A Science Lesson Plan Analysis Instrument for Formative and Summative Program Evaluation of a Teacher Education Program

CHRISTINA L. JACOBS
Department of Chemistry, Penn Science Teacher Institute, University of Pennsylvania, Philadelphia, PA 19104, USA

SONYA N. MARTIN
School of Education, Drexel University, Philadelphia, PA 19104, USA

TRACEY C. OTIENO
Department of Chemistry, Penn Science Teacher Institute, University of Pennsylvania, Philadelphia, PA 19104, USA

Received 29 June 2007; revised 29 December 2007, 25 January 2008; accepted 30 January 2008

DOI 10.1002/sce.20277
Published online 22 April 2008 in Wiley InterScience (www.interscience.wiley.com).

ABSTRACT: In evaluating the success of teacher development programs, valid and scalable measures of teaching practice are needed. We have developed and validated the Science Lesson Plan Analysis Instrument (SLPAI) for quantitative evaluation of teacher-generated multiday lesson plans. This paper presents the SLPAI as a complement to surveys and classroom observation, and demonstrates its use in 2 pilot studies. The SLPAI was used

Correspondence to: Christina L. Jacobs, Department of Science, Baldwin School, 701 Montgomery Avenue, Bryn Mawr, PA 19010; e-mail: jacobscl@sas.upenn.edu
Present address of Tracey C. Otieno: Department of Science, Germantown Friends School, Philadelphia, PA.
Contract grant sponsor: National Science Foundation.
Contract grant number: Math-Science Partnership program (EHR 0412404).

© 2008 Wiley Periodicals, Inc.
formatively to measure the teaching practices of incoming program cohorts and tailor program instruction. It was also used to track changes in teaching practice and pedagogical knowledge of participants over time, providing summative evidence of program effectiveness. © 2008 Wiley Periodicals, Inc. Sci Ed 92:1096–1126, 2008

INTRODUCTION

Recent years have seen a surge of public interest in the training and subject matter competency of secondary mathematics and science teachers; this interest has resulted in congressional hearings and high-profile publications such as Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future (National Research Council, 2006). Politicians, business leaders, and institutes of higher education have taken notice of the crucial link between the training of qualified secondary teachers and the production of a well-prepared college student population and workforce. The Math-Science Partnership (MSP) granting program at the National Science Foundation (NSF) is a direct outgrowth of this movement. Universities, school districts, and other partners have been awarded large sums of money on the basis of the assumption that the activities of such partnerships will result in a higher quality teaching force, which will, in turn, result in better prepared students.

However, to test these assumptions and validate the expenditures of the MSP program, MSP researchers and evaluators must detect and measure changes in teachers and their students and be able to attribute these changes to the activities of the MSP. For example, the institutional context for this study, referred to herein as “the Institute,” aims to “increase the content knowledge of science teachers, and change the teaching and learning methodologies used in secondary science classrooms to research-based promising pedagogical practices” (PennSTI, n.d.). These two proximal goals are hypothesized to support “the overriding goal [which] is to increase the interest in and learning of science by all students” (PennSTI). The Institute has been touted as a model program for other universities that wish to strengthen their contribution to teacher preparation (National Research Council, 2006). However, at this early stage of program implementation, we cannot offer the MSP or higher education communities a great deal in the way of evidence that our approach is better able to support students’ learning of, and interest in, science than other established teacher education programs in many places around the country. Each individual MSP grantee, as well as the research projects that are evaluating the MSP program as a whole, must continue to collect meaningful data and analyze its findings to test and hopefully support the basic assumptions of the MSP funding program.

A problem we, and presumably other MSP-funded programs face, is how to evaluate our progress toward these goals. A commonly used approach relies on standardized and usually decontextualized tasks such as surveys, tests, and other assessments for teachers, and student’s beliefs and knowledge, often used in a pre–post comparison fashion. This approach can produce large data sets in order to detect and measure change due to an intervention; however, such tasks usually occur outside the normal practices of a classroom teacher and his or her students, and therefore may not be reliable predictors of a teacher’s or a student’s actions in context. On the other hand, truly authentic assessment or evaluation depends so heavily on context that any sense of control of extraneous factors becomes impossible. Given these seemingly incompatible goals of inference of causality and authenticity, program evaluation appears at first glance to be an impossible task.

Obviously, though, program evaluation is alive and well. Mixed methods research and evaluation have gained attention as a way to address these issues by using a combination of evaluation techniques and instruments. For example, the evaluation plan for the Institute
includes yearly student and teacher surveys of practices and attitudes, student achievement tests, annual classroom observation, and interviews of a subset of teachers using both a quantitative evaluation protocol and qualitative field notes. Upon examining this set of tools, however, we found a weakness in our ability to evaluate the quality of teaching. Although classroom observations are useful in this regard, they cannot be implemented at a large enough scale to get a clear idea of even a subset of teachers’ classroom practices. Another, more scalable and broader lens with which to measure teaching behaviors and the beliefs they evince was needed. This need led to the development of the Science Lesson Plan Analysis Instrument (SLPAI),\(^1\) which provides several benefits as an addition to the comprehensive evaluation plan. It allows for evaluation of longer “chunks” of planned instruction, allowing insight into the teachers’ decisions about sequence of and relationships between activities and topics as well as their assessment strategies, neither of which are commonly evident when observing a single class period (Appendix). We do not claim that lesson plan analysis is a suitable replacement for either in-depth site visits coupled with classroom observation or a large-scale administration of pre–post surveys and tests. In an ideal situation, all of these protocols would be used in combination, in a manner responsive to the size and staffing of the project, to provide the most reliable and useful information.

This paper describes the SLPAI and its use in evaluating the work of the Institute’s teacher participants before and during their involvement in the program. We discuss how results from this instrument fit with information gathered using other established protocols, as well as how it contributes unique and useful data about our participants and our program.

**BACKGROUND**

Previous investigations of changing teacher practice as a result of professional development have utilized several techniques to measure teacher practice, either directly or indirectly. Self-reported questionnaires have the benefit of ease of administration and the possibility of large sample size aiding statistical analysis. Self-reported questionnaires, triangulated with data from student surveys, were successfully utilized to monitor the effects of a statewide systemic initiative on teachers’ practices and students’ attitudes (Scantlebury, Boone, Butler-Kahle, & Fraser, 2001). However, research clearly indicates that how a teacher understands and implements a reform, not merely its presence in his or her practice, influences the effectiveness of that reform in the classroom (Brown & Campione, 1996). One limitation of using survey data alone, therefore, is that the researcher cannot distinguish between high and low implementation quality of a strategy based only on a teacher’s assertion that the strategy is utilized. Several studies have indicated that the degree to which self-reported data is valid is limited to the teacher’s ability to accurately describe his or her instructional practices. Andrew Porter addressed the validity of data produced through tools for measuring content of instruction in his 2002 Presidential Address at the annual meeting of the American Educational Research Association. Porter raised issue with self-reported instruments, specifically he noted, “Teachers may report what they think is appropriate, even if what they report does not accurately depict their own practice” (2002, p. 8). Additional studies have shown that survey respondents may not be clear on the terminology used in the survey instrument (Schepenzeel & Saris, 1997; Sudman, Bradburn, & Schwarz, 1996) and may inadvertently misrepresent their teaching practices based on the misunderstanding. Porter adds that even when “teachers may understand what content is wanted and believe they are teaching that content, they may in fact not [be]” (2002, p. 8).

\(^1\) For ease of reading, as well as to reduce the number of acronyms in the paper, the Science Lesson Plan Analysis Instrument is referred to as both SLPAI and the instrument.
For example, one survey item used in the 2001 study cited above asked teachers to rate the extent to which their students “use data to justify responses to questions” on a scale from “almost never” to “very often.” Assuming the item is understood by the teacher in the intended way, such “justification” of students’ conclusions could reflect greatly varying expectations of students. Triangulation with a similar question for students, “In this class my teacher asks me to give reasons for my answers,” is of little use in this case because the student item does not refer to the use of data and could be interpreted differently by students, depending on their past experiences with writing explanations. From the students’ point of view, a “reason” could simply be “because of gravity” or “because I learned it last year.”

