

Museum-Based Teacher Professional Development: Peabody Fellows in Earth Science

Jane Pickering,^{1,a} Jay J. Ague,² Kenneth A. Rath,³ David M. Heiser,¹ and James N. Sirch¹

ABSTRACT

The Peabody Fellows in Earth Science program was a professional development opportunity for middle and high school teachers to enhance their knowledge of, and teaching skills in, the Earth sciences. It combined a summer institute and academic year workshops with the production of new curricular resources on the interpretation of landforms in Connecticut. Teachers implemented these materials with the aid of an accompanying classroom kit. The program included in-depth evaluation of the project outcomes for teachers and students; focused on tectonic processes; and on erosion, weathering, and glaciers. Forty-seven teachers participated in the institute, 30 taught the full curriculum, and 21 completed all the evaluation activities. Teachers reported that they had significantly increased their geoscience content knowledge as well as their ability to teach geoscience-related skills, particularly in guiding students to make observations and inferences about the local landscape. In the majority of cases for teachers that completed all evaluation activities, there was a significant increase in student performance in at least one learning goal as a result of the teachers' participation in the program. These data show this type of informal-formal educational initiative can be highly successful in improving teacher competence and student learning in the geosciences. They also provide evidence that positive proximal outcomes for the teachers are reflected in the ultimate outcomes for students. © 2012 National Association of Geoscience Teachers. [DOI: 10.5408/11-241.1]

Key words: teacher professional development, museum, informal science education

INTRODUCTION

“All young Americans should be educated to be ‘STEM [Science, Technology, Engineering, Mathematics] capable’ no matter where they live, what educational path they pursue, or in which field they choose to work.” Few would disagree with this statement from the report of the Carnegie Corporation of New York and the Institute for Advanced Study on the ultimate goal of math and science education in the United States (Griffiths and Cahill, 2009). However, while it is almost universally acknowledged that math and science need to be at the center of the educational enterprise, there is equal acknowledgement that science education needs transformational improvement to prepare students to be scientifically literate (e.g., National Academy of Sciences, 2007; President’s Council, 2010). Perhaps nowhere else is this directive more important than with the Earth sciences, which are vital to understand and mitigate the growing problems of resource depletion, energy sustainability, environmental degradation, and climate change (National Science Foundation, 2009). However, “Despite the relevance and increasing importance of the geosciences to everyday matters, few Americans understand the fundamental workings of their planet” (National Science Foundation, 2009), and the Earth sciences are relatively “invisible” in the K–12 curriculum (McNeal, 2010). For example, the 2007 State

Indicators of Math and Science Education (Council of Chief State School Officers, 2007) show, on average, only 21% of high school students take an Earth science course, compared with over 95% taking a biology course. The Peabody Fellows in Earth Science was designed to address this critical need to engage students in the Earth sciences (Barstow and Geary, 2001; Scotchmoor et al. 2005; McNeal, 2010).

An editorial in *Science* opined that “the most important element of any education system is a highly skilled teacher” (Alberts, 2009). Research consistently demonstrates that teacher expertise and quality is a critical, if not *the* critical factor in student success (e.g., National Academy of Sciences, 2007; Griffiths and Cahill, 2009) and that the quality of teaching has a huge impact on student learning (National Academy of Sciences, 2010; National Science Board, 2010). Unfortunately recent Science and Engineering Indicators (National Science Board, 2010) show that only 40% of fifth graders across the nation are taught science by in-field certified teachers, and the physical and space sciences are often taught by teachers with minimal qualifications (Ward et al., 2008). In addition, by any measure (e.g., certification, experience) low-income, black, and Hispanic students are significantly less likely to be taught science by a well-qualified teacher (Lankford et al., 2002; Corocan and Evans, 2008; Tomás Rivera Policy Institute, 2008; National Science Board, 2010). Nearly half of all new teachers in urban public schools have been in their school for 3 years or less, and high-poverty schools experience higher annual turnover rates than do low-poverty schools (Howey et al., 2008).

The professional development opportunity described here addressed these challenges by using the resources of a university-based museum to support Connecticut teachers from a variety of public school districts. Specifically the program’s goal was to enhance teachers’ knowledge of, and teaching skills in, the Earth sciences and therefore to

Received Received 6 May 2011; revised 29 March and 11 July 2012; accepted 13 July 2012; published online 6 November 2012.

¹Yale Peabody Museum of Natural History, P.O. Box 208118, New Haven, Connecticut 06520-8118, USA

²Department of Geology and Geophysics, Yale University, P.O. Box 208109, New Haven, Connecticut 06520-8109, USA

³SageFox Consulting Group, 30 Boltwood Walk, Amherst, Massachusetts 01002, USA

^aAuthor to whom correspondence should be addressed. Electronic mail: jane.pickering@yale.edu

improve their students' skills and knowledge. The main outcomes centered on teachers' and students' ability to interpret and investigate their local landscape and the geologic forces that shaped that landscape as follows.

Teachers would:

- Know how to interpret landforms that result from different geologic forces, both in the State of Connecticut and across Earth, and have the resources and skills to effectively teach associated topics in the framework of local, state, and national science standards.
- Learn how to use observation, field sites, and research results to understand Earth sciences.
- Increase their use of interdisciplinary and inquiry-based teaching methods in the classroom, for example, through the incorporation of hands-on materials into the teaching of their curriculum.

Students would:

- Learn that their local environment, the State of Connecticut, provides numerous windows into the past, where one can observe evidence of constructive and destructive geological forces including tectonic processes, and glaciation, weathering, and erosion.

The program, through collaboration with school districts that have high enrollments of students of color, was designed to support African-American and Hispanic students who are grossly under-represented in all parts of the science, technology, engineering, and mathematics (STEM) education pipeline and science careers, especially those involving the physical sciences (Czujko *et al.*, 2008; Tomás Rivera Policy Institute, 2008). The instructional approach taken was to develop and evaluate a program model that brought together best practices in professional development design with the resources of a university-based natural history museum (collections, exhibits, staff, faculty). Informal science institutions (ISIs)¹ make significant contributions to the K–12 learning environment. A survey of 475 informal science centers in the United States (Phillips *et al.*, 2007) discovered that over 70% of ISIs have programs specifically designed for school audiences. However, that same survey revealed only 16% offered teacher institutes with the 40 h (or more) contact time recommended for effective professional development (for example, see Penuel *et al.*, 2009). Those that do are almost all very large museums (for example, the Institute for Inquiry at the Exploratorium, MUSE at the Science Museum of Minnesota, and Urban Advantage at the American Museum of Natural History) that have substantial staff and other resources not available to many smaller museums and their communities. The program aimed to pilot an in-depth professional development model that could be implemented at small-/medium-sized institutions. It also aimed to investigate teachers' perceptions of how the program supported their instruction, and study the outcomes for the students themselves, both to test the model

and to contribute to the currently very limited quantitative evidence of the effectiveness (or not) of professional development based in a museum.

CONTEXT

"Great" STEM teachers have a deep content knowledge and a strong pedagogical training specific to STEM (President's Council, 2010). The nationwide lack of teachers with Earth science credentials is considered to be a major factor in the lack of awareness of, and interest in, the geosciences among students (Huntoon *et al.*, 2005; McNeal, 2010). In many states (including Connecticut) the Earth sciences are most commonly taught during the eighth grade, and certification standards for middle school mean most teachers have only introductory coursework in the Earth sciences (American Geological Institute, 2007). With little preparation, many teachers have misconceptions about fundamental topics, such as plate tectonic theory or the rock cycle, that greatly inhibit geoscience understanding (King, 2008). An unfamiliar content area also means teachers might not use the hands-on, experiential teaching methods that effectively engage students. It is vital that in-service teachers are able to maintain continuous contact with fresh content to keep their knowledge up to date (Griffiths and Cahill, 2009) as well as to improve their understanding of strategies that prepare and inspire students to be successful in science (National Academy of Sciences, 2010). Science teachers themselves report that they believe ongoing professional development opportunities in the geosciences are essential for them to teach the subject effectively (Science Education Resource Center, 2003).

