The Impact of a Multiyear Systemic Reform Effort on Rural Elementary School Students' Science Achievement

James Shymansky
University of Missouri-St. Louis

Tzu-Ling Wang
National Hsinchu University of Education

Leonard Annetta
George Mason University

Susan Everett
University of Michigan-Dearborn

Larry D. Yore
University of Victoria

This paper is a report of the impact of an externally funded, multiyear systemic reform project on students' science achievement on a modified version of the Third International Mathematics and Science Study (TIMSS) test in 33 small, rural school districts in two Midwest states. The systemic reform effort utilized a cascading leadership strategy of professional development delivered at summer workshops and through distance technologies and local leadership groups that focused on helping teachers work in communities of practice to adapt science inquiry lessons to teach and reinforce strategies and skills in language arts in the lessons. Science achievement scores of Grade 3 and Grade 6 student cohorts on the two forms of the TIMSS administered at the beginning, middle, and end of the professional development effort revealed a V-shaped pattern of scores, suggesting that teachers struggled with the newly adapted science inquiries at first but then became more effective in their use. The impact of the adaptation strategy on the students' achievement, questions about the time needed for new instructional strategies to be embraced by teachers, and the wisdom of using "low stakes" achievement tests in studies are discussed.

A Rationale

Between 1991 and 2006, significant U.S. National Science Foundation (NSF) funds were spent on various "systemic change" projects at the state, urban, and local levels, referred to respectively as "state systemic initiatives," "urban systemic initiatives," and finally "local systemic change" (LSC) projects. In each of these system-wide initiatives, energies and funds were focused primarily on the professional development (PD) of teachers—enhancing their content and pedagogical content knowledge (PCK) and inquiry-based teaching practices—on the assumption that these efforts in conjunction with high-quality instructional resources would translate into enhanced classroom practices and most importantly improved student performance. However, it was not until 2001, in the final cycle of LSC project funding, that principal investigators were required to include systematic studies of the project's impact on student achievement—and then, only four LSC studies to date have been reported in the literature (Czerniak, Beltyukova, Struble, Haney, & Lumpe, 2006; Johnson, Fargo, & Kahle, 2010; Klentschy & Molina-De La Torre, 2004; Revak & Kuerbis, 2008). Why is there seeming lack of research in refereed journals that study the system-level PD's impact on student-level achievement if it is assumed that teachers who are better prepared in content matter are more effective in the classroom at enhancing learning? The position taken by Loucks-Horsley and Matsumoto (1999) offers some insight; they claimed that to focus on student outcomes somehow devalues such things as enhanced teacher knowledge and practice, improved leadership, changes in pedagogical strategies, and the implementation of new programs in classrooms. While these arguments sound appealing, the reality is that administrators, teachers, school board members, parents, legislators, and funding agents want and need to see evidence that money spent on system-level teacher PD ultimately does yield benefits at the student level and in particular, in increased student achievement. Equally important, funding agency and school district leaders need to know which models of system-level PD have the greatest potential for producing the greatest student achievement gains, those by which they and their schools are judged in annual evaluations United States Department of Education (2002).

The research reported here focuses on the impact on student achievement in a two-state, multi-school district LSC project "adapting science inquiry lessons" (ASIL). Specifically, this report addresses the question: *What is the impact of an ASIL model of professional development on students' science achievement in rural school districts?*

Theoretical Framework

Attempting to link systemic PD (preservice or in-service) to changes in teacher knowledge and practice and in turn, to changes in student achievement is a daunting task (Holloway, 2002), but not without theoretical support in the literature (Wayne, Yoon, Zhu, Cronen, & Garet, 2008). Investigators studying these links no doubt have to recognize the complexity of large-scale PD efforts and the reality that the education system represents systems and subsystems nested within one another when the states, districts, schools, and classrooms are considered, i.e., (a) that students are nested in groups of other students and teachers comprising the classroom systems; (b) that classroom systems interact with other classrooms to make up the school building system, that then interact with various administrative and political entities to make up the district system, that in turn interact with other district systems to make up the state system, and in some countries, unlike the United States; and (c) that state/provincial systems interact to make up federal/national systems. The complexity of school systems notwithstanding, however, the impact of PD efforts on students should not go unstudied. After all, the link between what a teacher does and what a student learns is the basis of all formal programs of teacher education, certification, and licensure (Kennedy, 1998). However, systemic research requires the use of proxies such as random samples, exemplar cases, selected workshops, and activities to represent the quality of PD, random classroom observations of the teacher collective to represent instruction, the availability of instruction materials to represent the usage of resources, and baseline and annual grade-level performances to represent student learning. In this time of teacher accountability based on student performance on high-stakes tests, it is crucial for the researchers to investigate the optimal PD techniques and longevity to best support teachers.

The theoretical basis for the research reported here has two components. The first examines the literature linking PD practices to changes in teacher knowledge and classroom instruction (Wayne et al., 2008). The second examines the literature establishing a rationale for the ASIL strategy studied and reported here.

Numerous papers have been published on *why* PD is needed (Elmore, 2002; Guskey & Sparks, 2002) and others on *how* PD efforts should be conducted (Loucks-Horsley, Love, Stiles, Mundry, & Hewson, 2003). Borko (2004) sees research on PD occurring in three distinct phases: research on single-site studies focused on a given program's impact on the participating teachers as learners, research on a given program enacted by multiple facilitators at multiple sites, and research comparing multiple PD programs enacted at multiple sites. The phases place the focus on teachers—how they are impacted, assuming again that having teachers who are better prepared logically translates to better student performance.

