

Designing and Using Program-specific Evaluation Instruments

Leslie Garrison
San Diego State University

Olga Amaral
San Diego State University

Abstract

This paper describes the development and use of a program-specific evaluation instrument designed to measure the impact of professional development on classroom practice. The evaluation rubric describes a sequence of skills or proficiencies that guides teachers and professional developers toward improved practice. The design and implementation of the instrument: defines the program instructional targets, guides professional development, makes target instructional skills explicit to teachers, and aligns program evaluation with program instructional targets.

Introduction

The U.S. Department of Education, *No Child Left Behind* legislation, the National Science Foundation, and state Departments of Education have all called for the implementation of research based instructional practices in the programs they fund. In addition, funded programs are often required to do formal evaluations that expand research in the field. To this end, evaluations must include empirical evidence of the program's impact on classroom practice and student learning.

Few instruments are available that document changes in teacher practice from professional development activities. The generic measures of teacher practice which can be found, are seldom targeted on the specific goals of the professional development. For example, the Observation and Analytic Protocol, developed by Horizon Research (Weiss, Pasley, Smith, Banilower, & Heck, 2003) contains an impressive list of 29 aspects of classroom instruction. While all of these characteristics are important in lesson delivery, the list does not necessarily match with the specific training focus of an individual program. This paper provides an example of how an instrument, based on the key elements of the professional development, was designed and used for program formative evaluations.

The context for this project was a science reform initiative in the Imperial Valley of southern California. The science program there has been shown to be effective in providing professional development that leads to student achievement (Amaral, Garrison & Klentschy, 2002; Klentschy & Molina de la Torre, 2005). Researchers from San Diego State University's Imperial Valley Campus have worked with the science reform project to develop the Teacher Behavior Continuum, an instrument targeted at classroom practice (Table 1). This instrument is used to document teacher instructional practices when conducting inquiry-based science lessons and is aligned to the goals and objectives of the project.

Creating a Common Vision

Increasing the pedagogical content knowledge of teachers (Shulman, 1986), or in other words, the skills necessary to effectively teach science, comprises the focus of most professional development in the project. Pedagogical content knowledge in science has been described more broadly (Morine-Dershimer & Kent, 2002) as not only pedagogical knowledge and content knowledge, but also including knowledge of the learners, knowledge of the curriculum, and knowledge of the educational goals of the course.

Reform movements often have at least three groups that work together to improve student performance: program designers, staff developers, and teachers. Program designers envision the reform, write the grants or otherwise secure funding, as well as set the goals and activity framework. This vision is passed along to staff developers who, in turn, interpret the language of the reform, into professional development activities. They most often recruit participants for the professional development, as well as conduct training sessions and follow-up activities with teachers. The teachers and recipients of the professional development activities interpret what they have learned through individual lenses, those which embed their background experiences, classroom practices, and knowledge of their students. In short, they translate reform ideas between the professional development activities and their students. By the time the initial vision reaches the final stage of interpretation and implementation at the classroom level, it may have become almost unrecognizable from the original vision. The challenge for reform programs, therefore, is to establish and maintain a common vision among the three major groups to ensure unified goals and outcomes.

Establishing a common vision is best achieved by involving all three groups from the planning stages, including the development of the vision, goals, and activities. The vision can be formed by defining the goals of the project in terms of target teacher behaviors that can be observed in the classroom. Once the target behavior has been defined, levels of performance that build toward the target behavior can be established. The behavior descriptions for each level must be sequenced in ways that accommodate the novice teacher, that is, the instructional practice of the teacher for whom the target behavior is a new concept, and to the experienced teacher who is fully integrating the target behavior in daily practice. These successive levels building toward the target behaviors comprise the Teacher Behavior Continuum (Table 1).

Uses of the Behavior Continuum

Once established, the Teacher Behavior Continuum plays a central role in program design, implementation, and evaluation. First, it creates a common vision which all participants can support and strive to implement. This vision provides a focus and direction for the program leadership and drives orientation and activities for staff development. It can also be used to inform teachers about expected goals, serving as a self-assessment for teacher practice, as well as a vehicle to guide reflection and discourse of pedagogical content. It helps make concepts about the instructional process more concrete for practitioners. Finally, it provides a template for program evaluation, one that allows for ongoing assessment of implementation of strategies over time.