A second limitation of self-reported survey data is that a teacher with a higher level of awareness of teaching reforms, and the philosophical and sociological arguments that underlie them, would be expected to have a better understanding of the survey items’ intent. This would likely result in more accurate self-reported data; such teachers would be less likely to rate themselves positively (or negatively) for the wrong reasons. Teachers with little or no background in educational theory, on the other hand, would be vulnerable to misreporting. Using the previous survey item as an example, a teacher may not be aware of how to structure student inquiries to allow for students to support their conclusions effectively using evidence. Instead, they might have students treat textbook information or their lecture notes as “evidence.” Since the way a teacher interprets this and similar items depends on his or her own level of knowledge about the topic and/or effective strategies for teaching the topic, survey results alone may not give an accurate picture of the extent of reform-oriented teaching in such a teacher’s classroom. Alternatively, a teacher with a more sophisticated understanding of teaching reforms and theory is likely to know what evaluators are looking for, and might consciously or subconsciously give the “best answers,” thus skewing the data.

Direct observation by a trained evaluator using an instrument, such as the Reformed Teaching Observation Protocol\(^2\) (RTOP; Piburn et al., 2002; Sawada, Piburn, Judson, et al., 2002; Sawada, Piburn, Turley, et al., 2002) or the Approaches to Teaching Inventory\(^3\) (ATI; Trigwell & Prosser, 2004), provides an excellent solution to these issues. The primary drawback of this approach is that classroom evaluation of teachers is resource intense, and therefore not highly scalable for evaluative purposes in large programs. In our specific case, our program will have at maximum 120 current teacher participants (6 cohorts of 20 each) under evaluation during an academic year. Given current staffing levels and the geographic area over which our teachers are spread, we are able to observe and use RTOP to evaluate only about one third of the participants from each cohort. Furthermore, since each teacher selected for observation can be visited only once per year, observation data cannot provide a truly representative picture of teaching practices for any individual teacher or the cohort as a whole. For example, a hypothetical lesson sequence might include a day of mainly lecture

\(^{2}\) The RTOP (Sawada et al., 2002) was created by the Evaluation Facilitation Group of the Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT) as a valid and reliable observation instrument designed to provide a standardized means for detecting the degree to which K-20 classroom instruction in mathematics or science can be characterized as reformed teaching. The instrument items are based on reform documents in science and mathematics, such as Project 2061: Science for All Americans, National Science Education Standards, and Principles and Standards for School Mathematics. Items measured are organized into five categories, including (1) lesson design and implementation; (2) content: propositional pedagogical knowledge; (3) content: procedural pedagogic knowledge; (4) classroom culture: communicative interactions; and (5) classroom culture: student–teacher relationships.

\(^{3}\) The ATI (Trigwell, Prosser, & Prosser, 1999) was designed to measure the ways in which teachers approach their teaching in relation to their conceptions of teaching and learning. In addition, the ATI measures teaching in context by exploring the relations between teachers’ approaches to teaching and students’ approaches to learning.
as new topics are introduced, followed by a laboratory experience on the second day, and a discussion of the laboratory results on Day 3. Depending on which day of the sequence was observed, a teacher could receive very different ratings of classroom practice. We have found this to be true in our own program as well; one Institute instructor was observed teaching the same students on 3 different days during a semester, and the resulting RTOP scores were quite varied (63, 40, and 77 of 100) because of the daily activity structure.

What is needed to address these problems is an evaluation method with a closer link to actual teaching practices than survey results, but that allows evaluation of all teachers in the program, and provides a longer time frame than a single day of instruction. We believe that the SLPAI fulfills these needs. Using the SLPAI in concert with teacher and student questionnaires for all participants, in addition to RTOP for a subset, therefore allows us to develop a more complete and accurate picture of the effects our programs have in the classrooms of our teacher participants.

The lesson plans generally required of teachers by school administrators for evaluation purposes are typically brief outlines emphasizing procedural aspects of lesson planning (Danielson & McNeal, 2000; Halverson, Kelley, & Kimball, 2004; Kyle, 1980). Gail McCutcheon (1980) referred to these lessons as “planbook planning,” which she characterized as including information such as references to page numbers to be covered in the textbook, problems to be assigned for homework, and lists of standards or objectives to be covered during the lesson. Several studies have indicated that administrators collect these teacher planbooks either weekly or monthly with the intent of monitoring classroom curriculum implementation (Berger, 1990; Fleck, 2005; McCutcheon, 1980), as well as to ensure that plans are available for substitutes should the teacher be absent. In contrast with these planbook lessons, our conception of the “lesson plan” focuses on the pedagogical knowledge and decisions of the teacher as embodied in lesson planning documents that are more fully developed by teachers for their day-to-day teaching practice. Various studies on teacher planning (Brown, 1993, 1998; Kagan & Tippins, 1992; Yinger, 1980) describe teachers’ daily plans as involving more attention to content goals, knowledge, and sequencing, as well as activity procedure, implementation, and assessment.

A few examples of lesson plan evaluation are present in education literature, both general and science specific. One such instrument, the Science Lesson Plan Rating Instrument\(^4\) (SLPRI; Hacker & Sova, 1998), focused on procedural aspects of lesson planning such as identification of resources used, timing estimates for activities, and inclusion of lesson objectives on the course syllabus. Of the 34 equally weighted items on this instrument, 15 of them address substantive issues about how science is taught. We drew on these categories in developing the SLPAI; for example, “Have key questions for students been identified?” was folded into our items dealing with classroom discourse and goal orientation, and “Are the selected activities at the correct level of difficulty for the class?” is consonant with our “Content presentation” item.

Regardless of the measurement instrument used, its alignment with the reform agenda or teacher education curriculum being studied is vital. To this end, we have utilized contemporary educational research and reform documents that underpin the mission of the Institute to inform development of our instrument. We were influenced by the description of learner-, knowledge-, and assessment-centered learning in *How People Learn* (Bransford, Brown, & Cocking, 1999). Learner-centered instruction requires an understanding by the teacher

---

\(^4\) The SLPRI (Hacker & Sova, 1998) was designed to evaluate the efficacy of a computer-mediated courseware module created to develop the lesson planning skills of preservice science teachers. Categories included clarity of objectives, use of resources, alignment of instructional strategies to content and objectives, and continuity.
TABLE 1
National Science Education Teaching Standards (A–E) With Selected Substandards

A. Teachers plan an inquiry-based science program for their students
   Select science content and curricula to meet the interests, knowledge, understanding, abilities, and experiences of students
   Support the development of student understanding and nurture a community of science learners
   Work with colleagues within and across disciplines and grade levels

B. Teachers guide and facilitate science learning
   Focus and support inquiries while interacting with students
   Orchestrate discourse among students about scientific ideas
   Recognize and respond to student diversity and encourage all students to participate fully in science learning
   Encourage and model the skills of scientific inquiry, as well as the curiosity, openness to new ideas and data, and skepticism that characterize science

C. Teachers engage in ongoing assessment of their teaching and of students’ learning
   Use multiple methods and systematically gather data about student understanding and ability
   Analyze assessment data to guide teaching
   Guide students in self-assessment

D. Teachers design and manage learning environments that provide students with the time, space, and resources needed for learning science
   Structure the time available so that students are able to engage in extended investigations
   Make the available science tools, materials, media, and technological resources accessible to students
   Engage students in designing the learning environment

E. Teachers develop communities of science learners that reflect the intellectual rigor of scientific inquiry and the attitudes and social values conducive to science learning.
   Display and demand respect for the diverse ideas, skills, and experiences of all students
   Nurture collaboration among students
   Structure and facilitate ongoing discussion based on a shared understanding of rules of scientific discourse
   Model the skills, attitudes, and values of scientific inquiry


of how the knowledge and beliefs that students bring to the classroom will affect how they interpret new information and construct new knowledge (Bransford et al.). The concept of knowledge-centeredness, which emphasizes the need for connected and conceptual understanding over “acquisition of disconnected sets of facts and skills” (Bransford et al., p. 153), is central to several instrument items. Finally, the SLPAI considers the assessment practices of the lesson plan in terms of both summative and formative assessments, and requires evaluation of conceptual understanding over factual knowledge as reflected in the goals of the lesson. These features correspond with the description of “assessment to support learning” recommended in Understanding by Design (Wiggins & McTighe, 2001).

The instrument is aligned with the Science Teaching Standards (A–E) from the National Science Education Standards (National Research Council, 1996). These standards, with relevant sections summarized in Table 1, constitute a clear call for inquiry-based science instruction and illuminate the features that identify such teaching practices.

Science Education
Brown and Campione’s review (1996) of the critical features of powerful learning environments described the basic components of such an environment. Their model, Fostering Communities of Learners, requires that students (1) undertake research, (2) share the resulting information with each other, (3) use shared information to perform a meaningful task, and (4) engage in self-conscious reflection during the learning process, all with the aim to (5) understand deep disciplinary content. These five aspects of a learning environment were each central in development and refinement of the instrument.