This focus on teachers has resulted in a collection of studies reporting the impact of PD on teaching practices (Banilower, Heck, & Weiss, 2007; Supovitz & Turner, 2000). Garet et al. (1999) found that teachers participating in PD projects supported by Eisenhower funding consistently self-reported the greater use of inquiry teaching strategies and improvement in their own content knowledge. Supovitz and Turner (2000) suggested that a direct relationship exists between hours of PD and teaching practice, based again on self-reported teacher use of inquiry-based teaching practices and levels of investigative culture in the classroom. They actually went on to suggest 80 PD hours as the point at which PD begins to produce significant differences in both of the self-reported outcomes. Banilower et al. (2007) reported similar positive relationships between hours of PD and self-reported teacher attitudes, beliefs, time spent teaching science, and the perceptions of preparedness to teach inquiry-based science, but they did not identify a specific PD hour threshold. The results of these studies are promising, but all of them relied on self-reported data on teaching practices, and none of them reported anything on the effects of the self-reported changes in teachers on student achievement.

Banilower et al. (2007) proposed a theory for action that assumed (a) high-quality PD focused on science content and PCK and (b) the availability and utilization of high-quality instructional resources would lead to (c) improved inquiry-based teaching practices translating into (d) improved student performance. Links among the first three factors have been documented across several large-scale projects and case studies of elementary and middle school science (Akerson & Hanuscin, 2007; Banilower, Boyd, Pasley, & Weiss, 2006; Johnson, 2006, 2007a, 2007b). These studies indicated that high-quality PD and resources increase the use of inquiry science teaching approaches. However, few published large-scale studies have addressed the link between the clusters of these three factors and factor 4—student performance measured by high-stakes tests (Czerniak et al., 2006; Johnson, Kahle, & Fargo, 2007; Johnson et al., 2010; Klentschy & Molina-De La Torre, 2004; Revak & Kuerbis, 2008; Shymansky, Yore, Annetta, & Everett, 2008). The study reported in this article attempted to connect all four factors in the theory of action by documenting the PD quality, resources usage, classroom practices and student learning using direct observations, teacher and principal self-reports, and student performance on a popular international science assessment.

Despite the numerous articles and books published on why PD is critical and how it should be structured and claims that " . . . student learning outcomes should provide the starting point for *all* educational improvement efforts and PD activities" (Guskey & Sparks, 2002, p. 5), studies on the impact of specific models or strategies of PD on student achievement in science are effectively missing in the post-*Nation at Risk* literature. Reports found that included data on student learning were designed to reveal trends of multisite, but not necessarily single-model PD efforts (Darling-Hammond & Ball, 1998). These reports, although potentially valuable as resources for policy makers, were not intended to and cannot serve as blueprints for designing and conducting system-level PD that will impact student achievement at a given grade level or grade span.

The "adaptation" strategy characterizing the ASIL PD draws theoretical support from two areas in the research literature: (a) using science inquiry as a context in which to teach cross-curricular objectives, and (b) strategies for organizing and facilitating effective PD. The PD strategy of "adapting" science inquiry kits rather than "adopting" them addresses the first area. The idea of adapting lessons in science kits to get "more bang for the buck" when doing hands-on inquiry science is what makes the approach both effective and attractive. It was in an earlier funded PD project (Yore, Shymansky, & Anderson, 2004) that *teachers came up with the idea* of adapting science kits to address language arts objectives to find time for doing science. Children's books were used as a springboard for doing inquiry science, and the inquiry science was used as a springboard for teaching language arts and mathematics. The adaptation strategy used in the earlier project proved successful in two important ways: (a) it brought to the PD program K-6 teachers who had previously been reluctant to teach science or participate in in-service science activities—resulting in more K-6 teachers spending more time on science, and (b) children's attitudes toward and achievement scores in science improved (Shymansky, Yore, & Hand, 2000; Yore et al., 2004), outcomes were viewed very positively by school administrators and parents as well as teachers.

The research of Romance and Vitale (1992) provided the basis for the authors' previous PD project and the follow-up ASIL project reported here. They showed in a comparative study of classrooms integrating vs. separating reading and hands-on science instruction that both reading and science achievement were significantly improved when reading and science instruction were integrated. In the ASIL strategy, science inquiry kits from Lawrence Hall of Science (2003), the National Science Resources Center (2009), and the Education Development Center (1997) were adapted to include discernable reading and writing learning objectives. The strategy caused teachers to think about ways to connect and integrate the science inquiry across the curriculum—to use the inquiry to get students to think, read, write, discuss, and apply the science ideas before, during, and after the inquiry. As in the earlier project (Yore et al., 2004) and consistent with the Romance and Vitale study and other similar studies of integrated strategies (Klentschy & Molina-De La Torre, 2004), the ASIL PD strategy resonated strongly with the K-6 teachers, knowing that doing the science inquiry would help to improve student performance in other critical areas. The focus of the ASIL strategy is also consistent with the call for teacher education programs that " . . . prepare teachers to think about the enterprise of teaching as building on the existing knowledge base and preconceptions of students [and that] convey a model of teachers as learner, who continually develops expertise that is flexible and adaptive" (Donovan, Bransford, & Pellegrino, 1999, p. 46).

