The Iterative Process of Developing an Inquiry Science Classroom Observation Protocol

Alice K. H. Taum
Paul R. Brandon
University of Hawai‘i at Mānoa

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April 9, 2006.

Purpose

With the support of a National Science Foundation grant (No. REC 0228158), a team of educational researchers, curriculum developers, and teachers from the Curriculum Research & Development Group (CRDG) at the University of Hawai‘i at Mānoa have developed an inquiry science classroom observation instrument. The instrument is designed to determine the extent to which science teachers engage in discussions with students using three specific questioning strategies that are outlined in the Foundational Approaches in Science Teaching (FAST) program Instructional Guide (Young & Pottenger, 1983). The questioning strategies being measured were introduced to teachers during their participation in a professional development training institute and supported through written instructional materials provided to the teachers following the training. The purpose of this paper is to present some primary considerations when developing a classroom observation instrument, to provide some examples of our instrument as it progressed through many iterations, and to discuss some of the challenges that we encountered along the way.

Background

The FAST Program

FAST is an interdisciplinary middle school science program developed at CRDG. It consists of three inquiry courses entitled, “The Local Environment” (FAST 1), “Matter and Energy in the Biosphere” (FAST 2), and “Change Over Time” (FAST 3). FAST is aligned with the National Science Education Standards (CRDG, 1996; Rogg & Kahle, 1997) and has received more national recognitions than any other middle school science program (Building Science and Engineering Talent, 2004; Killion, 1999; Northwest

Regional Educational Laboratory, 1998; Office of Educational Research and Improvement, 1990, 1994; U.S. Department of Education Mathematics and Science Education Expert Panel, 2001). Science teachers using the FAST model employ questioning strategies while engaging students in scientific inquiries. Serving as "research managers," the teachers use questioning strategies to help students develop hypotheses, design experiments, describe data, develop conclusions, and generate new hypotheses.

FAST was developed in the mid-1970s. Many science curricula at this time were not implemented by teachers as intended by the developers (Loucks-Horsley & Bybee, 1998). The FAST program was unique in its design, helping to ensure full program implementation by providing in-depth professional development (PD) in addition to curriculum materials. In FAST training institutes, teachers actively work on many of the investigations that they eventually teach their students. Participation in PD institutes, three decades later, is still a required feature for all new FAST teachers, thereby helping to ensure classroom implementation as the developers intended.

Project Instruments

Our NSF grant provided funding for us to prepare and validate instruments studying the effects of variations in PD on the implementation and outcomes of the FAST program. Three implementation instruments were developed, including a teacher questionnaire, a teacher log, and the instrument that is the focus of this paper: a classroom observation coding sheet. Student assessments were developed, as well. The aspects of implementation that are measured with the implementation instruments are *adherence*, *exposure*, and *quality* (Dane & Schneider, 1998, and Dusenbury, Brannigan, Falco, & Hansen, 2003). Adherence is the extent to which program implementation follows the prescribed sequence of a program's procedures, lessons, and steps; exposure is the number of procedures, lessons, or steps that are implemented and their duration; and quality is the implementation skill and knowledge shown by the service deliverer--in our case, the teachers. The questionnaire and log measure adherence and exposure, and the observation coding instrument measures quality.

The observation instrument is designed to code low-inference teacher behaviors reflecting teaching quality. Our rationale for choosing a checklist to measure quality was that the frequencies of teacher questions and teacher-student interactions are indications of the extent to which the teacher engages in inquiry practices. By recording frequency of teacher behaviors instead of rating teacher quality, we avoid the central-tendency errors and leniency errors that sometimes occur with rating scales (Evertson & Green, 1986).

Considerations When Developing Observation Instruments

As in all education and social science research, the primary considerations for developing adequate instruments are validity and reliability. (In this paper, we address validity issues; we have reported on reliability elsewhere; see Taum & Brandon, 2005.) In research on an educational program, an observation instrument is valid to the extent to which it fully addresses the target construct and does not measure aspects of the program irrelevant to the construct (Messick, 1989). Our observation instrument is valid to the extent to which it assesses teaching quality in an inquiry-based classroom, focusing on those teacher behaviors that have been identified as important. In our description of the development of the instrument in this paper, we show the steps we took to define quality in the context of an inquiry-based classroom and how we refined the checklist to hone in on the central aspects of quality instruction.