As previously noted, the approach to curriculum and lesson design as measured by the SLPAI was also guided by Understanding by Design (Wiggins & McTighe, 2001), specifically the emphasis on preassessment, goal-oriented instruction, and alternative assessment. Finally, the SLPAI “Nature of science” item was developed out of the extensive literature on teachers’ beliefs and instructional practices around the nature of science (Brickhouse, 1990; Chinn & Malhotra, 2002; Crowther, Lederman, & Lederman, 2005). The emphasis of these sources on the evidence-based and tentative and social nature of scientific knowledge development is in contrast to the objectivist and fact-focused science teaching beliefs and practices we noted in preliminary observations and survey response analyses. Many of these sources also informed the development of the other instruments utilized in our evaluation, including the observation protocol and surveys, making this instrument a theoretical complement to these methods.

METHODS

Setting and Participants

Study participants were in-service teachers enrolled in an MSP Institute program funded by the NSF. The Institute is composed of two master’s degree–granting programs: one for high school chemistry teachers (HSChem) and the other for middle-grade science teachers (MSSci). Each program spans three summers and the two intervening academic years, and requires completion of eight science content courses and two science education courses over this 30-month period. The science content courses have been specifically designed by faculty members for these two programs and enroll only MSP participants. Content faculty, who engage in MSP-sponsored professional development concerning pedagogical issues, such as integrating a cross-discipline approach and using inquiry methods, teach these courses. The science education courses are also specific to the MSP, and they require participants to engage in classroom-based research. In addition, these courses focus on improving the science pedagogy and leadership skills of participants. As a whole, these courses are meant to facilitate the Institute’s programmatic goals: to increase the content knowledge of science teachers, increase their use of research-based promising pedagogical practices, and enable participants to become science education teacher leaders (PennSTI, n.d.).

5 The HSChem participants are referred to as high school teachers and the MSSci participants are referred to as middle-school teachers in an attempt to reduce the number of acronyms used in this paper. However, readers are reminded that there are teacher participants in each of these programs who may teach students outside this descriptor. For example, some of the MSSci participants teach students in elementary grades and some of the HSChem participants teach students in the middle grades. In addition, not all high school teachers exclusively teach chemistry courses; some also teach mathematics and other science disciplines.
All teacher participants from the 2005 incoming Institute cohorts who were enrolled from partner schools and complied with lesson plan submission guidelines were included in this study. In addition, data from participants selected for classroom observation from the 2006 incoming cohorts were utilized in the RTOP validation portion of the study only. The analyses presented here are therefore based on 20 MSSci and 8 HSChem teacher participants in the 2005 cohort and 7 MSSci and 7 HSChem teacher participants in the 2006 cohort. Only teachers from partnered school districts were included in this study because of their availability for classroom visitation as well as program staffing limitations that precluded data collection beyond the scope required for our external evaluation. Some participants who were otherwise qualified were omitted from certain analyses to allow paired statistical testing.

These participants taught in three states and many different districts; we evaluated science lesson plans ranging from Grades 5 to 12. While a large proportion of the teacher participants worked in urban schools, suburban and rural schools were also represented in the sample. The teachers had a wide range of teaching experience (2–25 years). They were also diverse with respect to their prior science backgrounds, having taken from 1 to 14 post-secondary science courses before enrolling in the middle-grade or high school programs.

The authors were employed by the project, either at the time of the data collection (C.L.J. and T.C.O.) or previously (S.N.M.), as internal evaluators. As such, they were involved in gathering information for the dual purposes of informing program development and implementation in a formative sense and formal summative program evaluation. C.L.J. and T.C.O. performed lesson plan analysis and classroom observations.

**Instrument Development and Testing**

The SLPAI was developed by adaptation from a general lesson plan review protocol provided by the Institute’s external evaluation team. The development process drew heavily on the literature and policy documents, and used results from pilot implementation to refine the wording of several items and add items specifically dealing with science instruction. During the pilot phase, we also significantly modified the scoring mechanism to avoid subjective holistic evaluation, and therefore improve interrater reliability. Instrument development was an iterative process in which reviews of lesson plans from teachers not involved in this study were used to refine and specify the rubric wording, organization, and scoring protocol.

The instrument consists of four major subscales: Alignment with Endorsed Practices (AEP), Lesson Design and Implementation—Cognitive and Metacognitive Issues (CMI), Lesson Design and Implementation—Sociocultural and Affective Issues (SCAI), and Portrayal and Uses of the Practices of Science (PUPS). The full list of item titles by category is provided in Table 2. A sample item with rating descriptors is shown in Table 3.

For each item, teachers could be rated as exemplary (2 points), making progress (1 point), or needs improvement (0 point), or as intermediate between two of these categories. Raw scores were multiplied by item weight coefficients (range = 1–3), which were determined

---

6 Partner school districts sign formal agreements allowing program evaluators to collect data on teacher participants and their students through surveys, content examinations, video taping, interviews, focus groups, etc., for the formative and external evaluations of the program.

7 Teacher participant applicants are requested to submit a portfolio of teaching artifacts, including a sample of a previously enacted, unit lesson plan of approximately 5 days in length that they were likely to continue teaching in the future. In addition to a description of the daily lesson activities, they were asked to include copies of assignments, handouts, laboratories, and assessments, with examples of graded student work if possible.
### TABLE 2
SLPAI Items by Category With Scoring Weights

<table>
<thead>
<tr>
<th>Subscale With Item</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alignment with Endorsed Practices</strong></td>
<td></td>
</tr>
<tr>
<td>Alignment with standards</td>
<td>1</td>
</tr>
<tr>
<td>Awareness of science education research</td>
<td>1</td>
</tr>
<tr>
<td><strong>Lesson Design and Implementation—Cognitive and Metacognitive Issues</strong></td>
<td></td>
</tr>
<tr>
<td>Goal orientation</td>
<td>3</td>
</tr>
<tr>
<td>Content accuracy</td>
<td>2</td>
</tr>
<tr>
<td>Content presentation</td>
<td>3</td>
</tr>
<tr>
<td>Preassessment</td>
<td>2</td>
</tr>
<tr>
<td>Meaningful application</td>
<td>2</td>
</tr>
<tr>
<td>Student reflection</td>
<td>2</td>
</tr>
<tr>
<td>Assessment</td>
<td>3</td>
</tr>
<tr>
<td><strong>Lesson Design and Implementation—Sociocultural and Affective Issues</strong></td>
<td></td>
</tr>
<tr>
<td>Equity</td>
<td>1</td>
</tr>
<tr>
<td>Student engagement</td>
<td>2</td>
</tr>
<tr>
<td>Appropriate use of technology</td>
<td>1</td>
</tr>
<tr>
<td>Adaptability</td>
<td>1</td>
</tr>
<tr>
<td>Classroom discourse—Fostering a community of learners</td>
<td>3</td>
</tr>
<tr>
<td>Variety and innovation</td>
<td>2</td>
</tr>
<tr>
<td><strong>Portrayal and Use of the Practices of Science</strong></td>
<td></td>
</tr>
<tr>
<td>Hands-on exploration</td>
<td>2</td>
</tr>
<tr>
<td>Nature of science</td>
<td>3</td>
</tr>
<tr>
<td>Student practitioners of scientific inquiry</td>
<td>3</td>
</tr>
<tr>
<td>Analytical skills</td>
<td>3</td>
</tr>
<tr>
<td>Error analysis</td>
<td>1</td>
</tr>
</tbody>
</table>

### TABLE 3
Example SLPAI Item

<table>
<thead>
<tr>
<th>Item</th>
<th>Exemplary</th>
<th>Making Progress</th>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student practitioners of scientific inquiry (weight = 3)</td>
<td>Students are consistently engaged in learning content through inquiry or doing, rather than being told “answers”; inquiry process skills are taught in context</td>
<td>Some effort at engaging students in doing science is evident, with an emphasis on telling students science or Inquiry is taught out of context as a separate content area, rather than as a set of process skills to be applied</td>
<td>Students learn science exclusively by being told accepted scientific knowledge without discussion of how the knowledge was developed</td>
</tr>
</tbody>
</table>
by evaluators according to the goals of the Institute. (The use of weights as multipliers was intended to provide flexibility in adapting the instrument to contexts other than the MSP.) The weighted item scores were added and the point total was normalized to give a score out of 100, so that nonapplicable items could be excluded when not appropriate without negatively affecting the overall score.

The reliability of the SLPAI was examined using independent double scoring of 25% of the lesson plans (10/40) by the codevelopers of the instrument. The average interrater agreement in this test was 96%. In other words, the average difference between the two reviewers’ scores for these 10 plans was 4 of 100 possible points.