The current environment of accountability and test taking in the nation's public schools mandated by U.S. department of Education's No Child Left Behind legislation has had a profound impact on all aspects of K–12 education. One of these is that the increasing emphasis on teaching to the test often results in the loss of a connection to the world of resources outside of the classroom (Bevan *et al.*, 2010). As it becomes increasingly hard for teachers to take students away from school, museum–school relationships must expand from the field trip model to one of an ongoing resource partnership to continue to play a significant role in supporting and improving formal K–12 education (e.g., Astor-Jack *et al.*, 2007; Bamberger and Tal, 2007; National Research Council, 2009). In their seminal report on Earth science education "Blueprint for Change," Barstow and Geary (2001) state that museums have the capacity to be key providers of high-quality materials that support national science standards as well as of exemplary teacher professional development opportunities. Teachers, too, see museums as places to learn science content, to improve their pedagogical skills, to help them develop resources and strategies that engage and sustain student interest, and to interact with science-rich professional communities (Phillips *et al.*, 2007; MacDonald *et al.*, 2008).

In an in-depth study on the role ISIs can play in K–12 education (Bevan *et al.* 2010), the authors identify a number of structural features and affordances² these institutions

¹ The term "Informal Science Institution" is used throughout the paper to refer to science centers, museums, and other science-rich cultural institutions.

² Affordances are the opportunities for acting, thinking, or feeling that are provided by a given environment. See Bevan *et al.*, 2010, p. 23, for a discussion.

provide and how they complement the school environment. ISIs routinely provide three-dimensional exhibits, real objects, and immersive experiences that afford different types of student engagement and understanding than that provided by school. They also have access to different types of data and connections to current science that can engage students in current research. In particular, the object-based and place-based approaches that are fundamental to ISI's pedagogy naturally promote the scientific processes of inquiry. For teachers, ISIs can provide resources, including staff expertise and availability, for ongoing and sustained professional development support, as well as curricular materials and site-based programs that are integrated into the core academic curriculum. ISIs also tend to focus more on affective and practical approaches to professional development that complement the more cognitive approach of other providers (Astor Jack et al., 2007). In summary, formal–informal collaborations lead to conceptually rich and compelling science learning programs, as well as to the creation of learning communities that bring together teachers from different districts and informal educators.

However, as noted earlier, few ISIs are involved in in-depth teacher professional development. The models that do exist (e.g., MacDonald et al., 2008) are far beyond most institutions' capacity to implement. In addition, and very importantly, research on the effect of informal science institutions' efforts in teacher professional development is extremely scarce (Astor-Jack et al., 2007; Bevan et al., 2010). The ultimate goal of any professional development is one step removed from the program—its effect on student learning. While it is generally accepted that the quantity of professional development (and its content) has effects on teaching practice and classroom culture and is an important predictor of student outcomes (Desimone et al., 2002; Banilower et al., 2007; Yoon et al., 2007), studies that look directly at the effect on student learning are far fewer (Yoon et al., 2007; Blank and de las Alas, 2009). As summarized in a recent National Academies report on learning in informal science environments (National Research Council 2009, p. 132–134), most research on the impact of ISIs on K–12 education has focused on field trips. This work has not been extended to examine the impact (or not) of professional development programs and the impact on the students themselves. There is urgent need for more documentation and evidence of the outcomes of ISI activities in this area (Bevan et al., 2010).

Program Design

The Peabody Fellows program was designed according to best practices for effective professional development in the sciences as well as taking full advantage of its setting in a museum: It contained rigorous subject content, addressed science standards and was focused on classroom practice, encouraged inquiry-oriented teaching techniques through the use of place-based and object-based learning, and incorporated support over time for teachers to integrate new skills into their teaching (see, for example, Banilower et al., 2007; Blank and de las Alas, 2009; Bevan et al., 2010; National Science Board, 2010). In addition, it incorporated elements recognized as being particularly important for Earth science professional development: It provided field experiences, introduced teachers to geoscience professionals, and provided an integrated geoscience perspective (Science

Education Resource Center, 2003; Huntoon et al., 2005). Since the Museum recruited teachers primarily from urban school districts that have high enrollments of students of color, strategies shown to be especially effective for this audience were used (Science Education Resource Center, 2009).

The primary pedagogical model the program used was to adopt a “place-based” approach that uses the local community and environment as a starting point to teach concepts across the curriculum, and has been shown to be particularly effective approach for urban youth (Barstow and Geary, 2001; Harnik and Ross, 2004; Science Education Resource Center, 2009). This addressed the fact that most urban teachers are not aware of the many different ways that *local* geology can be used as an effective context for Earth science topics (Kean and Enochs, 2001; Birnbaum, 2004). It also allows teachers to address the misconception many urban students have that Earth science is not relevant to their lives (as summarized by the Science Education Resource Center, 2009).

Another hallmark of the program was the significant involvement of teachers in its development, and in its acknowledgement (both through explicit statements and through ensuring sufficient time during the institute) of the importance of the teachers' own contributions and decisions on designing the students' educational experiences. Teacher input into curriculum design (as well as assessment) is vital in today's high-stakes testing environment when teachers in different states, and even in neighboring districts, operate in different environments both in terms of content and pacing through the school year. It is also recognized that involving teachers in curriculum decisions is far more effective for their students (for a summary discussion of this topic, see Penuel and Gallagher, 2009). The Peabody Fellows program's model provided opportunities for teachers to collaborate with their peers and interact with expert mentors in the form of museum educators and university science faculty. Interactions with their peers help teachers make their interpretation and instructional decisions more explicit, and mentors help teachers address gaps between curriculum writers' design intentions and teacher enactment (Penuel and Gallagher, 2009). There are extensive and excellent Earth science K–12 activities on the Web and in print. However, they are rarely aligned with curricula from the local school district or state standards, and they need to be linked to those standards to make them attractive and useful for teachers (Penuel et al., 2007; Mervis, 2009). Given this wealth of resources, rather than developing new materials and activities, the project team decided to bring together a set of nationally recognized educational activities to produce a customized and locally relevant curriculum for Connecticut teachers.

Finally, based on the Museum's prior experience, the program model paid attention to several practical issues. Teachers were offered stipends in recognition of their commitment and to encourage completion of evaluation activities. Another important activity was involving teachers and school administrators, particularly science supervisors, from the beginning of the project to ensure its relevance and build ownership within the schools and districts. This is important to foster teachers' perception of coherence i.e., how well-aligned the activities are with their own goals for learning. This coherence is closely linked to the teachers'

understanding of their own district's goals, level of administrative support, and the social pressures within the schools; a program that supports this is more likely to foster change in practice, knowledge and support curriculum implementation (Penuel *et al.*, 2009). In addition, the assessment instrument for the program was modeled on the State's test instruments, as research has shown that providing assessment to school districts that demonstrate how the program impacts student learning increases participation (Bevan, 2007).

PROGRAM AUDIENCE

The Peabody Fellows in Earth Science served 47 teachers, all of whom were teaching in Connecticut. We recruited teachers from a variety of districts, about half of which were from Title 1 high-needs districts. The teachers were generally experienced, with only 12% having less than 4 years of teaching experience, compared with 62% with more than 9 years of teaching experience. Eighty percent were teaching students from grades six to eight. Just over 70% were female, and just under 90% were white.