Validity concerns are paramount in developing an instrument, but they must be addressed within the constraints of the resources of a project. There are two major feasibility considerations when preparing to use classroom observations as a data source. The first and arguably most important consideration is fiscal feasibility. Hiring videographers and coders, investing in video equipment such as video cameras, digital cassette tapes, DVDs and other miscellaneous equipment to simplify and expedite the tape-to-DVD transfer are all costly investments. The second consideration is logistical. Observations involve extensive data collection, developing and monitoring inventory processes, training videographers and coders, and the lengthy coding process, each of which demands a great deal of time. Videotaping multiple teachers on multiple occasions throughout the year, transferring data from videotapes to DVDs, producing multiple copies of DVDs for multiple coders, and organizing the DVDs following clear and consistent labeling procedures are tasks that require extensive time commitments. Much training time is required for both videographers and coders as the former explore and test cameras and other equipment while the latter concentrate on becoming proficient in understanding and using the coding instrument. The training-time required for coders tends to extend beyond what is needed for videographers, mostly due to instrumentation modifications, which are an inevitable part of the instrument development process.

Development of the Observation Instrument

Over the course of one year, our research team collaboratively identified the pedagogical features that were essential to implementing the FAST program in public and private, high-, medium-, and low-socioeconomic, culturally diverse, middle-school physical science classrooms. The core research team included eight University of Hawaiʻi CRDG members, with external consultation from additional psychometricians and science assessment developers. The core team met regularly to collaborate, reviewing and providing feedback on the observation instrument, resulting in an extensive list of essential characteristics for implementation of the FAST program. From this list, early versions of what is now called the Inquiry Science Observation Code Sheet (ISOCS) were developed. The ISOCS is the heart (i.e., the measurement piece) of a full observation guide, currently undergoing final revisions, which will include a complete manual of videographer and coder training procedures and descriptions of instrument validity and reliability analyses.

Early drafts of the ISOCS were developed using an observation rubric from the Center for Research on Education, Diversity and Excellence (CREDE) as a framework. The rubric, called the Standards Performance Continuum (SPC), was designed to measure five pedagogical approaches (Five Standards for Effective Pedagogy) that are believed to promote student learning. Each of the pedagogical approaches, as well as the goal to promote student learning, were viewed as shared objectives between both the FAST program and CREDE's model. FAST was developed using a constructivist approach in which students constantly build on their own knowledge, and the Five Standards model was designed from Vygotsky's (1978) sociocultural perspective, which suggests that all higher psychological functioning has its roots in social interaction. Other studies, such as those conducted by She and Fisher (1999), Winters and Hollweg (2001), Keys and Bryan (2000), and the National Research Council (2000) also emphasize the value of teacher communication within the science classroom. Both the FAST and CREDE perspectives share the belief that learning takes place through collaboration, whether through informal social interaction or a more formal scientific community of classroom learners; therefore, the SPC was a logical starting point for instrument developers to begin designing a science observation prototype (Taum, 2004).

Nineteen teachers were videotaped between one and five times throughout the school year while teaching science using the FAST curriculum. The tapes were transferred to DVDs, then labeled, filed and stored for later coding. Two graduate students were trained to pilot-test the ISOCS while viewing and coding classroom lessons. Using a glossary of terms developed specifically for use with the instrument and a FAST student book, each student independently went through a series of training DVDs, viewing and coding each lesson.