Subsequent to the completion of the studies presented in this paper, a researcher who had not been involved in the instrument’s development was trained to use the instrument to more convincingly verify instrument reliability. Using a new set of lesson plans, one of the instrument developers achieved 89% interrater agreement with the new researcher on 30% (8/27) of plans submitted by teacher participants who were not subjects of this study. The average difference between the two reviewers’ scores for these 8 plans was 11 of 100 possible points.

**Baseline Diagnostic Pilot Study**

Before beginning their Institute programs, teacher participants from the 2005 cohorts were asked to submit a sample of a previously enacted unit lesson plan of approximately 5 days in length that they were likely to continue teaching in future years. In addition to a description of the daily lesson activities, they were asked to include copies of assignments, handouts, laboratories, and assessments, with examples of graded student work if possible. In this way, planning and some aspects of enactment of the lesson unit could be measured, either directly or indirectly. Our aims were to determine the content and pedagogical areas of strength and weakness of the cohort teacher participants according to their average total and item scores and to provide a baseline measure of teaching practice to detect change over the course of their studies.

All “baseline” lesson plans that were submitted with sufficient information for review (17/21 MSSci plans and 8/14 HSChem plans) were evaluated using the instrument. The remaining participants either did not submit baseline lesson plans or submitted materials without enough detail for review using the SLPAI; for example, several teachers simply photocopied their district curriculum guide without indicating which of the suggested activities or assessments were actually used. These teachers were omitted from the analysis.

**Teacher Change Pilot Study**

Middle-school teacher participants in the 2005 cohort also submitted lesson plans to fulfill a science education course assignment near the end of their first full year in the program. At this time, the participants had completed one course each in physics and mathematics, and nearly finished courses in chemistry and science education. The science education course instructor provided the researchers with copies of these teacher participants Year 1 plans, which covered 2–3 days of instruction, and they were scored using the instrument. Total and subscore averages were compared with the baseline scores for significant change using \( t \) tests (baseline: \( N = 17 \), Year 1: \( N = 18 \)). Additional item-level analysis using repeated measurement analysis of variance (ANOVA) was also carried out to find specific areas of change; for this analysis, only teachers for whom we were supplied with both baseline and Year 1 plans were included (\( N = 15 \)).
Because the course assignment that was used as the Year 1 data source did not require that the lesson had been implemented in a classroom or entail submission of graded student work, analysis of these plans could not provide any information about lesson enactment. Furthermore, these lesson plans were submitted for a different purpose (graded course requirement vs. ungraded baseline establishment). For these reasons as well as the difference in length of the plans, we treated the differences between baseline and Year 1 lesson plans conservatively when attempting to draw conclusions about teacher change.

Other Data Sources

Validity of the SLPAI was examined by triangulation of the results with other measures of teaching practice, including direct observation of a subset of teachers using the Standards-Based Teaching Practices Questionnaire (SBTPQ; Scantlebury et al., 2001), and RTOP (Sawada, Piburn, Judson, et al., 2002; Sawada, Piburn, Turley, et al., 2002). Validation against the SBTPQ was conducted by comparing cohort-level conclusions generated from the 2005 baseline administration of that survey to conclusions reached using the SLPAI. Validation against the RTOP was conducted at the level of the individual teacher by testing for correlation between SLPAI scores and RTOP scores on related items.

RESULTS AND DISCUSSION

Validation of the SLPAI

Teacher participants in the 2005 middle-grade and high school programs who were evaluated using the instrument were also administered the previously validated SBTPQ before their participation in the program. The results from the independent SBTPQ analysis were compared with our instrument data, and similar but not entirely overlapping conclusions were reached. We present here the comparison between SBTPQ responses and SLPAI baseline data for both high school and middle-grade teachers as one means for instrument validation.

Using the SBTPQ, external evaluators found that middle-grade teachers reported significantly more frequent use of standards-based teaching practices in comparison with the high school teachers (Table 4), in terms of both what they do in class and what their students do. Middle-grade teachers were significantly more likely than high school teachers to report arranging seating to facilitate student discussion, using open-ended questions, requiring students to supply evidence to support their claims, encouraging students to consider alternative explanations, and using nontraditional or authentic assessments. Middle-grade teachers also reported that their students were more likely than high school teachers’ students to design activities both to test their own ideas and to talk with one another to promote learning.

The conclusions drawn from the SBTPQ measure were also supported by the baseline SLPAI data. Four SLPAI items were identified to address the same teaching practices as those listed in Table 5: Student engagement, Classroom discourse, Student practitioners of scientific inquiry, and Analytical skills. The baseline results for both middle-grade and high school teachers in these four items were analyzed using t tests to detect significant score differences between cohorts (Table 5). We found that middle-grade teachers scored significantly higher than high school teachers on all items. The largest average score difference between teachers was in the promotion of active student engagement ($p < .01$). These results indicate that the newly developed SLPAI has diagnostic overlap with the well-studied SBTPQ, thereby giving us confidence in the validity of the four instrument items that were compared.
### TABLE 4
SBTPQ Items With Significant Baseline Differences for MSSci and HSChem Teachers

<table>
<thead>
<tr>
<th>SBTPQ Item</th>
<th>MSSci (N = 23)</th>
<th>HSChem (N = 18)</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>I arrange seating to facilitate discussion</td>
<td>4.17 1.07</td>
<td>3.06 1.26</td>
<td>9.43**</td>
</tr>
<tr>
<td>I use open-ended questions</td>
<td>4.26 0.75</td>
<td>3.61 0.85</td>
<td>6.73*</td>
</tr>
<tr>
<td>I require that my students supply evidence to support their claims</td>
<td>4.35 0.65</td>
<td>3.83 0.86</td>
<td>4.80*</td>
</tr>
<tr>
<td>I encourage my students to consider alternative explanations</td>
<td>3.73 0.88</td>
<td>3.44 1.15</td>
<td>5.52*</td>
</tr>
<tr>
<td>My students design activities to test their own ideas</td>
<td>2.82 0.80</td>
<td>2.11 0.90</td>
<td>6.95*</td>
</tr>
<tr>
<td>My students talk with one another to promote learning</td>
<td>4.14 0.77</td>
<td>3.56 0.92</td>
<td>4.69*</td>
</tr>
</tbody>
</table>

*Note. The items were rated on a 1- to 5-point Likert-type scale, where 1 = “almost never” and 5 = “very often.” Adapted from Evaluation of the University of Pennsylvania Science Teacher Institute—2005–6 (p. 15), by J. B. Kahle and K. C. Scantlebury, 2006, Oxford, OH: Miami University, Ohio’s Evaluation & Assessment Center for Mathematics and Science Education.

* p < .05. ** p < .01.

A second validation test of our instrument was conducted using RTOP data. Teacher participants who also submitted lesson plans for review were observed either via video footage from their baseline portfolio (7 middle-grade and 7 high school teachers from the 2006 cohort) or in person during their first academic year in the program (8 middle-grade teachers from the 2005 cohort). It is important to note that the lessons that were observed were for the most part not taken from the same unit covered by their evaluated lesson plans, but were generated during the same school year as the directly observed lesson. Participants’ RTOP scores on items that had previously been determined by the researchers to conceptually align with SLPAI categories were tested for correlation with the teacher participants’ SLPAI scores, and the significance level (p value) was calculated. The results of this analysis are given in Table 6.

Moderate (.30 < r < .50) to large (r > .50) positive, statistically significant (p < .05) correlations were detected between the SLPAI items “Nature of science” and “Student inquiry” and their conceptually related RTOP items. We were not surprised that these significant correlations were seen between items clearly observable in both written lesson plan and direct observation formats, such as student use of manipulative techniques to represent phenomena.

In addition, several other SLPAI items exhibited low to medium positive correlations that fell below the standard threshold for significance (.15 < r < .50, p > .05) with directly
TABLE 5
Baseline Results for MSSci and HSChem Teachers on SLPAI Items Related to the SBTPQ Items in Table 2

<table>
<thead>
<tr>
<th>SLPAI Item and Description</th>
<th>MSSci (N = 17)</th>
<th>HSChem (N = 8)</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Student engagement—Requires active participation of students in their own learning</td>
<td>1.51</td>
<td>0.49</td>
<td>1.00</td>
</tr>
<tr>
<td>Classroom discourse—Lesson is structured to require and promote sense-making discussion among students</td>
<td>1.29</td>
<td>0.58</td>
<td>0.84</td>
</tr>
<tr>
<td>Student practitioners of scientific inquiry—Inquiry skills are taught in context</td>
<td>1.06</td>
<td>0.77</td>
<td>0.50</td>
</tr>
<tr>
<td>Analytical skills—Students are supported in drawing or refuting conclusions based on evidence</td>
<td>1.07</td>
<td>0.65</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Note. Lesson plans were evaluated on a 0- to 2-point scale.
* p < .05. ** p < .01.

related RTOP items: Preassessment, Meaningful application, Student reflection, Student engagement, Classroom discourse, and Hands-on exploration. Given the fairly small sample size available for this validity analysis, we believe these correlations are worthy of further exploration. As an item that spans many aspects of the classroom environment, “Classroom discourse” was tested for correlation with a number of RTOP items; as mentioned above, it did have some correlation with several of the RTOP items, but interestingly, it did not show any correlation with other items including Item 20, “There was a climate of respect for what others had to say,” which describe teaching practices that could be difficult or impossible to capture in a written lesson plan.