Case Study: Development and Use of the Teacher Behavior Continuum in Science Reform Efforts

Science reform efforts started in Imperial Valley in 1996 with assistance of a National Science Foundation Local Systemic Change grant. During the ensuing five years, the Valle Imperial Project in Science (VIPS) trained teachers on the implementation of inquiry-based science instruction. A cadre of five science resource teachers supported the change initiative. The National Research Council talks about how the responsibility for inquiry science instruction is shifting from teacher to student resulting in shifts in inquiry lessons from guided to open. (National Research Council, 2000). The Teacher Behavior Continuum employs a five-point rubric that reflects this transition. The ability of teachers to carry out this transformation in their practice depends on how effective the professional development experiences are in advancing their pedagogical content knowledge. The Teacher Behavior Continuum has proven to be an effective tool for the VIPS leadership team as they continue to develop the target skills among teachers by focusing on Lesson Design training.

Table 1:
Teacher Behavior Continuum Video Analysis Instrument

STRAND	I	II	III	IV	V
Lesson Focus	Tasks, intent and purpose of the lesson are unclear	Tasks made clear but not the intent or purpose of the lesson	Lesson tasks and intent are clear but not set within a larger frame	Some linkages are made between the current activity and the Big Ideas of the unit	Lesson tasks and intent is clearly evident within the key concepts of the unit
Student Engagement	Many students not actively engaged in the lesson	Engages most of the students to participate	Engages nearly all students to participate at various point in the lesson	Most students engaged physically and intellectually in the lesson	Engages nearly all students physically and intellectually to contribute consistently throughout the lesson
Data, Claims & Evidence	Teacher doesn't require and/or provide direction for data collection	Teacher requires data collection but without sufficient student support	Teacher monitors and guides students to clearly and accurately record data from the lesson	Teacher ensures that students record data clearly and accurately and can interpret data	Teacher ensures that students record data clearly and accurately, can interpret data, and relate findings to the key concepts
Discourse/ Discussion	Teacher talks, students listen	Teacher engages students in procedural and management discussions	Teacher asks students fact based questions about what they did and found in the lesson	Teacher poses questions to develop student thinking that begin to link the lesson to the key concepts	Teacher poses questions that connect lesson to key concepts and requires students to explain their responses with clear lines of evidence
Closure/ Conclusion	Lesson ends without closing activity	Procedures reviewed to handle and put away materials	Lesson's activities and findings were reviewed. Teacher directed all students to the same conclusion	Lesson's activities and findings were reviewed with some reference to the key concepts	Lesson's activities and findings were reviewed and tied to lesson intent, purpose, and key concepts
English Language Development	Vocabulary assistance needed but not addressed	Vocabulary defined only after student request	Vocabulary presented verbally or in written form with no elaboration	Vocabulary introduced with context using active student participation	Vocabulary defined and integrated throughout the lesson and students required to use new words
<i>Leslie Garrison and Olga Amaral</i>					

Lesson Design

Lesson Design is a process developed by VIPS and instructors from California Institute of Technology to identify central ideas within science units that students need to know. The first step in this process is the identification of lessons that are essential to maximize student understanding of concepts in science. The second step is to determine the areas within each lesson that can lead teachers to maximize instruction and convey the “big idea” of science. In addition, it was important to develop strategies that were based on scientific processes. The habits of science are defined by Connolly (1989) as the collection of data, suspension of premature closure, and the ability to make claims based on data and evidence. Building on this premise, the team decided to include the following elements for each lesson: Lesson Focus, Student Engagement, Data, Claims, and Evidence, Discourse/Discussion, and Closure/Conclusion. Teachers receive guidance on the use of each element during professional development sessions. Trained support personnel are also available to provide ongoing classroom support. The challenge, then, is to determine the level of impact of each element in target classrooms.