Reviewing and revising the instrument was an ongoing process. The instrument went through a series of expansions and compressions as we sorted through the constructs that were most critical to measure in a FAST classroom. We focus in this paper on four of the major transformation points during the year-long instrument development process, beginning with a 21-item coding sheet (see Appendix A). As seen on this coding sheet, the items were divided into five categories of activities, using key features of CREDE’s Five Standards for Effective Pedagogy: (a) students working together; (b) teacher introducing new science concepts, engaging students orally or through writing; (c) teacher making connections between what students are learning and what they already know; (d) teacher challenging students, and (e) teacher engaging in conversations with students. Although the instrument appears to be straightforward and simple, once it was tested, it proved the opposite. There were two problems identified with the instrument. First, although both coders acknowledged that the five activity categories were valuable teaching practices, they found that they did not serve a useful coding purpose. The reason for initially including these headings was to help chunk groups of items together by categories and simplify the coding process; however, coders did not find this grouping system to be very helpful in identifying the 21 coding options. Second, the coders realized that of the 21 listed items, multiple item activities would occur simultaneously, resulting in considerable variation between coders and producing unreliable results. For example, the following three activities (items) were observed happening at the same time: Item 14, "Teacher reviews concepts or skills learned in previous units (lessons) that are relevant to student;" Item 4, "Teacher uses 'key questions' from the student book to guide class discussions," and Item 8, "Teacher uses questioning strategies to further student, group, or class discussions and student understanding of science concepts." This problem was discovered only after comparing the two coders individual coding results and after the coders viewed the DVD together to identify the observed teachers' behaviors.

The next pivotal point in the instrument development process resulted in a complete transformation of the coding sheet. From 21 single items, the coding sheet was divided into six, broad, teacher-initiated, fill-in-the-blank activities as indicators of high-quality implementation of the FAST program, as well as 75 activity detail options (see Appendix B). The six activities included teachers' (a) use of direct instruction, (b) use of rhetorical and interactive questioning strategies, (c) use of engaged inquiry strategies, (d) degree of movement throughout the classroom during instruction, (e) making connections between new information and what students already know, and (f) using questions from the FAST student book (Taum & Brandon, 2005). Although coders believed that the revision of the instrument made coding somewhat easier than earlier versions, there were still problems. Activities 1, 2, and 3 all seemed to be problematic. All three activities focused on teacher behaviors: Activity 1 focused on direct instruction, Activity 2 focused on direct instruction, rhetorical questioning and interactive questioning, and Activity 3 focused on inquiry questioning. These categories overlapped: Activities 1 and 2 addressed direct instruction, and Activities 2 and 3 addressed interactive questioning and inquiry questioning, resulting in coding by a process of elimination—a one-or-the-other decision—rather than the accurate assignment of a single activity-detail option for an observed behavior.

To remedy this problem, our research team discussed possible options to better distinguish between the three pedagogical approaches: direct instruction, rhetorical questioning and interactive/inquiry questioning strategies. Using the same six activities and reducing the number of "activity details" to 69, which included some repeated descriptors across the six activities, the new instrument included identification labels for three levels of instruction: (1) authoritative inquiry, (2) descriptive inquiry, and (3) Socratic inquiry (see Appendix C). These three categories were discussed and identified while the lead FAST author was concurrently developing a monograph describing FAST inquiry (Pottenger, 2005). Authoritative inquiry was designed to measure the instances of when the "*Teacher directs student(s) to* A *to* B *relating to* C;" descriptive inquiry measuring "*Through* A, *the teacher* B *science* C;" and Socratic inquiry capturing when "*The teacher questions students through* A *and responds to student* B *by* C *the comment* *or question*," maintaining the multiple alpha-numeric options for each fill-in-alpha-blanks.

Challenges were still encountered using the new-and-improved version of the instrument. First, the duplicate codes across multiple activities were not always easily identifiable by activity, leaving each coder with the decision to select the most applicable code for the situation. For example, descriptors such as "procedures" and "explanations" were offered as options for both Activities 1 and 2, with the main difference between the two being whether the teacher was teaching through direct instruction or through student-engaged discussions. Because we did not want to rely on coders "best guesses," we eliminated the duplicate options across multiple activities.