Comparison of our instrument and classroom observation data for validation purposes also provided some unexpected results. It was assumed that the instrument item “Goal orientation,” which sought appropriate, explicit, comprehensive, and clear learning goals, would correlate with RTOP Items 6 and 7: “The lesson involved fundamental concepts of the subject” and “The lesson promoted strongly coherent conceptual understanding.” However, these items showed a negative correlation (not statistically significant) with the SLPAI item “Goal orientation.” Likewise, the correlation between the “Content presentation” SLPAI item and RTOP Item 9, dealing with the use of abstractions when presenting content, was negative (though also not statistically significant).

Finally, the SLPAI items “Content accuracy,” “Analytical skills,” and “Error analysis” were uncorrelated with their counterpart RTOP items. Again, lack of strong correlation is likely due in part to the limited sample, but other factors also probably play a role. Since lesson plan analysis and classroom observation of each individual teacher were not done on the same lesson in most cases for logistical reason, even scores on closely matched items from our instrument and RTOP could be expected to vary widely. A mismatch between the lesson plan evaluated and the lesson observed might cause variability in the teacher’s comfort with the lesson content and/or approach to teaching data analysis skills depending on the topic or student population. Furthermore, in the specific case of the “Error analysis”
### TABLE 6
Correlations Between SLPAI Items and Related RTOP Items ($N = 22$)

<table>
<thead>
<tr>
<th>SLPAI item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal orientation</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-.29</td>
<td>-.20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Content accuracy</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.07</td>
<td>-</td>
<td>.16</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Content presentation</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-.06</td>
<td>-</td>
<td>-.21</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Preassessment</td>
<td>-.27</td>
<td>-</td>
<td>-</td>
<td>.12</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.12</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Meaningful application</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Student reflection</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-.29</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Student engagement</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-.07</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Classroom discourse</td>
<td>-.29</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.24</td>
<td>.10</td>
<td>-.06</td>
<td>.29</td>
<td>.11</td>
<td>-.09</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hands-on exploration</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-.34</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Nature of science</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Student inquiry</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Analytical skills</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Error analysisa</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. RTOP item descriptions can be found in Piburn et al. (2002). Pearson correlation coefficients were calculated for pairs of items determined to be conceptually linked by the authors. RTOP Item 23 is omitted from the table because it was not predicted to correlate to any SLPAI item.

*p < .05. **p < .01.

*aLesson plans involving topics that did not allow for analysis of experimental error were omitted ($N = 15$).
item, scores were skewed toward a very low average (0.3/2), which would tend to minimize any possible correlation.

We do not feel that these issues decrease the utility of the SLPAI or lesson plan analysis more generally; on the contrary, we propose that the presence of any contradictions in the data may provide insight into the subtle but insurmountable distinction between the intended lesson plan and the enacted lesson. What a teacher puts down on paper often differs from what occurs in the classroom in real time; for a variety of reasons, all of which are not fully understood (Roth, 1998). Use of lesson plan analysis via an instrument such as the SLPAI, in concert with classroom observation, should provide a lens with which to address this perennial concern regarding evaluation of teacher practice. The data set presented here was sufficient to validate the utility of the lesson plan analysis instrument, and we believe that the anomalies in this small data set provide areas for future research, such as comparing same-lesson and cross-lesson evaluations using SLPAI and RTOP to determine how much variability teachers have in their enactment across topics or contexts. Such research would require large projects such as the MSP to increase the resources allocated toward program evaluation.

Baseline Diagnostic Pilot Study

Baseline lesson plans from two cohorts of teachers were analyzed using the SLPAI to provide information on the strengths and weaknesses of the incoming teachers’ knowledge and practices. This analysis was intended to provide program faculty with useful knowledge about the skills and practices of their students in order to gear their instruction to be most effective. Table 7 presents the results for items with either low or high cohort averages (<1.0 or >1.5, respectively) and with a scoring weight of greater than 1.

From these data, we see that both cohorts were very strong in the areas of content accuracy and content presentation in a self-selected topic area. However, teachers in neither program showed evidence of attention to representing the nature of science, as demonstrated by their low average scores on this item. In addition, middle-grade teachers also performed well on the item dealing with student engagement, but poorly on the preassessment item. High school teachers’ lesson plans were below an average score of 1.0 in the areas of classroom discourse, variety, student inquiry, analytical skills, and student reflection.

These data suggest that teachers enter our program with established practical knowledge and experience that can be used as a foundation for further growth. We found both groups of teachers to utilize fairly accurate science information and present science topics in a relatively clear and appropriate manner, at least in the areas of science they chose to present in their baseline lesson plans. We believe that the intensive science coursework provided in the Institute will enable teachers to expand their comfort level with science, improve the accuracy of their teaching diverse topics, bring to the classroom topics they previously avoided, and gain skills and attitudes that favor life-long learning required of science teachers in a technologically oriented society.

We also found that in several of these areas our teachers’ lesson plans do not include evidence of the social constructivist teaching practices and beliefs espoused by the Institute. Teachers may be unaware of or inexperienced in implementing such practices, or their own beliefs about teaching and learning may not be congruent with those of the Institute. These results and possible interpretations point to the need for the Institute programs to address educational theory and the link between the content knowledge learned in the program and that which is utilized in the secondary classroom.
### TABLE 7
Item Analysis of Baseline SLPAI Results by Program

<table>
<thead>
<tr>
<th>SLPAI Item and Description</th>
<th>MSSci ($N = 17$)</th>
<th>HSChem ($N = 8$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>$SD$</td>
</tr>
<tr>
<td>Content accuracy</td>
<td>1.66</td>
<td>0.58</td>
</tr>
<tr>
<td>Content presentation—Level of detail and abstraction,</td>
<td>1.54</td>
<td>0.49</td>
</tr>
<tr>
<td>sequencing, and examples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nature of science—Tentative nature of knowledge based on</td>
<td>0.68</td>
<td>0.52</td>
</tr>
<tr>
<td>changing evidence, social process involving argumentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student engagement—Requires active participation of</td>
<td>1.51</td>
<td>0.49</td>
</tr>
<tr>
<td>students in their own learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preassessment—Teacher solicits student ideas in order to</td>
<td>0.32</td>
<td>0.56</td>
</tr>
<tr>
<td>plan instruction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom discourse—Lesson is structured to require and</td>
<td>1.29</td>
<td>0.58</td>
</tr>
<tr>
<td>promote sense-making discussion among students</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety—Teacher innovation or creativity keeps teacher</td>
<td>1.47</td>
<td>0.57</td>
</tr>
<tr>
<td>and students engaged</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student practitioners of scientific inquiry—inquiry skills</td>
<td>1.06</td>
<td>0.77</td>
</tr>
<tr>
<td>are taught in context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analytical skills—students are supported in drawing or</td>
<td>1.07</td>
<td>0.65</td>
</tr>
<tr>
<td>refuting conclusions based on evidence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student reflection—students reflect on and summarize their</td>
<td>1.19</td>
<td>0.65</td>
</tr>
<tr>
<td>understanding</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. SLPAI items were evaluated on a 0- to 2-point scale.

*The “Preassessment” item was added to the SLPAI during a later round of revisions after HSChem baseline lessons had been evaluated.*

Middle-Grade Master’s of Science Education (MSSci)
Teacher Change Pilot Study

The total score distributions of the MSSci baseline and Year 1 lesson plans are represented at a coarse level of detail in Figure 1. The number of participants in each score range is depicted for baseline and Year 1 lesson plans, showing the change in score distribution over time. The score distribution shifted upwards after 1 year of instruction, showing increases.
in both the lowest and the highest scores and an increase in the mean total score (Table 8). Broken down by rubric category, significant score increases were seen in the AEP and CMI categories using unpaired t tests, and smaller increases were measured in the SCAI category and the total score (Table 9). No change was seen in the PUPS category, although it had the lowest baseline category average. Note that this comparison is between nonpaired samples (some teacher participants were included in only one analysis) and that the Year 1 plans were submitted for a course grade; these differences could account for some of the change seen at this level of analysis. These data indicate that the MSSci teacher participants had made gains in some areas of content and pedagogical knowledge during their first year of Institute instruction, but that certain areas of practice were unaffected thus far.