The Elements

Each element is itself rooted in research. For example, the effectiveness of embedding classroom discourse in science instructional practices is well established in the literature. Teachers need to improve the intellectual level of tasks as well as the level of student communication and discourse (Ruiz-Primo, Li, & Shavelson, 2002). Rowell and Ebbers (2004) recognized three different categories of explanations developed in discourse: descriptive and relational explanations and explanatory models. Newton and Newton (2001) reported that there are certain types of oral discourse that can improve learning. Other elements have also been the focus of many studies: student engagement (Herrenkohl & Guerra, 1998; Brophy & Good, 1986); lesson focus (Garrison, 2004; Williams and Sternberg, 1993); and closure/conclusion (Baxter, Bass, and Glaser, 2000). An understanding of the research in the field provided the foundation upon which the program was built.

Guiding Question

The question that guided the work of the VIPS reform effects was: How does the professional development in Lesson Design impact the delivery of classroom instruction for grade 4-8 classrooms? The reform movement set out to help teachers implement lessons which were clear in focus, inquiry-based, enriched with discussion, inclusive of English learners, and engaged student thinking. The development of the Teacher Behavior Continuum helped the leadership team reach a common understanding and definition of the target strategies in that it required the group too not only operationalize each strategy, but to describe the sequence of skills that led toward its implementation. Videos of science lessons were made and evaluated for the presence of target instructional elements that promote student thinking and learning. The evaluation of the professional development centered on the impact of Lesson Design training on teacher practice.

Instrument

Teacher Behavior Continuum for Video Analysis

An instrument was designed for use in the evaluation of professional development relative to Lesson Design. The Teacher Behavior Continuum (Table 1) was developed collaboratively by a team of program designers, program evaluators, staff developers, and classroom teachers. The initial task before the group was to identify the central elements of lesson design and lesson delivery that the program aimed to establish. Specifically, what were the target instructional areas addressed by the program; and what specific teacher behaviors did the program seek to establish?

The areas for analysis on the rubric were agreed upon collaboratively among the program researchers, evaluators, and program personnel. These areas were: Lesson Focus, Student Engagement, Data, Claims, and Evidence, Discourse/Discussion, Closure/Conclusion, and English Language Development. Initially each area was discussed and described in a narrative form, which defined in general terms the behaviors that the group felt best described it. A summary of the groups' thoughts follows.

Lesson Focus: At the beginning of the lesson, the teacher needs to provide students with the opportunity to be clear about the tasks of the lesson as well as its intent. Students need to understand both what to do and why they were doing it.

Student Engagement: To learn, students need to be engaged in the lesson. Engagement has two parts, physical engagement and intellectual engagement. The former is indicated by the number of students who are 'busy' working on the task at hand. Intellectual engagement can be determined by the questions students asked and the questions that teachers posed. Tasks needed to be designed so that students had to analyze and to reflect as they worked.

Data, Claims and Evidence: Inquiry is the heart of the Lesson Design program. Therefore, student work needs to center around the collection and interpretation of data. Through this analytic process, students need to tie the information they gather to the central scientific ideas of the unit.

Discourse/Discussion: An important instructional principle of the Lesson Design professional development is the use of scientific discourse. The teacher should pose questions that deepen student understanding and connect the lesson generated data to the central ideas of the unit. Students are asked to explain their responses using clear lines of evidence.

Closure/Conclusion: Teachers are often very conscientious about having students put away the materials used in a lesson. Lesson Design training also encourages teachers to help students put the findings and ideas of the lesson in order as well. To do this, teachers review the lesson focus with students as well as the data collected, even if the collection process has not been completed. The day's activities are set within the larger context and next steps are reviewed.

English Language Development: Instructional strategies for English learners are assessed in both the formal introduction of new words and on how well the teachers integrate them throughout the lesson. The goal is to have the vocabulary and related concept development be addressed directly by the teacher and then observe the opportunities for the students to use the terms throughout the lesson.

Five point rubrics were developed for each of the areas, starting with Level 1, where the desired trait is least evident to a Level 5 where it is fully present. Level 1 presents a stage of development rather than a description of an inferior teacher. For example, Level I in Discourse/Discussion describes the condition where the teacher talks and the students listen. The approach to teaching described in Level I is not deemed inferior. It is merely the initial stage for many teachers as they start to incorporate discourse and discussion into their instruction.