The second area of research on which the project strategy draws support is in the literature on what Wenger (1998) and others refer to as "communities of practice," defined as "groups of people informally bound together by shared expertise and passion for a joint enterprise" (Wenger & Snyder, 2000, p. 139). Much of what was attempted and hoped for in the "systemic" part of the LSCs was to develop school district-level communities of practice that would continue after the external funding had ceased. In the ASIL communities of practice, teachers collaborated on preparing or modifying units, analyzing student work, developing assessments, providing feedback, and generally supporting one another as suggested by Duschl, Schweingruber, and Shouse (2007). Working within a community of practice seems to be an especially effective way to share and build on teachers' prior knowledge, local insights, and leadership capacity. Such a community builds feelings of respect and confidence in what teachers know about teaching and learning and is consistent with the National Staff Development Council's conclusion that teachers are in the ideal position to support the PD of colleagues because of their deep understanding of the subjects they teach (National Center for Education Statistics, 2004).

The ASIL PD Strategy

In this section, the key characteristics of the ASIL project are presented along with evaluation data related to its implementation. All of the data presented in this section were collected directly by the project's external evaluator or through procedures administered by an evaluation contractor hired by the agency funding the project.

The ASIL PD project was a multiyear effort targeting the K-6 teaching staff in 33 small, rural school districts scattered across approximately 40,000 square miles of two Midwestern states. The project utilized summer workshops, interactive television (ITV) technology, and local leadership team support during the school year. The goals of the project were consistent with the NSF LSC initiative (i.e., promote inquiry science, use innovative programs, enhance PCK, examine impact, and create infrastructure and support networks). Within these LSC goals, the project staff and participating teachers recognized from the onset that the single most important competence that K-6 teachers lack and need to acquire is the PCK (Shulman, 1987) to make the hands-on science inquiry lessons meaningful for students (Guess-Newsome, 2001).[1] But project staff also recognized that enhanced PCK is not enough to convince K-6 teachers to teach science. Knowing how to make time for science in an already packed elementary school day is equally important. The essence of the ASIL PD plan then was to build a sustainable PD team that could facilitate enhanced PCK on selected Full Option Science System (FOSS), Science and Technology Concepts (STC), and Insights science units and show how science lessons could be adapted to teach more than science during the classroom time allotted for science instruction.

The ASIL PD was provided via a cascading leadership strategy that involved a gradual transfer of responsibilities across the project from experts external to the district to teacher and administrator leadership teams within each district. Educational change is a complex process involving leadership at different levels with different roles (Fullan, 2005). Cascading leadership is based on models for change that recognize that sponsors of change like academics, school district staff, and building principals can promote and authorize changes but cannot affect these changes at the classroom level. They must recruit or empower others to be the change agents or advocates that model, implement, and support the desired change among the targets of change—classroom teachers. The first full year of PD focused exclusively on the leadership teams whose teacher members were referred to as "advocates." The advocates were selected as much for their enthusiasm and interest in teaching science as for their science competence or years of classroom experience. Throughout the project, advocates participated in special retreats and PD sessions to learn the process of adapting inquiry lessons.[2]

ASIL PD, after Year 1, took the form of teachers working in small groups to form regional multi-school district communities of practice focused on building portfolios of adapted lesson plans on selected science topics—portfolios that grew and evolved over the duration of the project. The science topic communities shared their knowledge, set their own agendas for goals to accomplish during the workshop, and worked passionately to be able to teach the science content of the kits more effectively in developing the portfolio. What made the portfolios a powerful and unifying focus for working with teachers from the multiple school districts was that the portfolios of adapted lessons were not specific to publisher kits; they were built specific to science topics, related misconceptions, and associated state standards. This facilitated sharing adapted lessons across publisher kits and building vibrant communities of practice. A sample lesson plan from one of the portfolio sets compiled by an ASIL teacher is shown in the Appendix.

An "interactive television" network, using Polycom Vu-stations (Polycom, San Jose, CA, USA), provided the mechanism for continuing communities of practice activities into the school year when time and distance did not allow travel to workshops. Through this distance learning delivery strategy, teachers participated in a minimum of four two-hour ITV sessions that were held weekly during the school year.[3] ITV sessions focused on a range of activities including presentations by teachers, project staff, and invited experts. The real-time audio/video medium facilitated lively discussions on both teacher challenges and successes in implementing the science kits and the adapted lessons. Ultimately, these sessions were designed to enhance teacher PCK by bolstering content knowledge from expert science faculty doing cutting edge research with enhanced pedagogy through communities of practice discussion on how to integrate the content knowledge learned in an inquiry manner. The local leadership teams composed of district staff, school administrators, and advocates worked to set district-specific priorities and goals and to provide ongoing support for participating teachers as they implemented the adapted lessons in their classrooms.

Collectively, these PD delivery modes and the local leadership team helped establish and maintain communities of teachers within the school districts and schools and across different school districts in the project. These likeminded teachers customized, taught, reviewed, and revised the science portfolios for two or more grade-specific science kits with student ideas, cross-curricular connections, relevant applications, and local resource people, and provided peer support for implementation of these adapted inquiry science kits.

