multiplicative rule that the problem demanded.

Using TIMSS released data, ASSET students' performance on the ATA items were compared against the performance of students from five nations (Singapore, Korea, Japan, Czech Republic, and Hungary) that outperformed U.S. students at the middle-school level (Peak, 1996, p. 23). ASSET students' performance was not significantly different from those of these countries on fourth- or seventh-grade items. In view of the fact that ASSET students did not take the full TIMSS test, single-item comparisons must be considered with caution. Nevertheless, these results suggest that, on questions addressing topics covered in ASSET modules, ASSET fifth graders perform as well as seventh graders in high-performing countries.

One objection to this study could be that TIMSS assessments are not aligned to a single curriculum but, rather, provide a test of general science knowledge and understanding. It is worth noting that not all ASSET fifth graders had been exposed to all the topics addressed in the ATA. Items selected for the ATA, while aligned with topics covered in the ASSET modules, required students to extend and apply knowledge gained over 5 years. The purpose of the ATA was not to assess general scientific literacy but to determine whether ASSET instruction has an impact on student learning of targeted science concepts and processes. The TIMSS scores afforded a benchmark for comparisons.

Comparison of Cohorts

This section compares the performance of students from the two ASSET cohorts – districts with 4 to 5 years of involvement in the project versus districts with 1 to 2 years of involvement. To ascertain that the student populations of the two cohorts were similar and the

comparison between cohorts was valid, SES levels were compared, as were school averages on the 1998-1999 and 1999-2000 Pennsylvania System of School Assessment (PSSA) on reading and mathematics (See Table 3 for details). There was no statistical difference between the cohorts in socioeconomic status or in students' mathematics and reading abilities.

Data collected from the teachers indicate that, on average, three to four class periods were devoted to science instruction in both cohorts. On the ATA, Cohort 1 students scored half a standard deviation higher than their Cohort 2 counterparts (t (63) = 2.037$; p < 0$ .05$;$ Effect Size = 0.5; using intact classes as units of comparison). Individual scores of boys and girls from the two cohorts were compared. Cohort 1 boys scored significantly higher than Cohort 2 boys, and Cohort 1 girls outperformed Cohort 2 girls. There was no significant difference in the performances of girls and boys within cohorts.

Table 4 displays the mean scores of the two cohorts on individual ATA items. Cohort 1 students performed significantly higher on nine items, marked with asterisks in the table. Using the classification schema provided by TIMSS, test items were grouped into topic and skill areas. Figure 3 compares the performances of the two groups by topic areas and Figure 4 by skill areas. Cohort 1 students appear to have a slight edge over Cohort 2 students in physical, earth, and life science areas, and they seem slightly better at using tools or science processes, understanding complex information, and theorizing and problem solving. Similar differences are observed when comparing performances on multiple choice and free response items. The consistency of the

Table 3. *Academic and Demographic Information About Cohorts 1 and 2.*

	Cohort 1	Cohort 2
PSSA Mathematics (Max. 1600) 1998-1999	1353	1344
PSSA Mathematics (Max. 1600) 1999-2000	1358	1344
PSSA Reading (Max. 1600) 1998-1999	1361	1351
PSSA Reading (Max. 1600) 1999-2000	1362	1343
SES	24%	32%

Table 4. *Comparison of Cohort 1 and Cohort 2 Students' Mean Scores on the ATA Items.*

Item Description	Cohort 1	Cohort 2
Buoyancy and density *	47	32
Separation of mixtures using filtration	35	31
Solubility rates of loose sugar vs sugar cube	71	68
Arm and elbow as a lever system *	59	54
Interpretation of experimental results about seed germination	55	53
Electric circuits and proportional reasoning	27	23
Science process - observation of a chemical process *	72	65
Explaining why a plain near a river is good for farming	93	92
Explaining why a plain near a river is not good for farming *	52	45
Science process - using the senses for observation	83	83
Role of plants in an aquarium ecosystem *	81	72
Role of light in an aquarium ecosystem *	35	24
Provide evidence for the existence of air	54	52
Identification of materials that conduct electricity *	80	71
Design experiments with plants - controlling variables	34	35
Chemical/physical changes	38	34
Stability of ecosystem	34	31
Conservation of mass *	32	25
Experimental design - seed growth	77	75
Mass comparison and logical reasoning *	74	70
Identification of animal groups using a chart	66	65
Suggest one way to reduce air pollution	82	83
Suggest two ways to reduce air pollution	54	58

Note: Items marked with * indicate significant difference at the 0.05 level.