Connecticut is home to some of the wealthiest communities in the United States as well as the poorest, and has one of the highest achievement gaps in the country between poor and nonpoor families, and between white students and their black and Hispanic peers (Ali and Dufresne, 2008). The program proactively recruited teachers from districts designated as "high needs" by No Child Left Behind legislation, and over half of participants were from such districts. For example, New Haven has a school district that educates almost 20,000 students, of whom 87% are students of color and 80% (up from 62% in 2006) are eligible for free/reduced-price meals (Strategic School Profile, 2010). The dropout rate is twice the State average. New Haven schoolchildren are educationally disadvantaged, consistently demonstrating low performance levels on mastery tests. Up to 75% of 4th–8th grade students perform below the State goals in reading, writing, and math; only 24.9% of 8th grade students and 15.5% of 10th grade students meet State goals in science (compared with 63 and 45% statewide). In Bridgeport public schools, another diverse urban environment, 98% of students are eligible for free/reduced-price lunch, 91% are students of color, and only 25% of 8th grade students meet State goals in science. Other school districts that participated include those in the Naugatuck River Valley, an economically depressed region of industrial and manufacturing towns such as the Meriden Public School District, in which 62% of students are eligible for free/reduced-price lunch, 61% are students of color, and only 34% of 8th grade students meet State goals for science.

In Connecticut, the implementation of a new State framework for science, together with the introduction of performance-based assessments in science for grades five and eight in 2008, has focused teachers' and administrators' attention on sciences and the Earth sciences in particular. Discussions between the Museum and a number of science supervisors across the State indicated that Earth science was the top priority for professional development. It should be noted that while many museum-based teacher professional development programs are under-enrolled, recruitment for the program was not difficult, as it closely matched the schools' expressed need.

METHODS

Program Content and Implementation

The program was explicitly built around a seventh grade Connecticut Science Framework 7.3: "Landforms are the result of the interaction of constructive and destructive forces over time." It includes content and activities that can be adapted for use in a sixth grade class as well high school Earth science classes. In Connecticut, as elsewhere, the science standards are structured around landscapes—it is the landscapes that are the result of the various tectonic and erosional processes that students are expected to be able to recognize and understand. For example, by the end of middle school Connecticut students are expected to understand how volcanic activity and the folding and faulting of rock layers during the shifting of the Earth's crust affect the formation of mountains, ridges, and valleys. Within 9 miles (14.4 km) of the center of New Haven, lies the Orange–Milford belt, the exposed remnant of a strongly folded and deformed rock unit that was at one time buried at depths of 30–40 km. Using real rock samples along with materials from author Ague's research on metamorphic processes within the Orange–Milford belt, teachers and students can investigate in their "own backyard" how rocks can be altered by heat and pressure deep within Earth and brought back to the surface.

The central 3-week curriculum unit was developed by Museum educators working closely with four very experienced classroom teachers. These "resource development teachers" were recruited the winter before the summer institute and met with Museum staff for six 4-h sessions to determine guiding themes, and choose new and existing core activities that integrated Connecticut State standards and expected performances. Rather than developing new materials and activities, the team decided to bring together a set of nationally recognized educational activities to produce a customized curriculum for Connecticut teachers. There were three sections of the unit: geologic time; tectonic processes; and glaciation, weathering, and erosion. A fourth section detailed additional resources that teachers could use to find suggested field sites in the State, online information and activities, and recommended publications. The detailed curriculum can be found online at: <http://peabody.yale.edu/sites/default/files/documents/education/Connecticut%20Geology%20Guide%2013OCT2011.pdf> and is summarized in Table I.

The program began with a 3-day summer institute for the teachers that included lectures, workshops, participatory demonstrations of classroom activities, and an all-day field trip to a variety of sites in Connecticut to view different landscapes and geological features (see the supplementary materials for details of the Institute Agenda; available at: <http://dx.doi.org/10.5408/11-241s1>). The activities were designed to introduce the curriculum and model its use in the classroom. Activities included how to use the Peabody Fellows geoscience kit in the classroom, current geoscience digital resources, and how to integrate Museum exhibits into the unit. Time was allowed for teachers to work with colleagues, and share ideas, activities, and teaching strategies. Particular attention was paid to topics subject to student confusion, such as the structure of the mantle, and the difference between weathering and erosion (King, 2000), as well as areas where teachers requested more content. The latter mostly surrounded teachers' knowledge of State

TABLE I: Summary of program's curriculum guide.

Curriculum section	Activity	Source
1. Geologic time	Geologic time scale	New activity: Earth Timeline and Connecticut Time Cards produced by the program
2. Tectonic processes	Global patterns of earthquakes and volcanoes	Gallery activity at the Peabody Museum; alternative activity involves study of http://www.johomaps.com/world/worldtecton.jpg
	A convection cell	Revised hands-on activity: convection currents in water were created using red dye, blue-dyed ice cubes, and hot cups
	Snack tectonics	National Earth Science Teachers Activity, see http://www.windows2universe.org
	Gelatin volcanoes	Hawaii Space Grant Consortium activity, see http://www.spacegrant.hawaii.edu/class_acts/GelVolTe.html
	Demonstrating geologic processes—modeling using foam pads	Professor L.W. Braile, see http://web.ics.purdue.edu/~braile/edumod/foammod/foammod.htm
3. Glaciation, weathering and erosion	Investigating how glacial features are formed in rocks	National Science Teachers Association, <i>Explaining Glaciers Accurately</i> , see <i>Science and Children</i> , May, 2009, NSTA, and Glacial Shuffle Activity from Yosemite Institute
	Delineating a watershed	Adapted from a fact sheet by the Natural Resources Conservation Service, see http://www.nh.nrcs.usda.gov/technical/Publications/Topowatershed.pdf
	Reading the landscape online!	New student worksheet for use with a variety of Earth observation Web sites e.g., National Geographic Map Machine
	Stream table investigations	Exploratorium, San Francisco, see http://www.exploratorium.edu/IFI/docs/Stream_Table.pdf
4. Additional resources	Not applicable	Resources include Web site for additional content on plate tectonics, summary of key geologic features in Connecticut, list of Connecticut geologic sites for student field trips, links to general Earth science teaching resources, list of publications on Connecticut geology

geology, and supported the need for the program's place-based pedagogical approach. For example, we found that while many teachers teach tectonic processes (e.g., volcanism), in general, they did not appear to have content knowledge to related tectonic processes affecting the landscape in their local areas.

The Museum then provided ongoing assistance throughout the school year, including two follow-up academic workshops, digital resources, consultation with staff, and a social networking site. Teachers were kept up to date on content, with geology lectures and field trips happening at the Museum, the Yale Department of Geology and Geophysics, and the Geological Society of Connecticut. The workshops, jointly led by faculty and Museum educators, included a lecture on extraordinarily preserved fossils, with a behind-the-scenes tour of the Peabody, a field trip to view polymetamorphism in northeastern Connecticut, and a talk on the late Paleozoic tectonic evolution of eastern New England. A special lecture was added on the structure and dynamics of the mantle in addition to covering the topic in the institute itself.

Teachers were also able to borrow a classroom kit that included the curriculum unit guide and resource materials; stream tables and other activity supplies; rock samples from throughout the State, with site photos and descriptions; hand lenses or loupes; guides to Connecticut geology; and PowerPoint presentations on various topics (see the supplementary materials for a photograph of the Classroom Kit; available at: <http://dx.doi.org/10.5408/11-241s2>). The kit

was extremely important for teachers to use the object-based approach modeled by Museum staff in the classroom.

Program Evaluation

The program evaluation had two parts, teacher surveys and a series of student tests.

Teachers

The teachers completed three surveys. The first, pre-institute survey was given in the spring before the institute and collected data on participant background, expectations of the institute, and baseline information on areas where the program hoped to engender change in teacher responses: knowledge of Earth sciences and comfort with inquiry-based teaching and state standards around the Earth sciences. The second, postinstitute survey collected data on respondents' reflections on the value of various aspects of the summer institute as well as how they expected to be able to use what they learned in the classroom setting. They were also asked to reflect on their new state of knowledge about Earth sciences. Finally, there was a post-lesson survey that gathered information about how well the materials had worked in respondents' classrooms, collected feedback on how things might have been improved, and investigated the long-term impacts of the materials had been on student and, more particularly, teacher learning. For the latter, a number of the responses on this survey were compared with those on earlier surveys to determine which attitudes and areas of knowledge showed change.