Research Design

The number of PD hours experienced by teachers has been traditionally seen as a key variable in studies and discussions of PD and systemic reform. *The PD hour data figured prominently in the research here as well*. According to Supovitz and Turner (2000), teachers need at least 80 hours of PD before any impact on their classroom practices can be seen. For the later rounds of the LSC projects, the NSF moved this figure to 100 and finally to 130 hours. In the ASIL project, 583 of the targeted 1,269 K-6 teachers (all K-6 teachers, not only science teachers, because science literacy for all students is viewed as a school-wide responsibility in elementary schools) in the school districts participated in 130 hours or more of PD. The actual mean was 148 with a standard deviation of 63 hours and a range of 53–339 hours. These participation figures compare very favorably to those of all 90+ LSC projects funded in which an average of only 13% of the teachers reached the 130-hour goal set by the NSF (Banilower et al., 2006).

The five-year duration of the ASIL effort lent itself to a "time series" research design involving snapshots of the school districts as the project unfolded. The four-component theory of action (i.e., that high-quality PD and resources lead to improve classroom practice and student achievement) served as the basic framework for the research agenda. Elsewhere, positive relationships among the quality of the PD, availability of high-quality resources, instructional time, and classroom practices have been established (Shymansky, Wang, Annetta, Yore, & Everett, 2011, 2012). In this time series design focused on the final relationship between PD and student achievement, the successive grade-level cohorts served as the control/comparative group in the time sequence. The school district-level science achievement for cohorts of Grade 3 and Grade 6 students was gathered in the baseline (Year 0), (early) mid-project (Year 2), and the final (Year 5) year of the project as the data for analysis to explore the systemic impact of the ASIL PD on learning.

Data Sources

The student achievement data for this report consisted of *school district-level scores* from a Grade 3 and Grade 6 form of the "Third International Mathematics and Science Study" (TIMSS) administered to the Grade 3 and Grade 6 students in all 33 school districts. The TIMSS tests consisted of sets of parallel tests generated predominantly (~95%) from released TIMSS items (International Association for the Evaluation of Educational Achievement, 2000) with a small portion (~5%) from the National Assessment of Educational Progress (NAEP) (National Center for Education Statistics, 2000). Four parallel Grade 3 and six parallel Grade 6 TIMSS forms were developed and administered. Each Grade 3 TIMSS test form consisted of 32 items: 25 multiple-choice items and 7 short-answer and extended-response items. Each Grade 6 TIMSS test form consisted of 37 multiple-choice and 5 short-answer and extended-response items. All TIMSS test forms had 10 common multiple-choice items and 3 common short-answer and extended-response items to facilitate the generation of equated scores for analyses. The reliability of the multiple TIMSS test forms for each grade level was analyzed for internal consistency within each form. The consistency values ranged from .73 to .77.

The multiple TIMSS test forms were randomly distributed across students in each classroom tested in a matrix pattern so that some students were assessed by each of the multiple forms in every teacher's classroom. One-way analyses of variance (ANOVAs) of the 2000 results within the two grade levels on the 13 common items across the random subgroups taking the various forms of the test in each grade level revealed no significant ($p < .001$) differences in either grade level. Therefore, we assumed that the four test versions in Grade 3 and six test versions in Grade 6 were valid in detecting similar achievement performances as predicted for the random samples of students with similar instructional backgrounds. Next, one-way ANOVAs were run on the 18 uncommon items in the four Grade 3 test versions and the 29 uncommon items in the six Grade 6 test versions to explore the performance within the larger participating districts where the sample sizes would allow. The ANOVAs results revealed slight but nonsignificant differences ($p < .05$) among test versions for each grade level as expected because the reported item difficulties were similar but not equal for each test version. Based on this evidence, it was assumed that the TIMSS test forms within a grade level were not biased or of different difficulties and that the results across all forms of the respective tests (percent correct) could be aggregated within the two respective grade levels.

Student responses to the multiple-choice items on the TIMSS tests were optically scanned and scored against a master key. Short-answer and extended-response items were hand-scored by teams of independent expert raters using appropriated TIMSS or NAEP scoring rubrics. A random sample of 10 responses per hundred responses for each of the short-answer and extended-response items was double-scored to monitor rater agreement. Exact point rater agreement for the short-answer and extended-response items across all items for all forms of both grade level tests averaged .90 reliability.

Finally, a team of three raters used the outlines of content objectives accompanying the respective FOSS, STC, and Insights kits to generate post hoc *aligned* forms of the achievement tests with items categorically matched to each school district's set of kits adapted and implemented over the course of the five-year project (Webb, 1999). The Grade 3 tests were aligned with the kits being utilized in grades K-3, and the Grade 6 tests were aligned with the kits being utilized across grades K-6. The aligned forms created then for a given school district always contained fewer items than the generalized, unaligned forms, but in all cases, a recheck of the aligned test forms showed internal consistencies of .70 or greater.

In responding to the research question, *what is the impact of an ASIL model of PD on students' science achievement in rural school districts*, baseline (Year 0), mid-project (Year 2), and final year (Year 5) percent correct scores on the two TIMSS tests were aggregated and weighted by each school district's enrollment to produce weighted mean *school district-level* scores for analysis. The cohorts used to determine the 33 district statistics for Grade 3 ranged from 16 to 85 students and for Grade 6 from 14 to 78 students. The weighted *school district means* for the Grade 3 and Grade 6 cohorts were subjected to ANOVA procedures and post hoc Scheffe analyses using grade level and student gender as blocking factors. Using the school district as the unit of analysis more appropriately recognized the systemic focus of the PD on succeeding cohorts of students than an analysis based on individual students across the five project years. The focus on individual students would have required tracking and testing students, an interesting study, but not within the scope, focus, or budget of this research.