Please note that the rubrics that are presented here reflect the target skills for one particular program. The rubrics can provide a model for other programs, but are not intended for direct application. For example, other programs may be focusing on students writing their own research questions, or may be targeting English learner strategies other than vocabulary development. When this is the case, the descriptors in the Data, Claims and Evidence and/or the English Learner sections would look very different.

Implementation

Participants

The 73 teachers participating in Lesson Design during the 2004-5 school year were self-selected from among the K-9 teachers in Imperial Valley. Demographics collected on the participants revealed that they had taught from 1.5 to 32 years with 9.45 representing the average number of years in the profession (Table 2). The teachers were a stable force having spent on average 80% of their career in the same school and nearly five years (4.84) in their current grade assignment. Twenty (20) teachers were very experienced with the science unit and had taught it as many as seven times while 30 others were attending the Lesson Design and content training prior to ever having taught the unit. One third of the teachers attended content training in an academic area that was not the same as the lesson they had videotaped. The classroom types also varied with 27 teaching in a regular classroom, and 46 were in classrooms that had Structured English Immersion (SEI) support. Four of the SEI classes were for newcomers, or students who were recent immigrants to the United States. Since in the Imperial Valley it is rare to have a class without any English learners, 64% of the teachers in this study had classes specially designed to address English Language Development. Class size ranged from seven (7) (private school) to 34 with an average class size of just over 21 students. Teachers reported teaching science, on average just over three times a week.

Table 2

Demographic Chart of Participating Teachers

Number of Teachers	Grade Range	Number Schools	Years Experience	Years at Site	Years at Grade	Experience with Unit	Lessons per wk	Class size	Reg/SEI Classes
73	K-9	27	9.45	6.96	4.84	1.3 times	3.14	21.37	27 / 46
Video Evaluation Group									
20	1-9	10	11.22	8.00	4.17	1.53	2.94	21.00	7 / 13

Data Collection

All teachers participated in the videotaped lessons and lesson evaluation. A random sample of 20 teachers was selected for videotape evaluation.

Data Analysis

The Teacher Behavior Continuum evaluation instrument was designed specifically for this program and matched to the program goals. After the first draft of the instrument was developed, the leadership team met and discussed the following: Did the rubrics describe what we intended? Did the descriptors describe what we wanted to see in the classrooms? Could the behaviors we did see in classrooms be rated according to the rubric categories? The instrument was further validated by through the iterative process of designing the rubrics, reviewing the design with the leadership team, using the rubrics to evaluate a range of video taped lesson, and then having the team suggest further modifications. This process clarified with the professional development team and the program evaluators the target skills of the project.

The reliability of the instruments was established through repeated calibrations. In these sessions, the leadership team and evaluators would watch the same video, rate it on the rubric and then share their scores. Discussions of scoring differences and the reasons that led to particular scores helped align the results. Program evaluators perform a recalibration when they have not used the instrument for several months.

Videos were viewed and individually analyzed by the researchers as per the Teacher Behavior Continuum. Findings were recorded using the rubric. Random sample lessons were chosen to be scored by both researchers in order to establish scoring consistency and reliability.

Teacher Behavior Continuum as an Instrument for Lesson Analysis and Evaluation

Twenty video tapes from randomly selected teachers were to be analyzed by the evaluation team, however only 19 were available. The teachers selected represented all grade levels from first grade through 9th grade. The lessons were analyzed for evidence of: Lesson Focus, Student Engagement, Data, Claims and Evidence, Discourse/Discussion, Closure/Conclusion, and English Language Development using the Teacher Behavior Continuum - Video Analysis Instrument. Results from the lesson analysis are summarized on Table 3.

Table 3:

Teacher Behavior Continuum Results

Teacher	Grade	Lesson	Lesson Focus	Student Engagement	Data, Claims Evidence	Discourse Discussion	Closure Conclusion	ELD
1	5	Water Cycle	2	2	1	1	1	1
2	9	Newton's 2ndLaw	1	3	3	3	1	1
3	1	Finding the Moon	3	2	1	1	2	1
4	3	Sound	4	3	3	2	2	5
5	5	Bones Skeletons	3	4	3	2	2	2
6	3	Brine Shrimp	3	3	3	3	3	5
7	3	Brine Shrimp	5	4	3	4	3	4
8	1	Soils	3	4	4	4	4	5
9	2	Sink & Float	1	3	2	3	4	3
10	5	Mixtures Solutions	4	4	4	4	4	5