In all cases, attitudinal items and items measuring their state of knowledge were structured such that the teachers were asked to indicate the extent to which they agreed with a series of statements, based on a five-point scale. Some of these answers were probed more deeply with open-ended questions.

Students

An identical student test was administered on three occasions during the program. The first, termed the “baseline posttest,” was administered in the teacher’s classroom during spring 2009, prior to attendance at the institute. These results from this first test provided data regarding how a (presumably) typical class would respond to the test in the absence of the teacher having participated in the program. The following academic year (2009–2010) the teachers administered the test to their class twice, once as a pretest at the beginning of the school year and again as a posttest after having taught the curriculum. Posttest scores were compared with both the pretest, to examine score increases as a function of having taken the class, and to the baseline posttest to examine increases compared with what might be expected at the end of the year in the absence of Peabody Fellows’ lessons.

The student test instrument (see the Student Assessment Instrument; available at: <http://dx.doi.org/10.5408/11-241s3>) was designed to mimic Connecticut Mastery Tests, and was written and piloted in collaboration with the resource development teachers. It included 13 multiple-choice questions and 3 short-answer science process skills questions. The teachers themselves scored the multiple-choice questions, while the short-answer questions were graded by trained volunteers at the Museum to ensure intergrader reliability.

The process of scoring the short-answer questions began with the production of a scoring rubric by the evaluator and the project team. Initially, the teachers were asked to score the short-answer items by using the rubric, but it became quickly apparent that the teachers were interpreting the rubric differently, and that inter-rater reliability was a serious problem. To combat this, four volunteers were selected at the Museum to act as scorers, and these four people scored all the baseline, pre-, and posttests. They were introduced to the task through three multihour training sessions with the evaluator and project leaders. There, the team worked together to develop a shared understanding of how the rubric should translate to scores on the items by independently scoring several students’ responses and then comparing the answers. Only when it became clear that there was a consistent understanding of what scores should be given to the individual items did the volunteers do any of the scoring on their own, and even then they worked in pairs so that they could discuss particularly difficult-to-score responses. Although no formal measure was done of inter-rater reliability, it was clear by the end that it was very high.

RESULTS

The evaluation results of the program provide the primary means for assessing its effectiveness, how the program could be implemented by others, and suggestions for the future. Of the 47 teachers who participated in the

institute, 44 completed the post-institute survey. Thirty teachers taught the full curriculum in their classroom, and 23 completed the post-lesson teacher survey form. Pre- and posttest student data were submitted by 27 of the teachers, representing over 500 students. Of the other 20 teachers, approximately 10 reported teaching elements of the material, and the other 10 dropped out of the program after the summer for personal reasons or because their grade assignments changed. Among the 27 submitting posttest student data, some teachers only submitted multiple-choice answers, reporting that students were too overwhelmed with other tasks to do the short answers, especially on the baseline posttest.

Teacher Results

Summer Institute

Overall, attitudes toward the institute were very encouraging. Teachers reported that they gained a good deal from the institute, particularly in reference to geological features of the Connecticut landscape, and were looking forward to applying what they had learned to their classrooms. They particularly appreciated being able to develop relationships with other teachers, build their knowledge of the Earth sciences, and understand the ability of the Peabody Museum to act as a resource. When asked specifically about Earth science, teacher responses indicated that the institutes were successful at increasing teachers’ comfort at guiding their students to make observations and inferences about the local landscape, as well as teaching their students to analyze and interpret collected data. Open-ended questions suggested that while participants were very excited about being able to bring the new materials to their students, as well as being able to draw upon the resources and expertise of the Museum, there was some concern that time and curricular constraints would not allow them to do as much with their students as they would like. More details and background data from the post-institute survey report can be found in the supplementary material. (See Teacher Survey Results; available at: <http://dx.doi.org/10.5408/11-241s4>.)

Lesson Materials

In general, the lessons were well received, and the teachers felt that they were able to teach them without substantial difficulty. Many comments centered on having more materials and activities available for specific topics. As a result of this teacher feedback, several activities were streamlined and the final curriculum added more activities for the classroom, as teachers reported they were having difficulties taking students into the field. Oddly, although 60% said they had been inspired to change the way they taught, only a little over 20% said that the project-designed lessons were significantly different from how they taught the rest of their classes. From the accompanying comments, it appears that some teachers had difficulty completing the inquiry portions of the lessons, which could partly explain the discrepancy between the two responses. More details and background data on the teachers’ response to the lesson materials, from the post-lesson survey report, can be found in the supplementary material. (See Teacher Survey Results; available at: <http://dx.doi.org/10.5408/11-241s4>.)

TABLE II: Comparison of teacher responses on the pre-workshop survey and the post-lesson survey on comfort levels with Earth science content and teaching ($n = 21$).¹

Response	Pre-institute survey	Post-lesson survey	p value
I know a great deal about plate tectonics and its impact on the Earth	3.57	4.57	< 0.001
I know a great deal about volcanos and their impact on the Earth	3.48	4.19	0.008
I know a great deal about how rivers transform the landscape around them	3.43	4.10	0.005
I know a great deal about how different types of rock are formed	3.24	4.14	< 0.001
I know a great deal about glaciers and their impact on the Earth	3.29	4.14	< 0.001
I understand how to go about determining how landforms were created	3.29	3.95	0.016
I know a great deal about local geological features and how they were formed	3.10	3.86	0.002
I am confident that I can teach my students how to analyze and interpret geological data	3.52	4.14	0.015
I am comfortable guiding my students to make inferences about the formation of geological features in the Connecticut landscape	3.24	4.10	< 0.001
I am comfortable guiding my students to make relevant observations about geological features in the Connecticut landscape	3.24	4.10	0.001

¹Questions 5–12, $n = 558$; questions 14–16, $n = 423$.

Overall Impact on Attitudes and Skills

The teachers had been asked to respond to a series of questions about their comfort with science and teaching science at both the beginning of the institute and at the time of the post-lesson survey. The pre-institute survey data suggested that the program was likely to have impact on two specific areas, teachers’ geoscience content knowledge and their comfort in teaching specific geoscience-related skills. These were two areas that teachers rated low compared with their general comfort with science content and teaching.

As expected, responses to the post-lesson survey indicated that the program’s impact on teachers’ comfort with general science and science teaching was limited. Teachers in the program already rated their skills in this area

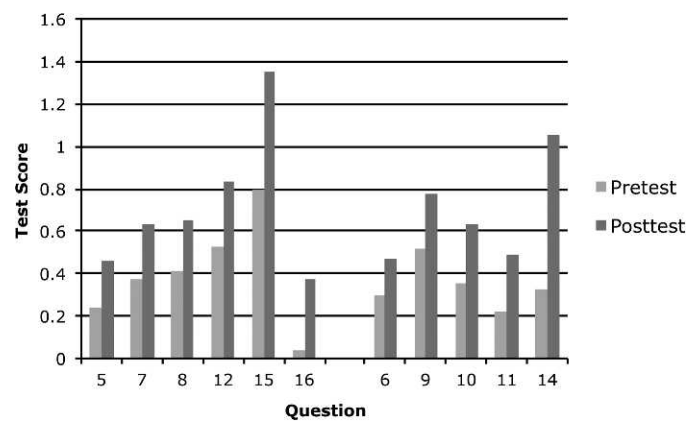


FIGURE 1: Increase in student test scores between the pre- and posttests, for all students who took both tests (questions 5–12, $n = 558$; questions 14–16, $n = 423$). Questions 5, 7, 8, 12, 15, and 16 addressed tectonic processes. Questions 6, 9–11, and 14 addressed glaciers, weathering, and erosion. The Student Assessment Instrument and the Standard Deviation Data information for each item can be found in the supplementary material (available at: <http://dx.doi.org/10.5408/11-241s3> and <http://dx.doi.org/10.5408/11-241s5>.)

very highly (see the Teacher Survey Results; available at: <http://dx.doi.org/10.5408/11-241s4>.) for supporting data). However, as Table II demonstrates, the teachers did show statistically significant increases regarding teaching their students to analyze and interpret data, and to make observations of and inferences about geological features. Similarly, the teachers were asked to indicate their level of knowledge in various areas of the Earth sciences both prior to the institute and on this survey. In every case, there was a statistically significant increase in knowledge. In particular there were very significant changes (at $p < 0.01$) in teacher responses about their knowledge of plate tectonics, rock formation, and the role of rivers and glaciers in shaping the landscape, which addressed the first outcome of the program. The teachers also responded positively to the program’s focus on developing relationships with their peers, felt it had increased their commitment to build their knowledge of Earth science, and were excited about bringing new materials into the classroom as well as being able to draw on the resources and expertise of the Peabody Museum.