Results

The ANOVA conducted on the weighted *school district-level mean* TIMSS student cohort scores for the pre-project, mid-project, and post-project years showed significant differences (at the $p < .05$ level) for year of test, grade level, gender, and gender by grade level. Figure 1 shows the plot of the weighted mean achievement scores for the three test periods, and Table 1 shows the significant ($p < .05$) pair-wise comparisons from the post hoc Scheffe analyses by grade level between gender groups in a given year and within gender groups across the testing years.

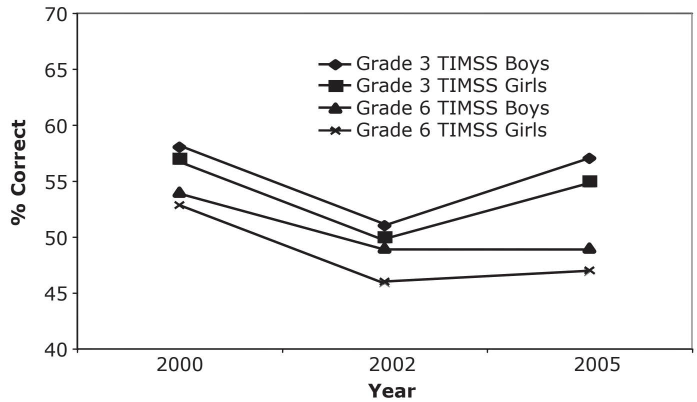

Figure 1. TIMSS school district ($N = 33$) mean total percent correct for Grade 3 and 6 boys and girls.

Table 1
Significant Pair-Wise Comparison of the TIMSS School District (N = 33) Mean Percent Correct Scores

Year	Grade	Gender	Mean (j)	Year	Grade	Gender	Mean (i)	(i – j)*
2000	3	M/F	57	2000	6	M/F	54	−3
2000	3	M	58	2002	3	M	51	−7
2002	3	M	51	2005	3	M	57	+6
2000	3	F	57	2002	3	F	50	−7
2002	3	F	50	2005	3	F	55	+5
2000	6	M	54	2002	6	M	49	−5
2000	6	M	54	2005	6	M	49	−5
2000	6	F	53	2002	6	F	46	−7
2002	6	F	53	2005	6	F	47	−6
2005	3	M/F	56	2005	6	M/F	48	−8

* $p < .05$.

Most remarkable about the plot in Figure 1 is how the mean scores declined for all the cohorts between the pre- and mid-project tests. However, the Grade 3 cohort scores for both boys and girls rebounded between the mid- and final-project years to the pre-project year level while the Grade 6 data for both the boys and the girls stayed at mid-project levels. The Grade 3 cohort scores for boys and girls started the project with higher mean scores than the Grade 6 boys and girls and increased that difference even further by project's end (a difference of three percentage points to eight percentage points).

Discussion

The analyses in response to the research question, what is the impact of an ASIL model of PD on students' science achievement in rural school districts, yielded results that were not totally unexpected. The initial decline followed by a rebound was similar to patterns observed in the earlier study in which PD focused on using fictional stories as a springboard for doing science inquiry (Yore et al., 2004). There are two likely reasons for the declines observed in this ASIL study. The first may be that kit-based science programs such as FOSS, STC, and Insights are designed to de-emphasize factual knowledge, but not to discourage the development of conceptual knowledge. The districts in the project went from a pre-project kit utilization average of <1 kit per Grade K-6 to an average of 2.7 kits per Grade K-2 and 5 kits per Grade 3–6 by 2005. As the teachers changed from the text-based curricula driving their pre-project instruction to the kit-based inquiry promoted in the project, they simply spent less time on the content knowledge needed to score well on tests such as the TIMSS and other high-stakes tests and spent more time using the inquiry-based PCK learned through the PD activities over the course of ASIL. The increased utilization of kits across the board in the districts suggests that teachers in general did embrace this emphasis on inquiry, but may have gotten confused about their role in the instruction and the role of various phases of guided inquiry.

It is well known that teachers who move to hands-on inquiry first focus on the "mechanical use" of activities in the kits and initially assume that students will learn science by just doing hands-on activities with little or no teacher involvement in the learning process (Smith, 2000). The result is that student scores on achievement tests initially plummet because tests such as TIMSS rely heavily on multiple-choice items that demand lower level content knowledge. This explanation is supported by the early annual reports of the project's external evaluator in which concerns were expressed that teachers were weakest in the final "consolidation" or "sense-making" phase of the learning cycle promoted in the project (Shymansky et al., 2008). It is in the consolidation phase that teachers help students review and reflect upon conceptual/content knowledge encountered—knowledge most akin to the kinds of things targeted in any of the achievement tests. The project staff took special note of the evaluator's comments and did place much greater emphasis on consolidating activities (classroom questioning, writing- and reading-to-learning activities, concept mapping, etc.) in the latter years of the project, but the consolidation phase remained the most challenging part of the modified learning cycle strategy for most teachers.