11	6	Measuring Time	3	4	4	4	4	3
12	7	Micro-Macro	3	4	2	4	4	3
13	2	Butterflies	4	4	3	3	5	5
14	4	Electricity	4	5	5	5	5	3
15	8	Properties Of Matter	5	5	5	5	5	3
16	5	Mixtures Solutions	1	3	3	4	4	4
17	5	Mixtures Solutions	2	3	3	3	1	3
18	3	Brine Shrimp	5	5	3	5	2	4
19	2	Soils	5	5	5	4	4	4
Average Score			3.21	3.68	3.16	3.37	3.16	3.37

Lesson Focus: Scores in this area ranged from a low of one to a high of five. In three lessons the tasks and intent were unclear to students (level 1) and in four examples the tasks and intent were clearly evident and set within the key concepts of the unit (level 5). Most lessons (74%) contained clear tasks and intent (level 3 and higher). Teachers frequently began the lesson with a focus question and organized the lesson's activities around it. The average score in this area was 3.21 on a 5 point scale.

Student Engagement: This was the strongest area in the lessons with an average score of 3.68. Most lessons received a score of 4 or 5 which indicated that students were engaged both physically and intellectually in the lesson. Physical activity was gauged by the on-task behavior exhibited by students in the video. Intellectual engagement was shown by the level of student questions and discourse with the teacher and among their peers. The lessons shown were all inquiry-based and required active student participation. Students exhibited an enthusiasm for learning in this way by being readily involved in the lesson activities. Some teachers are reluctant to use hands-on methods with their students as they are concerned that students will be off-task and 'playing' during instructional time. This fear was not evident in the videotaped lessons evaluated for this project. Students were focused, on-task, and involved with the lesson activities.

Data, Claims, and Evidence: The Teacher Behavior Continuum emphasized that students collect data and use that data to make claims which they backed up with evidence. At numerous points during the videos that were evaluated, teachers explicitly probed students about the data they were collecting and what claims they could make based on that data. When students offered a claim, the teacher would request the evidence that led the student to the claim. Two teachers did not adhere to the lesson protocol and therefore students did not have the opportunity to collect data and make claims in the portion of the lesson that was videotaped. The other seventeen lessons required students to collect data, 16 of them provided the necessary student support to ensure success. These teachers monitored and guided students to clearly and accurately record data from the lesson. Teachers actively moved throughout the class, talking to students about their work, answering questions, and providing quality control for the information students were gathering and organizing. Five (26%) of the teachers also worked with students to help them correctly interpret the data they gathered and in several instances ensured that students were also able to link their findings to the key concepts of the lesson.

Discussion/Discourse: Again two (11%) of the teachers did not teach through inquiry during the video section and therefore had nothing upon which to base discourse. Two of the remaining 17 teachers engaged students in only procedural and management type discussions. Four (21%) of the teachers asked students fact based questions about what they did and found in the lesson activities. While these were important in checking student data collection procedures and findings, the discussion didn't link to the interpretation and connection to the major concepts of the lesson. Nearly one-third of the teachers posed questions that developed student thinking and helped them connect their findings to the lesson's concepts. This infusion of classroom discourse into the lesson is markedly different than what is seen in a traditionally taught, lecture-based lesson. There, students are seldom engaged in discussion and frequently asked single answer, fact-based questions.

Closure/Conclusion: This was a challenging element to evaluate in the video lessons as many lessons continued over a two day (or longer) period and the complete lesson was not captured on tape. That notwithstanding, closure of the day's activities was still seen as a critical part of the lesson. Just as the students are asked to organize and put away the materials they used during the lesson, they should also be given the opportunity to review and bring at least partial closure to the progress they have made during the day's lesson. This is an excellent opportunity to start sense-making with the students, have them start to see patterns, and prepare them for the next day's activities. With this in mind, each lesson was evaluated for closure. Five (26%) of the teachers either failed to have a closing activity or merely reviewed how to put away the materials. An equal number of the teachers worked with students to arrive at one conclusion which the whole class was to copy into their individual notebooks. While this might be a bit more understandable at the kindergarten and first grade level, especially at the beginning of the year when students may lack the writing skills to develop an original conclusion, the videos showed teacher directed conclusions at the third grade level. Students at this age should be able to formulate original responses. Primary grade students were able to write their own responses with teacher support.