Student Results

The student test results were very encouraging and strongly suggested the program had impacted their understanding and knowledge of Earth science topics. The results were analyzed according to the learning goals of the program: (1) tectonic processes and (2) glaciers, weathering, and erosion. In developing the test instrument before the full implementation of the project (so we could implement the baseline posttest the preceding academic year) it became clear that some questions were less relevant to the lessons as they were ultimately enacted. To properly measure impact and capture the full measure of student gains, it is important the test instrument is fine tuned to capture the areas that the initiative is designed to impact (Blank and de las Alas, 2009). Therefore, questions 1–4 and 13 were excluded from the analysis. Figure 1 shows the average score for the students on the pre- and posttests for the remaining questions. Multiple-choice items, questions 5–12, had a maximum score of “1,” while the short-answer questions, 14–16, had

TABLE III: Effect sizes for the difference between the pre- and posttests, for all students who took both tests.

Test question (tectonic processes)	Effect size	Test question (glaciers and erosion)	Effect size
5	0.34	6	0.28
7	0.40	9	0.44
8	0.37	10	0.43
12	0.53	11	0.45
15	0.59	14	0.77
16	0.39		

maximal point values of “2,” “2,” and “3,” respectively. Student posttests were compared to their own pretests by using paired sample Student’s *t*-tests,³ so only students who completed both tests were included in the analysis. These tests show the likelihood that the difference between the two means would occur purely because of chance, with a statistically significant difference being one with a likelihood of the difference occurring by chance of less than 5% ($p = 0.05$). In this case, there were statistically significant increases on the posttest performance (at $p < 0.001$ or $< 0.1\%$ likelihood of occurring by chance) for all questions.

Table III shows the effect sizes of the differences between the pretest and the posttest for each item and for the test scores as a whole. While significance tests show the likelihood of the difference being because of chance, effect sizes show, on an absolute scale, how large the difference is.⁴ Typically, an effect size of 0.2 is considered small, one of 0.5 medium, and one of 0.8 large (Cohen, 1992). As the table demonstrates, most of the effect sizes for individual items are near or in the medium range—the students’ improvements on those items were not remarkable by themselves—but the effect sizes for the students’ combined scores on the questions in the two areas were quite large at 1.08 (tectonic processes) and 1.13 (glaciers, weathering, and erosion), respectively.

To put this into context, a recent meta-analysis of studies on student gains as a result of teacher professional development (Blank and de las Alas, 2009) found less than 25% demonstrated an effect size of 0.4 or greater. Hattie (2009) has made the case that the average overall impact of a teacher on student learning as measured by a change in performance on a pre- and posttest would be expected to be around an effect size of 0.4 in the absence of any effective intervention. Demonstrated differences larger than this would be considered noteworthy, while smaller ones would show an intervention that is not particularly effective when compared to the general teaching practice. Compared with

Hattie’s findings on the average *total* effect size, the effect sizes found for the students’ change in overall test scores are far and above Hattie’s criteria for a successful intervention.

The students were clearly able to answer the test questions far better after the lessons were introduced, and this change in ability seems to be far larger than one might expect from an average learning experience. Still, this is not conclusive. Since the test focused on information that is expected to be part of the classrooms’ curricula, it is possible that these gains are entirely associated with the overall experience of being in the class rather than because of the benefits of the teachers’ participation in the Fellows program, and it is possible that the teachers involved in the program are sufficiently above average to show effect sizes at such high levels. The more interesting analysis, therefore, involved looking at the pre- and posttests from 2009 to 2010 in comparison with the baseline posttests given in the spring of 2009, *prior* to the teachers’ participation in the program.

The expected result if the student groups were comparable from one year to the next (as is likely in the same school) would be that (1) the baseline posttest scores would be higher than the pretest scores from the fall of 2009, indicating that the students had learned some material examined on the test in the previous year, without the teachers having the benefit of the training, and (2) if the program was effective, the posttest scores in the spring of 2010 would be higher than *both* the pretest scores from fall 2009 and the baseline posttests from spring 2009. This latter result would indicate that the students both learned the material (since they did better than on the pretests) and they learned it *better* than they had the previous year (since they scored better than the students who took the baseline posttest). This design incorporates both the idea of a pre-/posttest and a historical comparison, both of which are common in educational research but which are rarely done in concert.

Figure 2 shows the results for the 21 teachers who submitted all three types of test. It presents the scores for the total test, including short-answer questions, with a maximum possible score of “9” for tectonic processes and “6” for erosion, weathering, and glaciation. The scores were analyzed by using independent sample Student’s *t*-tests. When the tests are all taken together (which ends up giving some classes more weight than others since they have more students than others), the trend matches expectations: the pretest scores are the lowest, followed by the baseline posttest scores, and the posttest scores are highest. The differences are highly statistically significant at $p < 0.001$.

The effect sizes are also telling. For the learning goal on tectonic processes, the effect size of the difference between the pre- and the posttest is 1.06 (slightly different from the effect found for the matched groups shown above, because these data include some students who took one test and not the other), which is a large effect as noted above. The difference between the baseline-post test and the posttest has an effect size of 0.43, a medium effect showing that the students in the year with the new materials performed at a substantially higher level. Similarly the effect size for questions on glaciers and erosion is 1.06 between the pre- and posttest, and 0.37 between the baseline test and posttest, again indicating significant improvement.

³ Paired sample Student *t*-tests look at the average difference between participants’ scores on one measure and the same participants’ scores on another measure. They are a more exact means of examining the difference between the measures than would be the case if all of the pretests were to be compared to the posttests without matching. As with all *t*-tests, they provide a means of determining the *p* value (likelihood of the difference being due to chance) of the difference between the two measures.

⁴ Effect sizes are a measure of the difference between two means divided by the standard deviation of the group. For paired samples, the measure is somewhat more complicated, but in essence, it is still a means of determining a difference scaled to the standard deviation of the scores.

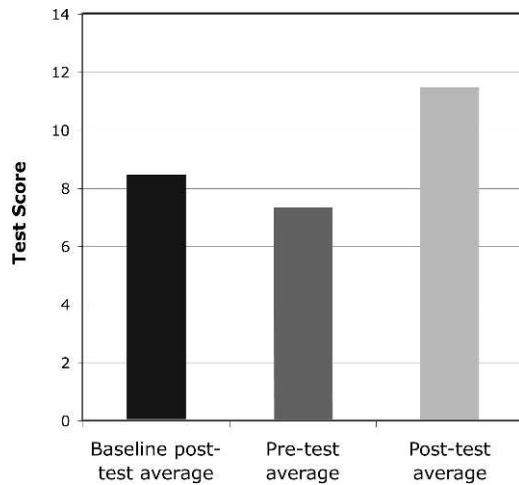


FIGURE 2: Test results for students in classes where teachers submitted the results of all three tests. The standard deviations for each item can be found in the supplemental material (See Standard Deviation Data; available at: <http://dx.doi.org/10.5408/11-241s5>.) ($n = 407, 416, 413$).