Another possible explanation is that the student cohorts were simply not being assessed as often in the early stages of the project as teachers moved more to the kit-based, modified learning cycle instruction. Although we have no formal data beyond those gathered by the external evaluator on teachers' use of assessments at the lesson level (the final phase of the modified learning cycle) or their use of more summative tests across the project, informal teacher feedback following the No Child Left Behind (NCLB) with its initial emphasis on reading and mathematics suggests that teachers did less testing in science immediately following the imposition of the NCLB even as they attempted to implement the inquiry kits. The TIMSS tests administered were "extra" tests done as a requirement for project participation—they were viewed as a low-stakes test with nothing riding on the results. While project teachers and administrators were interested in the baseline year results, it was clear that there was nothing at stake. Even though there was no evidence either way about how seriously project teachers or students were taking the test at the outset of ASIL (baseline data collection), things changed considerably by the spring of Year 2 when NCLB with its emphasis on reading and mathematics came on the scene. Early momentum and focus on changing the science curriculum shifted abruptly to worries about the reading and mathematics scores. As evidence of this shift, administrators at five of the project's initial roster of school districts decided to drop from the project entirely after one year, and six of the remaining districts secured "Reading First" grants at about the same time. Project wide, it was very clear that the "science" piece of the project lost its luster soon after it was launched. PD providers and researchers need to consider these other pressures in designing multiyear PD and interpreting the PD's impact within a context of changing priorities.

The results of this study provide more evidence that long-term PD is needed to support teachers in making significant changes in their classroom teaching as it pertains to student performance on standardized, norm-reference tests (Supovitz & Turner, 2000). Moving to inquiry-based science teaching requires multiple types of knowledge and skills from teachers: deeper understanding of science content knowledge to facilitate more student-centered investigations, skills to confidently use and teach K-6 students to handle science equipment and materials necessary for inquiry investigations, management of materials and students during hands-on activities, shifting of teaching strategies from very teacher directed to more open ended and student centered, and integrating across the curriculum, particularly with reading, writing and representing. These cannot be learned and implemented quickly with complete proficiency; however, the rebound in the test results shows that many of the participating teachers were apparently successful after having a greater amount of PD and continued opportunities to engage with their communities of practice to share the knowledge gained and to discuss the challenges encountered.

Science educators should note in their work with both preservice and practicing science teachers that the process of getting teachers to shift to or embrace new teaching strategies is not a one-shot deal—it is an evolutionary process that takes an extended time. One science methods course, no matter how well it is taught, will *not* be sufficient in helping all students to develop all of the skills required to teach inquiry-based science with complete proficiency. While preservice teachers may state that they can and will teach science using inquiry methods, in reality, it will take continued support and mentoring to ensure that the shift does indeed take place in the classroom.

Concluding Remarks

As argued in the introduction, the true success of a PD program is measured ultimately in terms of its impact on student learning and attitudes (Guskey & Sparks, 2002). In spite of the myriad uncontrollable factors and alternative explanations that make establishing clear and unequivocal links between systemic PD and system-wide student achievement possible, those who promote and provide PD must nonetheless tackle the task—not only at the classroom level, but at the school building, school district, and state levels as well. In the research reported here, we are fully aware of the problems in establishing the cause–effect link between an ASIL PD strategy provided in the project and student achievement, e.g., initiatives such as the NCLB, competing priorities, financial and staff changes, the unequal commitment and participation of teachers and administrators in the districts served, the changing attitudes and pressures on teachers, administrators, parents, and students across a five-year project, the reliance on measures of student achievement that may be insensitive to or not sufficiently representative of inquiry-based curricula and undervalued by teachers and students, even the problems of changing program officers at the funding agency. But even with these constraints and as tenuous as the links may be with these 33 school districts, these results are important if only to articulate the role of some of these constraints. It is also clear that the importance and normative value to the teachers and administrators of the outcome measure used to determine impact must be factored in when studies of reform are being designed. High-stakes tests—tests whose results are seen as important by teachers, students, administrators, parents, and policy makers should be used in conjunction with or even chosen over low-stakes tests (such as TIMSS) whose results may be of interest to researchers but not to those whose programs and practices we hope to influence.

As for the effectiveness of the specific ASIL PD strategy studied here, the most promising results from the effort, the one thing that kept the project viable for five years was its focus on connecting science to reading and writing—using fictional genre at first as a platform to stimulate science inquiry and later using science as a powerful context in which to teach discipline-specific reading and writing skills. The lesson learned (reinforced with us) is that science educators and funding agents would be wise to link K-6 science curriculum development and associated PD with reading and writing instruction rather than attempt to win over K-6 teachers with stand-alone inquiry activities. Elementary teachers have long recognized the importance of language across the school curriculum, and disciplinary literacy in science education is becoming more common and emphasized in recent standards documents. The constructive-interpretative language arts pairs (speaking–listening, writing–reading, representing–interpreting) are starting to be recognized for their essential roles in learning science and in reporting what we know about science. The Common Core Standards (Council of Chief State School Officers and National Governors Association, 2010) and New Science Education Standards (National Research Council, 2011) partially recognize the constructive and persuasive functions of language as well as the communicative function of language. Although these documents focus on reading and writing, they do recognize the importance of representing, modeling, arguing, seeking, and evaluating information using various print and digital sources in science and engineering. We found in this study and other parallel projects that exploratory talking and listening, graphic and symbolic representing, and interpreting prepared representations were also important in learning science and negotiating understanding in elementary and middle schools. The disciplinary literacy component of "science literacy for all" could do much to improve science understandings and applications of these understandings to contextual issues (Fang, Lamme, & Pringle, 2010).

References

Akerson, V. L., & Hanuscin, D. L. (2007). Teaching nature of science through inquiry: Results of a 3-year professional development project. *Journal of Research in Science Teaching*, *44*, 653–680.

Banilower, E. R., Boyd, S. E., Pasley, J. D., & Weiss, I. R. (2006). *Lessons from a decade of mathematics and science reform: A capstone report for the local systemic change through teacher enhancement initiative*. Chapel Hill, NC: Horizon Research.