English Language Development: Instructional strategies supportive of English Learners were present in 79% of the classrooms. The lesson vocabulary was presented in oral and written form in most of the regular classrooms, that is, classrooms not specifically designated for English learners. The ELD support strategies were even stronger in the Structured English Immersion classrooms. The average score on the rubric in the area of ELD for these classes was 3.85, a score that indicates that words are introduced in context and used active student participation. Students were also frequently expected to use the new words as the lesson progressed. The ELD strategies introduced in the Villas Institute such as working word walls and kit inventory lessons were evident in many classrooms.

The evaluation results made sense to the leadership team as they actively participated in the development of the instruments and spent many hours in teachers' classrooms so were aware of the range of teacher practices. In general, the evaluation results matched their observations, a condition that was created, at least in part by the earlier collaboration.

Teacher Behavior Continuum as an Instrument for Program Improvement

As can be seen from the descriptions above, the instrument was able to document teacher behavior in ways that led to both a numerical rating and an accurate depiction of classroom instruction. Since the Teacher Behavior Continuum was collaboratively developed, the areas

evaluated were previously agreed upon and understood by the team members. The following conclusions were substantiated through the Teacher Behavior Continuum and shared with the leadership team. The program leadership team stated that the findings were extremely helpful in planning program changes and improvement for the ensuing year.

At the request of the project leaders, evaluations were used in the aggregate to give the professional developers feedback and to help guide up coming sessions. While this was the preference for this project, the evaluations could be used to guide teacher self evaluation and these results compared to the outside evaluation scores. This could provide a springboard for a rich discussion of classroom practice. The authors have used a Teacher Behavior Continuum developed for another site for this purpose and found it to be a valuable tool for teacher reflection.

Program Strengths

- Teachers were doing inquiry-based lessons. Almost without exception, students were actively involved in inquiry practices collecting data and using the data to draw conclusions. Teachers were actively involved with the students; moving around the room as students worked; asking questions; giving direction.
- Students were actively involved in the lesson's activities, working cooperatively in groups on the lesson's tasks. Students seemed comfortable and knowledgeable in data collection procedures and were able to record the lesson's findings.

Developing Practices

- Instructional strategies inclusive of English learners were seen in most of the classrooms. Vocabulary was introduced, discussed, and in some instances woven into the lesson. Working word walls with key words were included in some of the lessons and referred to in the written analysis provided by each teacher.
- There was a clear trend toward providing direction and focus for the lessons. Terms such as "Big Idea" and "Focus Question" were used by many teachers in an attempt to set a larger frame and context for the student activities. Focus questions were frequently printed on the board and students were instructed to put them in their notebooks as part of the scientific process and as a guide to their thinking.
- A shift away from a lecture delivery model was evident. One indication of this shift was that teachers were posing questions to their students. However, questions frequently required only a factual explanation of what students had found. While factual knowledge can provide a critical first step in providing instruction based on student thinking, teachers need to practice posing questions that probe student reasoning. For example, "what happened" questions could be followed by "why" questions, that is, questions that ask students to look for patterns, trends, and relationships within the data that has been gathered.

Target for Improvement

- The area of greatest needed growth among the lessons presented is in lesson closure. Many of the lessons were not completed during the time allotted and teachers often left the day's accomplishments 'hanging.' Many teachers made no effort to set the activities in the context of

the question, to look for trends or consistency in the data, to recap the day's work and project the next steps for the following day. There were some instances, however, where closure did take place, even when the entire lesson was not complete. Here, students had the opportunity to examine the results they had compiled so far and start to understand them in the context of the lesson focus.

Program Changes Based on Teacher Behavior Continuum Results

The VIPS team used the above observations and incorporated the findings into an action plan for the upcoming year. They stated:

“Science resource teachers will focus on lesson closure as a significant part of classroom coaching. This along with time management will strengthen the overall delivery of science instruction in our classrooms. Questioning strategies will also be modeled in the content of “how” and “what” types of questions versus questions that may be simply answered “yes” or “no.” Videos from the digital video library demonstrating best practices from project teachers will also be available to model lesson closure and questioning strategies.” (Klentschy & Molina de la Torre, 2005, p. 10).