Figure 3. *Comparison of Cohort 1 and Cohort 2 student performance on indicated subject areas.*
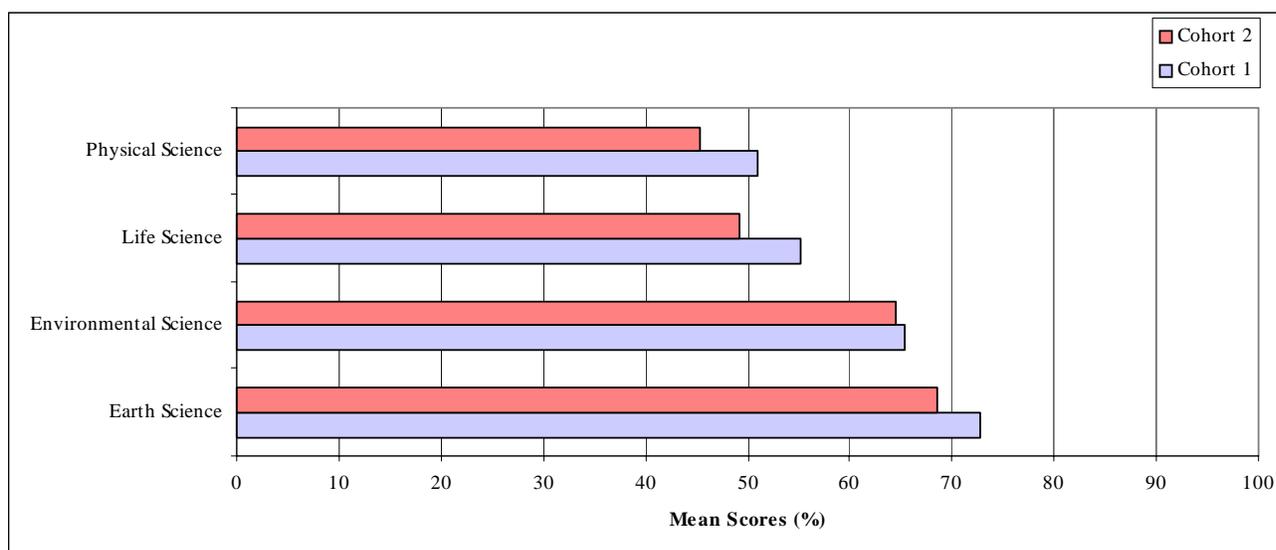
Figure 4. *Comparison of Cohort 1 and Cohort 2 student performance on indicated skill areas.*
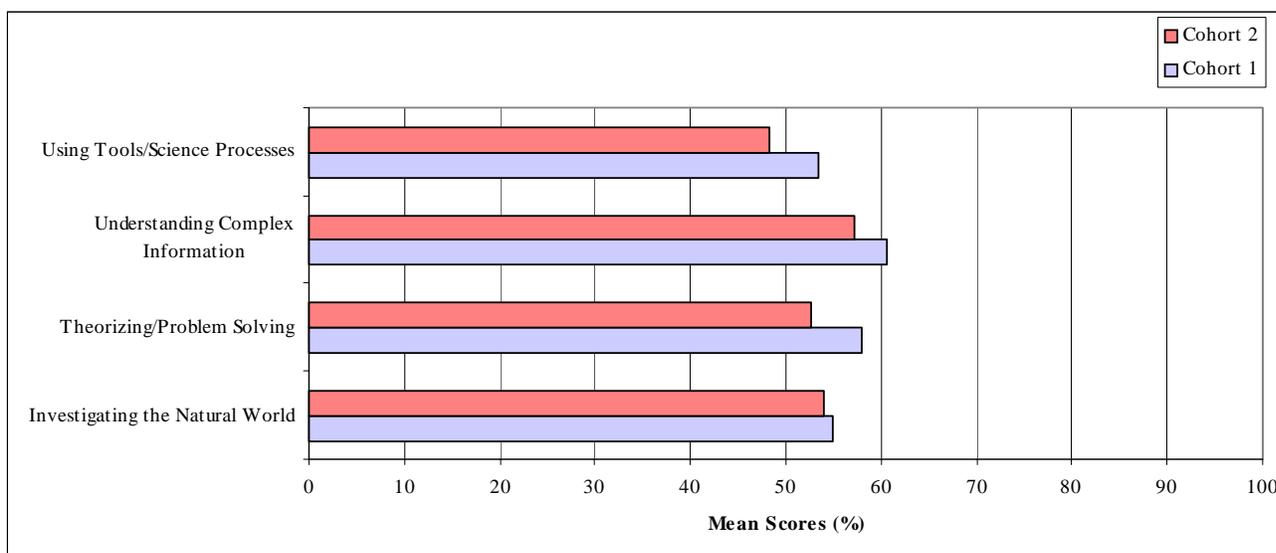
difference across every category indicates that long-term involvement in the ASSET project has a wide-ranging impact on student learning.