MSSci baseline and Year 1 averages on individual items were also compared to find areas of significant improvement. The items investigated were chosen because they fit at least one of two characteristics: items where the participants’ first-year MSSci coursework is hypothesized to have an impact and/or items with a low (<1.0) baseline average. Applying repeated measures ANOVA to the raw item score distributions, we found statistically significant score increases on several SLPAI items (Table 10). The alignment of MSSci teachers’ lesson plans with state or national standards had improved somewhat, probably due to an emphasis on the need for such alignment in the Science Education course assignment.

Similarly, teachers were more likely to attend to changing students’ attitudes or beliefs about science; however, the Year 1 cohort average score of 0.85 is still below the “Making progress” level. In contrast, teachers entered the program with a fairly high average score in the “Classroom discourse” category, and still were able to significantly improve this score, perhaps due to the emphasis placed on group inquiry learning in the Science Education course. Finally, teachers had previously researched the literature and interviewed their own students to understand common preconceptions in the topic covered by their lesson plan; this assignment likely accounted for the large and very statistically significant score gain on the preassessment item. Since the cohort average rose from close to zero to well above the “Making progress” mark, this one item represents the greatest impact of the MSSci course work thus far on the teachers’ practice as measured by the SLPAI. Although other aspects of the study design could account for some of the improvements mentioned here, our ability to directly connect these areas with the teacher participants’ program experiences allows us to be confident in attributing their participation with the changes described above.

The data in Table 10 also indicate that there were several areas of concern in which teachers began with a low item average score and did not show significant improvement.
### TABLE 8

Correlations Between SLPAI Items and Related RTOP Items ($N = 22$)

<table>
<thead>
<tr>
<th>SLPAI item</th>
<th>RTOP item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>24</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal orientation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content accuracy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content presentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preassessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meaningful application</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student reflection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student engagement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom discourse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hands-on exploration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nature of science</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student inquiry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analytical skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error analysis(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. RTOP item descriptions can be found in Piburn et al. (2002). Pearson correlation coefficients were calculated for pairs of items determined to be conceptually linked by the authors. RTOP Item 23 is omitted from the table because it was not predicted to correlate to any SLPAI item.

\(a\)Lesson plans involving topics that did not allow for analysis of experimental error were omitted ($N = 15$).
### TABLE 9
Test Results for MSSci Teacher Change in SLPAI Categories

<table>
<thead>
<tr>
<th>SLPAI Category</th>
<th>Baseline average ((N = 17))</th>
<th>Year 1 average ((N = 18))</th>
<th>(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alignment with endorsed practices</td>
<td>68</td>
<td>86</td>
<td>3.38**</td>
</tr>
<tr>
<td>Cognitive and metacognitive issues</td>
<td>57</td>
<td>66</td>
<td>2.09*</td>
</tr>
<tr>
<td>Sociocultural and affective issues</td>
<td>68</td>
<td>77</td>
<td>1.66a</td>
</tr>
<tr>
<td>Portrayal and uses of the practices of science</td>
<td>49</td>
<td>49</td>
<td>0.009</td>
</tr>
<tr>
<td>Total score</td>
<td>59</td>
<td>65</td>
<td>1.35</td>
</tr>
</tbody>
</table>

*\(p < .05\). **\(p < .01\).  
*aWelch correction applied because of nonnormal Year 1 distribution.*

### TABLE 10
MSSci Teacher Change on Key SLPAI Items by Repeated Measures ANOVA \((N = 15)\)

<table>
<thead>
<tr>
<th>SLPAI Item and Description</th>
<th>Mean</th>
<th>Baseline</th>
<th>Year 1</th>
<th>ANOVA (F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alignment with standards</td>
<td>1.30</td>
<td>1.72</td>
<td>4.73*</td>
<td></td>
</tr>
<tr>
<td>Awareness of science education research—Reflects knowledge and application of theory</td>
<td>1.40</td>
<td>1.67</td>
<td>2.37</td>
<td></td>
</tr>
<tr>
<td>Goal orientation—Includes changing student values, attitudes, or beliefs</td>
<td>0.00</td>
<td>0.85</td>
<td>61.30***</td>
<td></td>
</tr>
<tr>
<td>Preassessment—Teacher solicits student ideas to plan instruction</td>
<td>0.23</td>
<td>1.50</td>
<td>83.02***</td>
<td></td>
</tr>
<tr>
<td>Assessment—Emphasizes conceptual understanding, includes grading rubric</td>
<td>1.17</td>
<td>1.42</td>
<td>4.20</td>
<td></td>
</tr>
<tr>
<td>Equity—Attempts to address equity and access for underrepresented populations</td>
<td>0.95</td>
<td>1.10</td>
<td>2.39</td>
<td></td>
</tr>
<tr>
<td>Student engagement—Motivates students and requires active participation</td>
<td>1.45</td>
<td>1.55</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td>Classroom discourse—Fostering a community of learners</td>
<td>1.20</td>
<td>1.70</td>
<td>6.46*</td>
<td></td>
</tr>
<tr>
<td>Nature of science—Reflects tentative nature of knowledge based on changing evidence, social process involving argumentation</td>
<td>0.60</td>
<td>0.50</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>Analytical skills—Students learn to support conclusions with appropriate evidence</td>
<td>1.17</td>
<td>1.07</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>Analytical skills—The sources and effects of experimental error are discussed ((N = 8))</td>
<td>0.28</td>
<td>0.19</td>
<td>0.08</td>
<td></td>
</tr>
</tbody>
</table>

*\(p < .05\). ***\(p < .001\).*
These include attention to equity, the nature of science, and error analysis. Programmatic efforts to address these areas are ongoing, and preliminary responses by some faculty members to the nature of science and error analysis items are described below.

Instructor Responses to SLPAI Evaluation Data

As previously mentioned, the high school teachers submitted baseline lesson plans with low achievement on the nature of science and student inquiry items. The middle-grade teachers also performed poorly with respect to the nature of science and error analysis items, both in their baseline and Year 1 lesson plans. Since teacher knowledge, beliefs, and practices in these areas are of great importance to the Institute, and relevant to the teachers’ Institute content courses, these results were presented to Institute science faculty members during team meetings in the spring of 2006. We presented the claim that our teachers “fail to accurately portray science as a process of generating new knowledge, and fail to engage their students in the scientific process,” and supported this claim with SLPAI data. The science faculty members were posed the question, “How can the science content courses you teach contribute to helping improve the science-specific aspects of the teaching practices of our participants?” Responses discussed included a more conscious approach to modeling these behaviors as instructors, including the use of more historical information when discussing important scientific concepts, and using inquiry or student-centered teaching methods (rather than teacher-centered, knowledge transmission methods) for content instruction more frequently in Institute courses. The instructors also discussed their own feelings about the importance of understanding and experiencing how scientific knowledge is generated for students of science. Finally, possible reasons for the differences between high school and middle-school teachers were discussed.

Master’s of Integrated Science (MSSci) Content Instructor Responses to SLPAI Evaluation

In response to these meetings, several content instructors made conscious decisions about instruction. One pair of MSSci coinstructors chose to revise their course extensively, in part, to allow time for significant examination of the role of measurement, estimation, and error in physical sciences. In the first week of class, students worked in groups to measure a difficult and ill-defined quantity, such as the height of a large statue on campus, and then reported their methods and findings to the class. While planning their procedure, many students asked the instructors to clarify what should be measured, but the instructors left it up to students to define their problem and pointed out that this is always the first step in investigating a scientific question. Before their group presentations, the instructors made explicit that the exercise was a way to experience and understand the role of peer review in the scientific process. Students gave each other feedback about potential problems in their proposed procedures. During the presentations of final results, students often expressed the desire to know the “right answer,” revealing a problematic, naïve realist view of the nature of science typical for many teachers. The instructors responded that “there are no right, definitive answers,” putting the focus back on analysis and critique of the methods used, and insisted that the students, rather than the instructors, were the arbiters of how “right” an answer was. The messages inherent to this activity were reiterated to students in smaller ways throughout the course. We are interested to see whether this second cohort of students’ new appreciation for the roles of uncertainty and peer review in science will translate to their classroom teaching as evident in future lesson plan evaluations.
Another MSSci instructor, whose course had not yet been taught at the time of the faculty meeting, later reported, “Your presentation made me think that I wanted to put more (and more explicit) emphasis on the nature of science in my class.” She decided to begin her course by emphasizing that students will develop scientific models based on their own observations, rather than formulas from the instructor or textbook. Furthermore, she plans to be explicit with the teachers about the reasons for doing so: to support the development of deeper conceptual understanding as well as appreciation for the process of scientific knowledge generation. The students currently enrolled in this course are the same MSSci teachers used to generate the teacher change SLPAI data presented in this paper; it is hoped that their experiences this year will have an impact on the areas of nature of science and error analysis, which will be evident in future lesson plans. We plan to ask teachers to submit Year 2 lesson plans before they leave the program for further analysis.