Banilower, E. R., Heck, D. J., & Weiss, I. R. (2007). Can PD make the vision of the standards a reality? The impact of the National Science Foundation's local systemic change through teacher enhancement initiative. *Journal of Research in Science Teaching*, *44*, 375–395.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*, 3–15.

Council of Chief State School Officers and National Governors Association. (2010). *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*. Retrieved from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

Czerniak, C. M., Beltyukova, S., Struble, J., Haney, J. J., & Lumpe, A. T. (2006). Do you see what I see? The relationship between a professional development model and student achievement. In R. Yager (Ed.), *Exemplary science in grades 5-8: Standards-based success stories* (pp. 13–39). Arlington, VA: NSTA Press.

Darling-Hammond, L., & Ball, D. (1998). *Teaching for high standards: What policymakers need to know and be able to do*. Philadelphia, PA: Consortium for Policy Research in Education.

Donovan, M. S., Bransford, J. D., & Pellegrino, J. W. (1999). *How people learn: Bridging research and practice*. Washington, DC: National Academy Press.

Duschl, R., Schweingruber, H., & Shouse, A. (Eds.) (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.

Education Development Center, Inc (1997). *Insights: An elementary hands-on inquiry science curriculum*. Dubuque, IA: Kendall/Hunt Publishing Company.

Elmore, R. F. (2002). *Bridging the gap between standards and achievement: The imperatives for PD in education*. Washington, DC: Shanker Institute. Retrieved from: http://www.shankerinstitute.org/Downloads/Bridging_Gap.pdf

Fang, Z., Lamme, L. L., & Pringle, R. M. (2010). *Language and literacy in inquiry-based science classroom, Grades 3-8*. Thousand Oaks, CA: Corwin/NSTA Press.

Fullan, M. (2005). The meaning of educational change: A quarter of a century of learning. In A. Lieberman (Ed.), *The roots of educational change: International handbook of education change* (pp. 202–216). Dordrecht, NL: Springer.

Garet, M., Birman, B., Porter, A., Desimone, L., Herman, R., & Yoon, K. (1999). *Designing effective PD: Lessons from the Eisenhower program*. Washington, DC: U.S. Department of Education.

Guess-Newsome, J. (2001). Pedagogical content knowledge: Introduction and orientation. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge: The construct and its implications for science education* (pp. 21–50). Dordrecht, NL: Kluwer.

Guskey, T., & Sparks, D. (2002). *Linking PD to improvements in student learning*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Holloway, D. L. (2002). Using research to ensure quality teaching in rural schools. *Journal of Research in Rural Education*, *17*(3), 138–153.

International Association for the Evaluation of Educational Achievement (2000). *TIMSS science items (third and fourth grades)*. Chestnut Hill, MA: Author.

Johnson, C. C. (2006). Effective professional development and change in practice: Barriers teachers encounter and implications for reform. *School Science and Mathematics*, *106*, 150–161.

Johnson, C. C. (2007a). Effective science teaching, professional development and No Child Left Behind: Barriers, dilemmas, and reality. *Journal of Science Teacher Education*, *18*, 133–136.

Johnson, C. C. (2007b). Whole-school collaborative sustained professional development and science teacher change: Signs of progress. *Journal of Science Teacher Education*, *18*, 629–661.

Johnson, C. C., Fargo, J. D., & Kahle, J. B. (2010). The cumulative and residual impact of a systemic reform program on teacher change and student learning in science. *School Science and Mathematics*, *110*, 144–159.

Johnson, C. C., Kahle, J. B., & Fargo, J. D. (2007). A study of the effect of sustained, whole-school professional development on student achievement in science. *Journal of Research in Science Teaching*, *44*, 775–786.

Kennedy, M. M. (1998). *Form and substance in in-service teacher education* [Research Monograph No. 13]. Arlington, VA: National Science Foundation.

Klentschy, M., & Molina-De La Torre, E. (2004). Students science notebooks and the inquiry process. In E. W. Saul (Ed.), *Crossing borders in literacy and science instruction: Perspectives on theory and practice* (pp. 340–354). Newark, DE: International Reading Association.

Lawrence Hall of Science, University of California, Berkeley (2003). *Full Option Science System*. Hudson, NH: Delta Education.

Loucks-Horsley, S., Love, N., Stiles, K., Mundry, S., & Hewson, P. (2003). *Designing PD for teachers of science and mathematics* (2nd ed.). Thousand Oaks, CA: Corwin Press.

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

National Center for Education Statistics (2000). *Main NAEP science questions (1990–2005; grades 4, 8, and 12)*. Washington, DC: U.S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/itmrls/searchresults.asp

National Center for Education Statistics (2004). *Highlights from the trends in International Mathematics and Science Study (TIMSS) 2003*. Washington, DC: U.S. Department of Education.

National Research Council (2011). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=13165

National Science Resources Center (2009). *Science and Technology for Children*. Burlington, NC: Carolina Biological Supply Company.

Revak, M., & Kuerbis, P. (2008 January). *The link from professional development to K-6 student achievement in science, math and literacy*. Paper presented at the meeting of the Association for Science Teacher Education, St. Louis, MO.

Romance, N. R., & Vitale, M. R. (1992). A curriculum strategy that expands time for in-depth elementary science instruction by using science-based reading strategies: Effects of a year-long study in grade four. *Journal of Research in Science Teaching*, *29*(6), 545–554.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*(1), 1–22.