Individual classes do not always follow the expected pattern. Fewer results from individual teachers are statistically significant, in part because the low numbers of students in each class mean that differences have to be greater. In some cases, however, they tell a story about what occurred in the class that is substantially different from expectations, as shown in Figures 3 and 4. The effect sizes for these classes are shown in Table IV. For tectonic processes (Fig. 3), seven teachers followed the predicted pattern of scores (the pretest scores are lowest, followed by baseline posttests and then posttests), and six showed posttest scores that were higher than both the pretest and baseline posttest scores, but there was no significant difference between the pretests and

baseline posttests. This latter case implies that there was little significant learning of the tested content prior to the introduction of the new materials and is thus not a criticism of the new materials. However seven teachers showed no significant difference between the baseline posttests and posttests (marked with an * on Fig. 3), and one showed no real difference between any of the scores (marked with a # on Fig. 3). For glaciers, weathering, and erosion (Fig. 4), 10 teachers followed the predicted pattern of scores (the pretest scores are lowest, followed by baseline posttests and then posttests) and 3 showed posttest scores that were higher than both the pretest and baseline posttest scores, but there was no significant difference between the pretests and baseline posttests. The lower numbers for this latter case, compared to tectonic processes, could be the result of teachers not teaching tectonic processes the previous year. However 4 teachers showed no significant difference between the baseline posttests and posttests (marked with an * on Fig. 4), and 4 showed either no real difference between any of the scores or baseline posttest scores that were higher than posttest scores (both cases marked with an * and # on Fig. 4). Overall the results also indicated that 7 teachers showed the predicted pattern for *both* learning goals, and 3 more showed posttest scores that were higher than pretest and baseline test, but there was no significant difference between the pretest and baseline test.

There are a number of possible reasons for why some of the classrooms might not have shown patterns indicative of an impact of the new materials beyond the obvious explanation that there was no such impact. In some cases, this year’s class performed better on the pretest than last year’s did on the baseline posttest. Given this, it is impossible to say whether the improvements in the 2009–2010 class year are because of the materials, because the two classes were so different in their ability to answer test items correctly. This could also explain the two cases for the glaciation scores, where the baseline posttests were higher than the posttests. It might well be that other classes showed

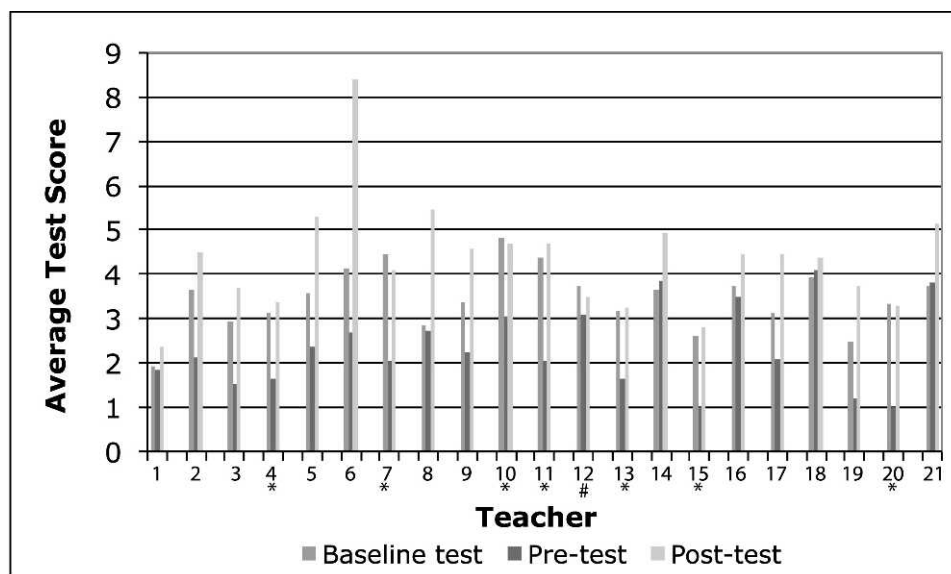


FIGURE 3: Test results for tectonic processes, from classes that completed all three tests, separated by teacher (n between 13 and 29). Maximum possible score was 9. Results that do not conform to our expectations are marked with an * and #, and are further explained in the text.

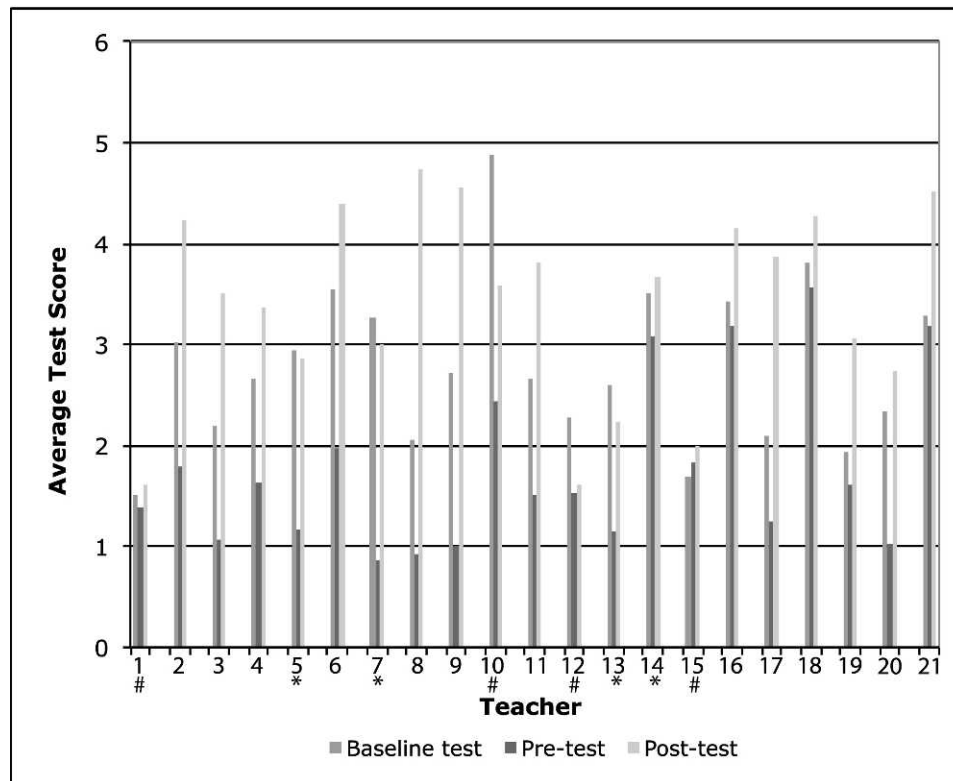


FIGURE 4: Test results for erosion, weathering, and glaciation from classes that completed all three tests, separated by teacher (n between 13 and 29). Maximum possible score was 6. Results that do not conform to our expectations are marked with an * and #, and are further explained in the text.

similar differences between the different years to make the individual comparisons insignificant. Another possible explanation for a lack of positive findings in those classrooms is that certain teachers might not have used the materials in the way or to the extent that was intended because of, for example, pressures from within their school, the needs of their students, or the constraints of the curriculum.

In summary, 16 of the 21 teachers in the analysis showed results that indicate an increase in performance in at least one learning goal for students that was associated with the introduction of the new materials. Of these, 10 showed positive results for both learning goals. This represents a positive result for the large majority of both teachers and students.