Master’s of Chemistry Education (HSChem) Content Instructor Responses to SLPAI Evaluation

As a group, the HSChem content instructors did not respond to the SLPAI data presented to them by considering or making changes to their courses. Since no Year 1 plans were available from the HSChem teachers, we did not have any evidence regarding whether program instruction had an impact on our participants’ lesson planning. Given this lack of motivating evidence, the HSChem content instructors’ reticence can probably be attributed to the fact that the HSChem program is already in its seventh year and most instructors are resistant to making changes to courses that have been developed and fine-tuned over many years. However, one HSChem content instructor described two historical data analysis projects that he has been using for a number of years. The goal of these assignments is to put students in the position of a scientist who has just done an experiment and collected data and now needs to establish criteria for determining whether the data illustrate a now-known law or mathematical relationship. These verification activities address some of the aspects of the nature of science that students often struggle with: that observations do not usually lead directly to conclusions and that some level of uncertainty accompanies inferences. Feedback from this year’s students influenced the instructor’s plan for a class discussion of the project next year, allowing a more explicit treatment of the purposes of the exercise with respect to the nature of science.

CONCLUSIONS

We conclude that the SLPAI, which utilizes artifacts of teaching and learning as data sources, is complementary but not redundant to other measures of teaching practice. The SLPAI specifically addresses issues particular to the nature of the science classroom and is a more easily scalable method than direct observation. An added benefit of lesson plan analysis is that it provides the researcher information about a larger unit of teaching than a 1-day observation, offering the researcher a more complete view of a teacher’s practices.

However, lesson plan review does present some unavoidable sources of imprecision as a measurement technique. A lesson plan, by definition, does not provide information about lesson enactment, unless postlesson information is also provided. We have also found that evaluators are often more critical of a lesson they have experienced teaching than one they have never taught or at least observed before; with previous experience, the evaluator has a range of possible teaching strategies to draw from and is more apt to note when instruction does not reach a previously set standard. For this reason, we recommend that evaluators using the SLPAI have classroom experience with the age level and science discipline being evaluated or extensive evaluation experience at the appropriate grade level.
This study also has implications for continued research in the area of teacher content planning and subsequent instruction. By utilizing the SLPAI to examine teacher participants’ lesson plans, the programs’ science education instructors are provided an opportunity to critically and explicitly discuss content planning in the context of the actual classrooms in which teachers are working. By explicitly engaging teachers in conversations around the SLPAI findings, teacher participants would be better supported to think about and plan for activities that may develop students’ understanding of the nature of science or that increase student discourse in science. This instrument may prove useful in helping teachers to identify areas in which they could improve and areas in which they already excel and can consciously choose to continually strengthen best practices. In addition, the SLPAI could be used as a professional development tool for program instructors as a means of evaluating their intended and enacted curriculum plans in the MSP courses. Conversations around the data generated from the SLPAI and RTOP scores of faculty members with relation to program goals for teacher participant development could strengthen course instruction and better align instructor classroom practices with program goals.

In addition to developing complementary instruments to support the triangulation of various data measurements, evaluation programs that rely on self-reported data with regard to content instruction would be well served to provide professional development to help program participants more accurately complete surveys. Porter (2002) suggests that training teachers to accurately describe their instructional practices could improve teacher reflection and teaching as well as provide researchers with an opportunity to gather information about how to develop instruments that are more reliable and valid. Thus, this is an area for future research in large teacher preparation programs such as this one. As teacher preparation programs expand and evaluation programs become more complex, funding agencies will need to consider providing more support for not only the evaluation of individual programs but also research on the effectiveness of the tools and methods that are being developed and utilized as part of these evaluations.

With these caveats in mind, the SLPAI is a unique and powerful tool for measuring teaching practices over time, especially when used in concert with other measurement tools. The SLPAI can be used for dual purposes: as a formative measure that informs program development and promotes effective instruction of the teacher participants, and as a summative measure that allows evaluators to provide a richer, more complete picture of program effectiveness. As MSP and other teacher professional development programs expand as a result of nationwide efforts to improve teacher quality, especially in science, technology, engineering, and mathematics fields, evaluation methods that can be used to triangulate other qualitative and quantitative measures will be needed.

**APPENDIX**

**Science Lesson Plan Analysis Instrument (Revised 2 October 2006)**

<table>
<thead>
<tr>
<th>Teacher Name:</th>
<th>School Name:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Course/Grade Level:</td>
<td>Lesson Title and Date:</td>
</tr>
<tr>
<td>Reviewed by:</td>
<td>Date of Review:</td>
</tr>
</tbody>
</table>

This lesson plan: ☐ contains clear explanation of planned activities and description of actual enactment, including student work.
☐ contains vague but sufficient explanation of planned activities and enactment.
☐ is too terse or unspecified to allow evaluation of the teacher’s work on lesson development and enactment.

*Science Education*
### Alignment with Endorsed Practices

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Exemplary</th>
<th>Making Progress</th>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alignment with standards</strong></td>
<td>Direct and explicit links to appropriate NSES, state, and/or district process AND content standards</td>
<td>Clearly contributes to students’ learning of one or more standards or benchmarks, which are not explicitly listed OR Either process or content standards are ignored</td>
<td>Not well aligned with standards</td>
</tr>
<tr>
<td><strong>Awareness of science education research</strong></td>
<td>The lesson plan gives evidence that the teacher is knowledgeable about contemporary science education research on learning, teaching, and/or curriculum, and implements these ideas regularly and effectively</td>
<td>The lesson plan gives evidence that the teacher is aware of contemporary science education research on learning, teaching, and/or curriculum, but may implement these ideas sporadically, ineffectively, or inappropriately</td>
<td>The lesson plan is antithetical to commonly accepted research findings in science education</td>
</tr>
</tbody>
</table>
### Lesson Design and Implementation—Cognitive and Metacognitive Issues

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Exemplary</th>
<th>Making Progress</th>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goal orientation</strong></td>
<td>Explicit learning goals and objectives for the unit are comprehensive and clearly comprise a big idea of science. Lesson activities clearly support goals and objectives</td>
<td>Learning goals and objectives are accurate but</td>
<td>Learning goals and objectives are not implied by the planned learning activities</td>
</tr>
<tr>
<td>(Weight 2)</td>
<td></td>
<td>○ are implied rather than explicitly stated</td>
<td>○ reflect an inaccurate understanding of the topic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>○ do not fully encompass the big idea of the topic area</td>
<td>○ do not help students attain understanding of the big idea</td>
</tr>
<tr>
<td></td>
<td></td>
<td>○ are too vague to assess or include inappropriate level of detail</td>
<td>○ are not reflected in the planned learning activities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>○ are not completely supported by the lesson activities</td>
<td></td>
</tr>
<tr>
<td><strong>Content accuracy</strong></td>
<td>Factual information is accurate and complete with respect to standards and objectives</td>
<td>Factual information is mostly accurate and may not completely reflect the learning goals or standards cited</td>
<td>Inaccurate factual information or other errors are present</td>
</tr>
<tr>
<td>(Weight 2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Content presentation</strong></td>
<td>○ Levels of detail and abstraction are challenging but accessible to most students; represents high expectations</td>
<td>○ Level of detail and/or abstraction is not challenging to a significant proportion of the class</td>
<td>○ Level of detail or abstraction is inappropriate for the course; reflects average to low expectations for all students</td>
</tr>
<tr>
<td>(Weight 3)</td>
<td>○ The sequence of topics is appropriate</td>
<td>○ Level of detail and/or abstraction is not accessible to a significant proportion of the class</td>
<td>○ The sequence of topics seems random or illogical</td>
</tr>
<tr>
<td></td>
<td>○ Appropriate examples are included</td>
<td>○ The sequence of topics is somewhat disjointed</td>
<td>○ Inappropriate examples are included</td>
</tr>
<tr>
<td></td>
<td></td>
<td>○ Appropriate examples are lacking</td>
<td></td>
</tr>
</tbody>
</table>