Shymansky, J. A., Wang, T. L., Annetta, L. A., Yore, L. D., & Everett, S. A. (2011). The impact of a multi-year, multi-school district K-6 professional development programme designed to integrate science inquiry and language arts on students' high stakes test scores. *International Journal of Science Education*, 1–24. Advance online publication. doi:10.1080/09500693.2011.589478

Shymansky, J. A., Wang, T. L., Annetta, L. A., Yore, L. D., & Everett, S. A. (2012). How much professional development is needed to effect positive gains in K-6 students' achievement on high stakes science tests? *International Journal of Science and Mathematics Education*, *10*(1), 1–19.

Shymansky, J. A., Yore, L. D., Annetta, L. A., & Everett, S. A. (2008). Missouri-Iowa Science Cooperative (Science Co-op): Rural schools—Urban universities collaborative project. *The Rural Educator*, *29*, 1–3.

Shymansky, J. A., Yore, L. D., & Hand, B. M. (2000). Empowering families in hands-on science programs. *School Science and Mathematics*, *100*(1), 48–56.

Smith, D. C. (2000). Content and pedagogical content knowledge for elementary science teacher educators: Knowing our students. *Journal of Science Teacher Education*, *11*(1), 27–46.

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

United States Department of Education (2002). *No Child Left Behind*. Retrieved from http://www.ed.gove/nclb/landing.jhtml

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher*, *37*(8), 469–479.

Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states* [Research Monograph No. 18]. Madison, WI: National Institute for Science Education.

Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge, UK: Cambridge University Press.

Wenger, E., & Snyder, W. (2000). Communities of practice: The organizational frontier. *Harvard Business Review*, *78*(1), 139–145.

Yore, L. D., Shymansky, J. A., & Anderson, J. O. (2004). The impact of a school district's science reform effort on the achievement and attitudes of third and fourth grade students. *Journal of Research in Science Teaching*, *41*(8), 771–790.

Authors' Notes

[1] PCK is defined as a construct that connects general pedagogy, contextual applications and specific content domain and topical knowledge. This integration of assessment strategies and classroom practices reflects the characteristics of nature of science, knowledge structures, and learning difficulties associated with science and specific topics or concepts that come together in effective instruction.

[2] Of the 65 teachers in the initial advocate group, 55 (85%) continued as advocates for the duration of the project.

[3] Hours of participation in ITV sessions were treated the same as hours of participation in the regional summer workshops in determining the total number of hours of PD experienced by individual teachers.

Appendix: Sample ASIL Lesson Plan

Return to Sender: Postcards that Share Information About the World's Biomes

Publisher and Kit Name: STC Ecosystems Kit—Lesson #1 used with 6th sixth graders.
Curriculum Connections: Reading (research skills) and writing (friendly letter/postcard).
Lesson Overview: This lesson provides an opportunity for the students to learn more about the world's ecosystems (biomes). After doing research about one of the world's biomes, groups of students create postcards about the biomes. Each student postcard focuses on one of the biomes with information about producers, consumers, temperature, precipitation, and location.
Objectives (Science): Students will identify regions of similar climates and communities called biomes and name the producers and consumers found in these biomes.
Objectives (Writing): Students will produce postcards that demonstrate clear and coherent writing appropriate to the task, purpose, and audience.
Student Preconceptions/Misconceptions (in Science): Students will have some misconceptions about the biomes; i.e., deserts are always hot, lions live in jungles (rainforests), bears eat coyotes.
Assessment Plan: A two-part rubric is used to determine that the students comprehend the essential features of their biome and can capture the big ideas of the biome in a postcard format.
Reflections on the Lesson: After Lesson 1, the class read pages 7–10 in the STC book *Ecosystems*. I then divided the class into seven groups and each group researched the characteristics of one of Earth's biomes: wetland, grassland, desert, temperate forest, rainforest, Arctic (tundra), marine (see suggested websites at the end of the lesson). Each group created a poster of the important features of their biome to share with the class (producers, consumers, temperature, precipitation, locations). Students then chose one of the biomes to make a postcard. Students were to follow the correct postcard format.

When I first taught the lesson, I found that it was difficult for students to decide what to include onto a postcard, especially those with larger handwriting. When I had students do a second draft of their postcard, I used a much larger note card, and I found that worked better. To help my students with their research, I also created a note-taking sheet to help them organize their facts. When I do this lesson next time, I might explore using a word processing program to create the postcards.

For more information about this adapted lesson, contact: (*teacher's email address*)

Rubric for Biome Postcard

Requirements	Exemplary 4	Accomplished 3	Developing 2	Beginning 1
Postcard content	Postcard includes 4 of the essential features: producers, consumers, temperature, precipitation	Postcard includes 3 of the essential features	Postcard includes 2 of the essential features	Postcard includes 1 or none of the essential features
Postcard illustration	Illustration is neatly & accurately colored	Illustration is neatly colored, but not completely accurate	Coloring and accuracy need to be improved	No illustration
Postcard format	Completely followed correct format	Has 1–2 format errors	Has 3–4 format errors	Has 5+ format errors
Conventions	Has 1–2 errors	Has 3–4 errors	Has 5–6 errors	Has 7+ errors
Presentation	Work is neat, shows creativity and effort	Work is neat and shows effort	Work is messy but shows some effort	Work needs more effort; work is messy

Total points possible: 20 points.