CONCLUSION

Museum-based teacher professional development programs provide an opportunity to engage teachers in up-to-date science content, allow them to interact with science-rich professional communities, enable them to improve their pedagogical skills, and help them develop resources and strategies that engage and sustain student interest (Finkelstein, 2006). The Peabody Fellows program succeeded in achieving both cognitive and affective teacher change, and improving teachers' geoscience knowledge and their comfort with specific geoscience teaching strategies. In terms of affective change, teachers felt more comfort in assisting students to analyze and interpret data, and to make observations of and inferences about geological features,

the first/core learning goal. They also indicated increased content knowledge in both areas that were a focus of the program, i.e., destructive and constructive geological forces.

This study also demonstrates that such programs also can have significant impact on the students themselves. The Peabody Fellows program has successfully improved student outcomes by enhancing their performance on both content and process questions associated with the Earth sciences. The data show a highly statistically significant difference (at a p value of less than 0.001) between students' performance on the baseline posttest and the posttest for 75% of the teachers who completed all evaluation activities, as well as effect sizes that demonstrate that these are substantial differences. Results were similar for both learning goals of the program. Students improved their understanding of both constructive and destructive geological forces within the context of their own environment.

The study design did have some limitations that need to be considered when drawing conclusions. The data on teacher outcomes is based on self-report surveys rather than independent testing or observation. However, it is generally accepted that such self-report data does provide an accurate predictor of outcomes (Banilower *et al.*, 2007). There is a potential selection bias in the data, as only 60% of teachers who participated in the summer institute delivered complete evaluation results, and those individuals might not be completely representative of the entire group. Finally, the design assumes that students from one year are a suitable control group for students the following year. Since the students are in the same school, same grade, and being taught by the same teacher this seems reasonable, although,

TABLE IV: Effect sizes for the three test comparisons for each teacher.

Teacher	Effect sizes for erosion, weathering, and glaciation questions			Effect sizes for tectonic processes questions		
	Baseline vs. pretest	Baseline vs. posttest	Baseline vs. pretest	Baseline vs. posttest	Baseline vs. pretest	Baseline vs. posttest
1	0.10	0.08	0.22	0.06	0.37	0.40
2	0.89	0.86	1.84	1.11	0.58	1.70
3	1.00	1.03	2.17	1.02	0.55	1.74
4	1.01	0.47	1.26	1.07	0.14	1.08
5	1.40	−0.05	1.39	0.70	0.94	1.60
6	1.27	0.68	2.18	1.24	4.30	5.56
7	2.20	−0.23	1.88	1.67	−0.28	1.51
8	1.19	3.27	4.15	0.09	1.97	2.08
9	1.92	1.66	3.25	0.93	0.82	1.61
10	2.09	−1.05	0.86	1.48	−0.08	1.13
11	1.03	1.03	1.88	1.78	0.24	1.88
12	0.48	−0.48	0.04	0.32	−0.13	0.18
13	1.30	−0.26	0.86	0.99	0.03	0.87
14	0.23	0.10	0.41	−0.16	1.12	0.77
15	−0.12	0.28	0.15	1.37	0.14	1.64
16	0.16	0.62	0.74	0.21	0.52	0.66
17	0.73	1.39	2.35	0.92	1.04	2.10
18	0.21	0.45	0.61	−0.11	0.36	0.22
19	0.33	1.25	1.49	0.81	0.67	1.55
20	0.86	0.26	1.26	1.62	−0.03	1.29
21	0.05	0.77	0.80	−0.06	1.35	1.33

some of our results indicate this might not be true in all individual classrooms. There could be a cohort effect of students, wherein the students in the year after the institute were academically stronger than those in the previous year, but the change in student composition would have to be correlated with the features of professional development across the different schools, which seems unlikely.

As noted before, the connection between teacher professional development and improved student achievement is intuitive but can be difficult to demonstrate (Yoon et al., 2007). Our evaluation protocol differed from standard practice for many professional development programs in quantifying student outcomes, as well as collecting self-reported data from the teachers themselves. The outcome measures were closely aligned with the program's goals for Earth science learning, with multiple-choice questions testing students' geoscience content knowledge, and the short-answer questions focused more on the inquiry-based science that the program promoted. Using these measures, our results affirm that positive proximal outcomes for the teachers mirror the ultimate outcomes for students. In addition, by including baseline tests from the teachers' previous year's classes, we were able to clearly relate these changes to participation in the Peabody Fellows program, which is often not possible when only pre- and posttests are used.

These results suggest that sustained, ISI-based professional development has positive impacts on both teacher

and student knowledge and understanding. As posited by Bevan et al. (2010), in their discussion of how ISIs support and complement K–12 education, our program supports the belief that formal–informal collaborations foster better learning outcomes. It would be interesting to more fully explore, perhaps through extensive interviewing, teachers' thoughts on the value of different aspects of the program compared with their other professional development experiences in Earth science. The literature and our experience would suggest that certain hallmarks of this type of professional development (place-based approach, use of inquiry and hands-on techniques, building a learning community facilitated by museum educators) would be picked out. However there needs to be more research to establish this.

Teacher surveys also identified a direction for future work on providing effective field trip experiences for students, including the need for field sites and resources that are close to the school. Field trips are fundamental to true Earth science literacy (see, for example, Kastens et al., 2009) but it is increasingly hard for teachers to bring students out of the classroom and the participants in the Peabody Fellows program are no exception. Teachers cited the institute field trips as the highlight of their personal learning experience: "The field trip was informative both in terms of introducing me to some excellent sites, as well as in inspiring me to want to better 'paint the picture' of Connecticut geological history to my students." But they still felt the need

for more support in this area, to improve their confidence in leading field trips, and to address the practical challenges of taking students out of the classroom. ISIs are in an excellent position to facilitate this. Not only are they often field trip sites themselves (especially those institutions with outdoor areas), but they also have knowledgeable staff and other resources to help identify appropriate field sites and related background information. This is an area where smaller ISIs can make a significant contribution to teacher professional development within their communities, as such resources must be generated locally rather than regionally or nationally.

Another area that would be interesting to explore in further research would be the effect of how teachers implement the materials. By observing how teachers used the new materials in their classes, it would be possible to determine if there are any differences that could explain why their use seemed to be more effective in some cases than in others. It would also be beneficial to investigate the longer-term impacts of the program on the participants' ongoing geoscience teaching activities. The Museum, through its ongoing networking activities, will keep in contact with the teachers in order to facilitate such a study.

Given the limited number of teachers who take Earth science courses as part of their professional preparation, effective in-service teacher professional development is vital. As McNeal (2010) describes in reference to the geosciences "gap" in K–12 education, "the exposure and knowledge a student receives about the geosciences is often dependent on their teacher's exposure and comfort with the field. This teacher–student phenomenon leads to a vicious cycle, a detrimental feedback loop of limited student geoscience experiences, frequent manifestations of commonly held misconceptions...and general devaluing of the field by the public and decision makers." The Peabody Fellows is a successful model for a program based in an ISI that provides the type of intensive, hands-on, classroom-focused experience to successfully address this challenge.

ACKNOWLEDGMENTS

The Peabody Fellows in Earth Science program was supported by a grant from the National Science Foundation, Directorate of Geosciences, GEO 0807864, under the direction of Jay J. Ague, Jane Pickering, and David M. Heiser. The authors thank the anonymous reviewers for their comments and suggestions.