Additional Evaluation Results

What are some of the factors that could be contributing to ASSET students' science learning? The standards-based curriculum modules certainly play a role. But to realize their full potential, the modules must engage students intellectually as well as physically. This means that, in addition to knowing the science content embedded in hands-on activities, elementary science teachers need to develop pedagogical content knowledge, the interpretations and transformations of science knowledge to facilitate student learning (Shulman, 1986). ASSET teacher questionnaires, professional development activities in which the teachers participate, and classroom observations provide additional evidence of changes that influence student learning.

Teacher questionnaires. Every year, as part of the project evaluation, questionnaires developed by Horizon Research, Inc., are sent to a random sample of 300 participating science teachers. The questionnaire asks teachers about their attitudes, beliefs, and teaching

practices, in addition to other demographic information. This was done during the 1999-2000 school year for both cohorts, with a response rate of about 80%.

Questionnaire items can be grouped into distinct factors or composites. One composite, *investigative teaching practices*, includes questions about the frequency with which students engage in hands-on activities, design or implement their own investigations, write reflections in notebooks or journals, and work on extended science investigations or projects. The second, *investigative classroom culture*, considers how frequently teachers arrange seating to facilitate student discussion, use open-ended questions, require students to supply evidence to support claims, encourage students to explain concepts to one another, and have students work in cooperative groups.

Between 1996 and 2000, both composites showed moderately significant increases at the 0.05 level within Cohort 1, and the investigative practices composite improved significantly within Cohort 2 between 1999 and 2000. Cohort 1 scored significantly higher than Cohort 2 on both composites in 2000. Engaging students in hands-on activities is a major contributor to the change in investigative practices composite of Cohort 2. A similar change was observed between the 1st and 2nd years of Cohort 1 on this composite and, after staying almost level for 2 years, increased significantly in 2000.

Although the scores are based on teachers' own perceptions, these results clearly indicate that sustained involvement in professional development activities increases teachers' awareness of effective practice and encourages them to gradually incorporate an investigative approach to instruction. However, compared to the scores on preparedness and investigative culture, the score for frequent use of investigative practices is low. This is further confirmed by the analysis of teacher questionnaires from 24 LSC projects conducted by Supovitz and Turner (2000), whose findings indicated a strong link between duration of professional development participation and investigative teaching practice and classroom culture. Overall, there has been no appreciable change in composite scores regarding teacher content preparedness over the years, and there is no significant difference between the cohorts.

Professional development. In addition to providing instructional materials, ASSET offers professional development. The professional development curriculum includes introduction to the Focus-Explore-Reflect-Apply model of teaching, a number of module-specific training sessions, and sessions on assessment, questioning strategies, inquiry teaching, and constructivist learning theory. During the initial years, the project's major focus was to get teachers to use hands-on instructional materials. Consequently, the module-specific training was a get-acquainted-with-the-module approach, with tips for classroom and materials management. During these years, most sessions on pedagogy were isolated from each other and from the context of the instructional modules. Although this approach allowed for focus on one pedagogical concept in depth, it did not facilitate easy transfer into classroom practice. However, in the past 2 years, to advance teachers' understanding of science content and the inquiry process, professional development sessions were redesigned to explicitly link module content, instructional strategies, and assessment. For example, during module training a content expert teamed up with the facilitator to answer content-related questions and clarify and extend participants' understanding of key ideas embedded in the module. In addition, week-long summer institutes were conducted to address major physical-, earth-, and life-science concepts and processes.

Classroom observations. As part of the project evaluation, 28 lessons (all from different districts) were observed and rated in the spring of 2000. Observed classrooms fell into two groups, based on the extent of the teachers' participation in ASSET professional development (PD). The high-PD group comprised 13 teachers who had or were likely to receive close to the targeted 100 hours of PD by the end of the 2000-2001 school year. The low-PD group consisted of 15 teachers who had not. The observed lessons were rated on a scale from 1 (*ineffective*) to 7 (*exemplary*) using the observation protocol developed for NSF by Horizon Research, Inc. A lesson rated 2 contains some features of effective practice, such as students working in cooperative groups, but there are several weaknesses, such as low-level questioning requiring one word responses. A lesson rated 4, on the other hand, may be purposeful but has problems in design, implementation, or content, resulting in a lack of integration of science content and processes. For example, a teacher might pose higher order questions but then short-circuit them by answering without adequate wait time.