This provides evidence of how data collected through the Teacher Behavior Continuum was used and incorporated in the decision making for program improvement for subsequent years. Information collected through the Continuum provided the type of detail that the program leadership team needed to make targeted and focused program changes.

Discussion

The Teacher Behavior Continuum holds potential for helping teams of educators interested in science reform to evaluate practices since they are developed locally and designed to meet specific program goals. The team designed instrument aligns to their specific needs when using the model provided here. The construction of the rubrics must take into account the range of possible teacher behaviors, from having a lack of knowledge about the topic to the highest level of knowledge, that of full integration in daily lessons. One of the powerful facets of using this approach is the interaction among the leadership team during in the development of the instrument. The team must reach consensus on the important program goals and what target teacher behaviors exemplify those goals. When goals are clearly established at the beginning of the program, they provide a roadmap and alignment for professional development and program evaluation. This article provides the instruments that were developed and used to evaluate an inquiry based science program. These tools can be used as models for other programs to design their own program-specific evaluation instruments.

References

- Amaral, O. M., Garrison, L., & Klentschy, M. (2002). Helping English learners increase achievement through inquiry-based science instruction. *Bilingual Research Journal*, 26(2), 213-240.
- Brophy, J. & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of Research on Teaching*. New York: Macmillan.

- Baxter, G.P., Bass, K.M., & Glaser, R. (2000). An analysis of notebook writing in elementary science classrooms. *CSE Technical Report 533*. Los Angeles: University of California, Los Angeles.
- Connolly, P. (1989). Writing and the ecology of learning. In P. Connolly & T. Vilardi (Eds.), *Writing to learn mathematics and science* (pp.1-14). New York: Teachers College Press.
- Garrison, W. (2004). Profiles of classroom practices in U. S. Public Schools. *School Effectiveness and School Improvement*. 15(3/4), 377-407.
- Herrenkohl, L. & Guerra, M. (1998). Participant structures, scientific discourse, and student engagement in fourth grade. *Cognition and Instruction*, 16(4), 431-474.
- Klentschy, M. & Molina de la Torre, E. (2004). Students' science notebooks and the inquiry process. In E. W. Saul (Ed.), *Crossing borders in literacy and science instruction*. Newark, DE: International Reading Association.
- Klentschy, M. & Molina de la Torre, E. (2005). *California Mathematics and Science Partnership Mid-Course Program Report*. Sacramento, CA: Department of Education.
- Morine-Dershimer, G. & Kent, T. (2002). The complex nature and sources of teacher pedagogical content knowledge. In J. Gess-Newsome, & N.G. Lederman, (Eds.). *Examining Pedagogical Content Knowledge: The Construct and Its Implications for Science Education*. Berlin, Germany: Kluwer Academic Publications:
- National Research Council. (2000). *Inquiry and the national science standards*. Washington, DC: National Academy Press.
- Newton, D. & Newton, L. (2001). Subject content knowledge and teacher talk in the primary science classroom. *European Journal of Teacher Education*. 24(3), 369-380.
- Rowell, P. & Ebbers, M. (2004). Constructing explanations of flight: A study of instructional discourse in primary science. *Language and Education*. 18(3), 264-281.
- Ruiz-Primo, M.A., Li, M., & Shavelson, R. J. (2002). Looking into students' science notebooks: What do teacher do with them? *CSE Technical Report 562*. Los Angeles: University of California, Los Angeles.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4-14.
- Weiss, I. R.; Pasley, J. D.; Smith, P. S.; Banilower, E. R, & Heck, D.J., (2003). Looking Inside the Classroom: A Study of K-12 Mathematics and Science Education in the United States. Retrieved on October 13, 2006.
<http://www.horizon-research.com/reports/2003/insidetheclassroom/looking.php>.

Williams, W. & Sternberg, R. (1993). Seven lessons for helping children make the most of their abilities. *Educational Psychology*. 13(3/4), 317-332.