REFERENCES

- Alberts, B. 2009. Making a science of education. *Science*, 323:15.
- American Geological Institute. 2007. Pulse of Earth science: national status of K–12 Earth science education, state by state, 2007. Available at <http://www.agiweb.org/education/statusreports/2007/index.html> (accessed 23 June 2012).
- Astor-Jack, T., McCallie, E., and Balcerzak, P. 2007. Academic and informal science education practitioner views about professional development in science education. *Science Education*, DOI: 10.1002/sce.20205.
- Bamberger, Y., and Tal, T. 2007. Learning in a personal context: Levels of choice in a free-choice learning environment in science and natural history museums. *Science Education*, 91:75–95.
- Banilower, E.R., Heck, D.J. and Weiss, I.R. 2007. Can professional development 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.
- Barstow, D., and Geary, E. 2001. Blueprint for change: Report from the National Conference on the Revolution in Earth and Space Science Education. Cambridge, UK: TERC. 100 p.
- Bevan, B. 2007. When schools meet museums: Finding a third way. *ASTC Dimensions*, May:3–7.
- Bevan, B., Dillon, J., Hein, G.E., Macdonald, M., Michalchik, V., Miller, D., Root, D., Rudder, L., Xanthoudaki, M., and Yoon, S., 2010. Making science matter: Collaborations between informal science education organizations and schools. A CAISE Inquiry Group report. Washington, DC: Center for Advancement of Informal Science Education. Available at <http://caise.insi.org/uploads/docs/MakingScienceMatter.pdf> (accessed 23 June 2012).
- Birnbaum, S. 2004. Overcoming the limitations of an urban setting through field-based Earth systems inquiry. *Journal of Geoscience Education*, 52:407–10.
- Blank, R.K., and de las Alas, N. 2009. Effects of teacher professional development on gains in student achievement. Washington DC: Council of Chief State School Officers. 62 p.
- Cohen, J. 1992. A power primer. *Psychological Bulletin*, 112:155–159.
- Corcoran, S.P., and Evans, W.N. 2008. The role of inequality in teacher quality. In Magnuson, K., and Waldfogel, J., eds., *Steady gains and stalled progress: Inequality and the black-white test score gap*. New York: Russell Sage Foundation, p. 212–249.
- Council of Chief State School Officers. 2007. State indicators of science and mathematics education 2007. Available at: <http://programs.ccsso.org/content/pdfs/SM%2007%20report%20part%201.pdf> (accessed 23 June 2012).
- Czujko, R., Ivie, R., and Stith, J.H. 2008. Untapped talent: The African-American presence in physics and the geosciences. Melville, NY: American Institute of Physics. 22 p.
- Desimone, L.M., Porter, A.C., Garet, M.S., Yoon, K.S., and Birman, B.F. 2002. Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, 24:81–112.
- Finkelstein, D. 2006. Science museums as resources for teachers: An exploratory study on what teachers believe. San Francisco, CA: Center for Informal Learning and Schools. Available at: <http://qt.exploratorium.edu/cils/resource.php?resourceID=301> (accessed 23 June 2012).
- Griffiths, P., and Cahill, M. 2009. The opportunity equation: Transforming mathematics and science education for citizenship and the global economy. New York: Carnegie Corporation of New York and Institute for Advanced Study. 63 p.
- Harnik, P.G., and Ross, R.R. 2004. Models of inquiry-based science outreach to urban schools. *Journal of Geoscience Education*, 52:420–428.
- Hattie, J. 2009. Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Oxford, UK: Routledge. 378 p.
- Howey, K., Brukardt, M.J., and Warren, B., eds., 2008. Putting a stop to dropouts. Access and success in urban schools: pre-school through college: Introduction. Cincinnati, OH: Jointly published by the Coalition of Urban Serving Universities, the Urban Educator Corps and the University of Cincinnati, p. v–xv.
- Huntoon, J., Peach, C., and Hopkins, J. 2005. Geoscience education and diversity: vision for the future and strategies for success: Report of the second Geoscience Education Working Group, September 2005. Washington DC: National Science Foundation. 24 p.
- Kastens, K.A., Manduca, C.A., Cervato, C., Frodeman, R., Goodwin, C., Liben, L.S., Mogk, D.W., Spangler, T.C., Stillings, N.A., and Titus, S. 2009. How geoscientists think and learn. *EOS*, 90:265–272.

- Kean, W., and Enochs, L. 2001. Urban field geology for K–8 teachers. *Journal of Geoscience Education*, 49:358–363.
- King, C. 2000. The Earth's mantle is solid: Teachers' misconceptions about the Earth and plate tectonics. *School Science Review*, 82:57–64.
- King, C. 2008. Geoscience education: An overview. *Studies in Science Education*, 44:187–222.
- Lankford, H., Loeb, S., and Wyckoff, J. 2002. Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24:37–62.
- Macdonald, M., Sloan, H., Miele, E., Powell, W., Silvernail, D., Kinzler, R., Hong, J., and Simon, C. 2008. Improving urban Earth science education: The TRUST model. *Journal of Geoscience Education*, 56:269–279.
- McNeal, K.S. 2010. Editorial: The geosciences gap in K–12 education. *Journal of Geoscience Education*, 58:197.
- Mervis, J. 2009. NSF rethinks its digital library. *Science*, 323:54–58.
- National Academy of Sciences. 2010. Preparing teachers: Building evidence for sound policy. Washington DC: National Academies Press. 324 p.
- National Academy of Sciences, Committee on Prospering in the Global Economy of the 21st Century. 2007. Rising above the gathering storm: Energizing and employing America for a brighter economic future. Washington DC: National Academies Press. 592 p.
- National Research Council. 2009. Learning science in informal environments. Washington DC: National Academies Press. 336 p.
- National Science Board. 2010. Science and engineering indicators. Available at <http://www.nsf.gov/statistics/seind10/start.htm> (accessed 23 June 2012).
- National Science Foundation. 2009. Geo Vision report. Available at <http://www.nsf.gov/geo/acgeo/geovision/start.jsp> (accessed 23 June 2012).
- Penuel, W.R., Fishman, B.J., Yamaguchi, R., and Gallagher, L.P. 2007. What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, 44:921–958.
- Penuel, W.R., and Gallagher, L.P. 2009. Preparing teachers to design instruction for deep understanding in middle school Earth science. *The Journal of the Learning Sciences*, 18:461–508.
- Penuel, W.R., McWilliams, H., McAuliffe, C., Benbow, A.E., Mably, C., and Hayden, M.M. 2009. Teaching for understanding in Earth science: Comparing impacts on planning and instruction in three professional development designs for middle school science teachers. *Journal of Science Teacher Education*, 20:415–436.
- Philips, M., Finkelstein, D., and Wever-Freichts, S. 2007. School site to museum floor: How informal science institutions work with schools. *International Journal of Science Education*, 29:1489–1507.
- President's Council of Advisors on Science and Technology. 2010. Prepare and inspire: K–12 education in science, technology, engineering and math (STEM) for America's future. Washington DC: Executive Office of the President of the United States. 108 p.
- Science Education Resource Center. 2003. Developing the Earth science teacher workforce: The role of geoscience departments and introductory courses. Available at <http://serc.carleton.edu/teacherprep/index.html> (accessed 23 June 2012).
- Science Education Resource Center. 2009. How to engage and support urban students. Available at <http://serc.carleton.edu/introgeo/urban/how.html> (accessed 23 June 2012).
- Scotchmoor, J., Marlino, M.R., and Low, R. 2005. Making a broader impact: Geoscience Education, Public Outreach, and Criterion 2: Report of a workshop held at University of California Museum of Paleontology, Berkeley, CA, 11–13 May 2005. Washington DC: National Science Foundation.
- Tomás Rivera Policy Institute. 2008. STEM professions: Opportunities and challenges for Latinos in science, technology, engineering, and mathematics. Los Angeles, CA: Tomás Rivera Policy Institute. Available at http://www.trpi.org/PDFs/STEM%20Lit_Final.pdf (accessed 23 June 2012).
- Ward, R.B., Sadler, P.M., and Shapiro, I.I. 2008. Learning physical science through astronomy activities: A comparison between constructivist and traditional approaches in grades 3–6. *Astronomy Education Review*, 6:1–16.
- Yoon, K.S., Duncan, T., Lee, S., Scarloss, B., and Shapley, K. 2007. Reviewing the evidence on how teacher professional development affects student achievement. Available at <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=REL2007033> (accessed 23 June 2012).