The mean rating for the low-PD group was 2.6, compared to 3.62 for the high-PD group. A Kruskal-Wallis ranked-sum test indicated a moderate but significant difference in the overall ratings of the high- and low-PD lessons ($H = 2.75$, $p < 0.05$, one-tailed). Although strengths and weaknesses were found in all lessons, strengths tended to outweigh weaknesses in high-PD lessons. Overall, high-PD lessons were likely to have a positive impact in enhancing students' learning

(Raghavan, 2000). These teachers appeared confident in their ability to teach science, displayed an understanding of science concepts in dialogues with students, and generally presented accurate information. In contrast, attention in low-PD lessons tended to focus more on procedure than content. In high-PD lessons, students were allowed a level of independence, and the teachers tended to facilitate rather than dictate activities. These observations suggest a relationship between teachers' professional development and the extent to which students were given responsibility for learning. They also clearly show that progress toward investigative instruction is not instantaneous but evolves slowly over time with continued professional development and classroom practice.

Discussion and Conclusion

ASSET fifth graders' overall performance, the difference in student performance between the cohorts, and the positive changes revealed by teacher questionnaire responses and classroom observations lend credence to the hypothesis that continuing professional development and implementation of standards-based curriculum modules impact positively on student learning. At the same time, they provide valuable insights that have implications for teaching.

Despite the creditable overall performance of ASSET students on the ATA, there is a parallel between weaknesses observed in ASSET classrooms and weak performance on ATA items. For example, observed teachers often failed to ask follow-up questions seeking clarification or elaboration of student responses. There was a general lack of explicit effort to help students see connections among the various components of the lesson, understand how those components relate to previous lessons, and build a foundation for future learning. For example, in the floating and sinking module, students learn that salt water is denser than fresh water, and they learned about displacement, weight, and buoyant force exerted on objects immersed in water, but there was no lesson that helped students pull all these ideas together or understand how they relate to floating and sinking. Student responses to the floating wood question mentioned earlier reflect this deficiency.

In classroom observations, rarely did teachers or students challenge one another's conclusions, and "what if" questions that could encourage students to consider pros and cons of a phenomenon were not normally used. As described earlier, ASSET students did poorly on an item that asked them to consider the repercussions of introducing a new species into an environment, despite the fact that opportunities to discuss such situations exist in the context of several modules, including Organisms, Insects, Animal Studies, Structures of Life, Environments, and Ecosystems. It may be true that some of these difficulties could be developmental. Still, the instructional modules provide a contextual framework within which students could be challenged to extend their thinking beyond a specific lesson or activity.

The results of this study indicate that the ASSET project has had an impact on student science learning. This is attributable at least in part to districts using quality science curriculum materials that focus on a limited number of concepts and skills. To the extent that ASSET professional developement has enabled teachers to gain confidence with instructional materials and adopt instructional strategies that promote inquiry, it has had an impact on student learning. Recent efforts by ASSET to improve its professional development design by integrating instructional strategies with module training will likely result in further improvements in student learning. A follow-up study, perhaps in 2 years, could test this hypothesis and provide more conclusive evidence of the connection between professional development and student learning. Even without such a study, the results discussed thus far give reason to believe that there is such a connection, underscoring the need for continued professional development to enhance teachers' competence in the art of interweaving pedagogical and disciplinary principles.

References

Banilower, E. (2000). *Local systemic change through teacher enhancement: A summary of project efforts to examine the impact of the LSC on student achievement* [Online]. Available: http://lsc-net.terc.edu/

Loucks-Horsley, S., & Matsumoto, C. (1999). Research on professional development for teachers of mathematics and science: The state of the scene. *School Science and Mathematics, 99*, 258-271.

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

Peak, L. (1996). *Pursuing excellence: A study of U.S. eighth-grade mathematics and science teaching, learning, curriculum, and achievement in international context.* Washington, DC: U.S. Government Printing Office.

Project 2061. (1993). *Benchmarks for science literacy*. American Association for the Advancement of Science. New York: Oxford University Press.

Raghavan, K. (2000). *ASSET 1999-2000 supplemental report on classroom observations*. (Technical Report Submitted to National Science Foundation).

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

Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching, 37*, 963-980.