*Continued*
<table>
<thead>
<tr>
<th>Criteria</th>
<th>Exemplary</th>
<th>Making Progress</th>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preassessment</strong></td>
<td>The lesson plan is structured to actively solicit students’ preconceptions at the start of a topic, and refers to possible ways in which instruction could be modified in response to preassessment information</td>
<td>The lesson plan does include preassessment activities, but information is not used to inform instruction OR teacher simply attempts to refute or replace misconceptions with correct information</td>
<td>The lesson does not reflect an understanding that students’ preconceptions can affect how they understand new information.</td>
</tr>
<tr>
<td><strong>Meaningful application</strong></td>
<td>○ Content is given a meaningful personal context for students ○ Content is portrayed as significant to real-world issues</td>
<td>Some attempt is made to give content a meaningful personal context or real-world significance</td>
<td>Content is largely devoid of ○ real-world relevance ○ student-engaging context</td>
</tr>
<tr>
<td><strong>Student reflection</strong></td>
<td>Either individually or as a class, students are required to reflect on and summarize their understanding verbally or in writing at an appropriate point(s) during the unit in order to build conceptual understanding</td>
<td>Lesson is structured to allow for (but not fully promote or support) meaningful student reflection or summation that furthers conceptual understanding</td>
<td>Time is not reserved for student summation or other reflective practices</td>
</tr>
<tr>
<td><strong>Assessment</strong></td>
<td>○ Includes effective tool(s) that assess for conceptual understanding ○ includes criteria and/or rubrics for performance-based assessments (reports, participation, etc.) if necessary ○ Assessment results used to modify the lesson as it is being taught, and as formative feedback to students</td>
<td>Includes tools or suggestions for assessment that may address conceptual understanding but emphasize factual recall</td>
<td>Assessment tools do not measure student conceptual understanding OR there is no assessment tool or method described</td>
</tr>
</tbody>
</table>
### Lesson Design and Implementation—Sociocultural and Affective Issues

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Exemplary</th>
<th>Making Progress</th>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equity (Weight 1)</td>
<td>Explicit attempts were made to address equity and access for underrepresented populations</td>
<td>No mention of issues of equity or access</td>
<td>The lesson suggests bias against identifiable populations</td>
</tr>
<tr>
<td>Student attitudes about science (Weight 1)</td>
<td>The teacher’s lesson objectives or activities are designed to affect a change in student values, attitudes, or beliefs about the importance and appeal of science, their ability or desire to learn science, etc. Student’s attitudes and beliefs are evaluated in order to measure progress toward these goals</td>
<td>The lesson objectives and/or activities imply a desire for changing student values, attitudes, or beliefs about science, but no means for measuring such change is utilized</td>
<td>Lesson objectives and activities are exclusively cognitive and include no implied desire for changing student values, attitudes, or beliefs about science</td>
</tr>
<tr>
<td>Student engagement (Weight 1)</td>
<td>Activities regularly engage students by promoting curiosity and/or motivating future learning</td>
<td>Students are sometimes but not consistently engaged by activities or material OR activities engage students in a manner unproductive to learning</td>
<td>Largely devoid of engaging or motivational content</td>
</tr>
<tr>
<td>Student participation (Weight 1)</td>
<td>Lesson regularly requires active participation of students in their own learning</td>
<td>Lesson involves some level of student participation OR students are allowed but not required to participate in class discussions</td>
<td>Little or no opportunity for student participation</td>
</tr>
<tr>
<td>Criteria</td>
<td>Exemplary</td>
<td>Making Progress</td>
<td>Needs Improvement</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Classroom discourse—Fostering a community of learners (WEIGHT 3)</td>
<td>Students are part of a learning community:</td>
<td>Lesson is structured to allow for (but not require) meaningful student discussion that furthers conceptual understanding</td>
<td>○ Lesson structure inhibits meaningful discussion</td>
</tr>
<tr>
<td></td>
<td>◦ Lesson is structured to require significant discussion among students focused on sense making</td>
<td></td>
<td>○ Teacher or text acts as authority figure who provides the “right answer” and curtails discussion</td>
</tr>
<tr>
<td></td>
<td>◦ Lesson promotes evidence-based debate among students</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>◦ Suggested open-ended questions for discussion are provided</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate use of technology (WEIGHT 1 IF APPLICABLE)</td>
<td>Appropriate use of available technology (e.g., digital projector, laboratory probes, Internet resources)</td>
<td>Could better utilize available technology resources</td>
<td>Inappropriate use of technology that distracts from learning goals</td>
</tr>
<tr>
<td>Adaptability (WEIGHT 1)</td>
<td>Discusses ways to adapt the lesson to a variety of types of students (i.e., varying levels of achievement and interest, grade level, etc.)</td>
<td>Has potential to be adaptable to various needs, but is not explicitly addressed</td>
<td>Narrow range of use (type of student, class size, etc.)</td>
</tr>
<tr>
<td>Variety (WEIGHT 2)</td>
<td>Innovative or creative approach, includes varying classroom activity to keep teacher and students engaged, including but not limited to: ◦ teacher-designed or modified activities ◦ interdisciplinary collaboration</td>
<td>May not be innovative or creative on the part of the teacher, but with enough variety to keep students engaged most of the time</td>
<td>Mundane or boring to most students, and showing a low level of engagement of the teacher with the material to be taught</td>
</tr>
</tbody>
</table>
# Portrait and Use of the Practices of Science

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Exemplary</th>
<th>Making Progress</th>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on exploration</td>
<td>Well utilized and integrated to promote student exploration and learning, and includes a meaningful assessment of that learning</td>
<td>Used to verify topics, illustrate or apply processes after instruction has taken place OR Promotes student exploration and learning of content, but is not meaningfully assessed</td>
<td>○ Used solely as diversions ○ Not integrated into the curriculum OR ○ There are no appropriate hands-on experiences included in the lesson</td>
</tr>
<tr>
<td>Nature of science</td>
<td>Reflects a sophisticated view of the nature and processes of science: ◦ Explicit mention of how theories are tentative and develop and change over time based on new evidence or new treatment of previous evidence ◦ Science is treated as a social endeavor involving argumentation and explanation</td>
<td>Reflects attempts to represent the nature of science: ◦ Some mention of the tentative nature of scientific knowledge ◦ Mixed messages about the nature of &quot;truth&quot; and the &quot;right answer&quot; ◦ Illustrates the tentative and social nature of science, though exposure to the history of science in lieu of students’ own experiences</td>
<td>Treats science exclusively as a body of factual knowledge to be committed to memory AND/OR Treats experimentation exclusively as a way to find the &quot;truth&quot;</td>
</tr>
<tr>
<td>Student practitioners of scientific inquiry</td>
<td>Students are consistently engaged firsthand in learning content through inquiry or doing science (questioning, experimental design, testing hypotheses or predictions, measurement and data analysis, drawing conclusions, etc.), rather than being told &quot;answers&quot;; inquiry process skills are taught in context</td>
<td>Students do not engage in inquiry themselves, but do learn about inquiry as a scientific practice OR Some effort at engaging students in doing science is evident, with an emphasis on telling students science</td>
<td>Students learn science exclusively by being told the accepted canon of scientific knowledge without discussion of how the knowledge was developed by scientists</td>
</tr>
</tbody>
</table>

*Continued*
### Analytical skills  
(Weight 1)  
(Weight 1 if applicable)

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Exemplary</th>
<th>Making Progress</th>
<th>Needs Improvement</th>
</tr>
</thead>
</table>
| OR Inquiry is taught out of context as a separate content area, rather than as a set of process skills to be applied | OR Students are supported in drawing (or refuting) conclusions based on evidence in order to develop their analytical skills; evidence may include quantitative data or qualitative observations | Students are asked to draw conclusions based on evidence without sufficient and accurate teacher support or guidance; choice of variables, type of observation, etc. are not scaffolded appropriately for the students’ level | Age-appropriate analytical skills are not developed because  
○ there is no opportunity provided for students to analyze qualitative or quantitative data  
○ students are allowed to draw conclusions based on opinion or outside information, rather than evidence  
○ students use quantitative data to “plug and chug” using formulas to arrive at the “right answer” |
| OR Students are supported in developing higher order quantitative problem solving skills | Sources of experimental error and their size and effect on the experimental results and conclusions are discussed | Sources of experimental error are treated simplistically                                             | Sources of experimental error are ignored or glossed over                                            |
The authors thank Eric Eslinger and Nicole Gillespie for their comments on drafts of this paper and Feng Wang for assistance with determining interrater reliability.

REFERENCES


Science Education