Another challenge was that the coders, whose academic background and professional experiences were primarily outside of education, stated that there were far too many variables to code reliably. This resulted in a modification of the coding process, but not the instrument. To simplify the process, as well as to improve the accuracy in recording the order in which activities were occurring, students began recording their observations by dividing them into two categories, "broader" and "other" (see Taum & Brandon, 2005). The broader activities were those that captured the larger activity in which the teacher is engaged, whereas the other activities include words or phrases that serve as descriptors to be inserted into the fill-in-the-blank broad activities. For example,

> Through interactive questioning (2A2b), the teacher introduces or provides an overview of (2B1) science investigation (2C4),

followed by the other activity,

> Teacher questions students through clarifying (3A1a) questioning and responds to student comment (3B1) by repeating (3C2) the comment and probing further (3C6).

(ISOCS codes are shown in parentheses.) This example illustrates the teacher initiating a discussion using higher-level, interactive questioning strategies (broader), while prolonging and deepening the discussion by using follow-up questions and deliberate responses to student questions (other).

Ongoing revisions of the observation instrument gradually shifted from measuring three activities to concentrating on one: Socratic inquiry. The research team concluded that this was the most important observable teacher behavior in a FAST science classroom (Taum & Brandon, 2005). The final draft of the ISOCS (see Appendix D) includes the specific type of questioning strategy used to begin an inquiry discussion, with many fewer activity details ($N = 29$) than in earlier versions ($N = 69$, $N = 75$) of the instrument. The fine-tuned version of the classroom observation instrument enables an observer to identify which of three questioning strategies, clarifying, lifting, or summarizing, is used and the context in which the question is initiated (Young & Pottenger, 1983). The teacher initiated question (Column A) is followed by a list of 14 different activities (Column B) in which students can be engaged, allowing for multiple activities to occur simultaneously, followed by a student's response (Column C), ranging from no response to a comment or question, and lastly, how the teacher responds to the student (Column D). Each coding sequence begins with a teacher question from Column A, noted by the inquiry start time in minutes and seconds, and is followed by "strings" of activities using Columns B–D.

Developing the Coding Process

A coder training session was designed and a training guide was developed through collaboration between the lead instrument developer (the first author) and one of the two initial pilot coders. A total of eight coders were hired and participated in a five-day, face-to-face, 16-hour training session, with an additional eight hours of home DVD-viewing time. The individuals included four who had teaching experience, another with a curriculum design background, a mechanical engineer, and a film festival project manager. As one would anticipate, teaching background knowledge affected coding. For example, the teaching-background group initially struggled with the notion of ignoring the students and focusing primarily on the teachers while viewing DVDs, despite repeated assurances that the student outcome measurements would be evaluated–but not using the ISOCS–whereas the non-educator-background group appreciated the simplified instructions to observe only the teachers and had little trouble ignoring the students.

The entire training process extended over a 6-week period, which included viewing and coding three teachers for a total of nine observed classroom periods. Throughout the training process, videotaped observations of teacher behaviors were first coded individually and then discussed within the larger coder group. The lead instrument developer facilitated the discussions, in which the goal was to reconcile final codings between coders for each teacher. After the training was completed, the lead instrument developer and the team of coders collaboratively revised the instrument, resulting in further iterations before the final version was accepted as reliable.

The coding process includes three general steps: a) viewing the DVD, b) coding observed teacher behaviors, and c) reconciling differences between pairs of coders. The first step requires that coders preview the entire lesson to establish a broad sense of the progression of how it unfolds. In the second step, coders carefully record all observable activities that can be captured using the ISOCS. The final step, the most time-consuming, requires that coders identify and discuss any differences in codings until consensus is reached. Reconciliation between coders is arguably the most demanding aspect of the entire coding process. It is during this time in which pairs of coders must rectify any differences in codings between them. The reconciliation process often requires a review of the DVD, with minute-by-minute references to exact time and code discrepancies. The entire process, from beginning to end takes approximately 4 to 6 hours per lesson. Throughout the training process, coders were encouraged to provide feedback regarding the utility of the code sheet, which ultimately drove the instrument to its final version.

Many lessons can be drawn from the experiences of designing an instrument, but it is the actual coding process that provided the greatest impetus in modifying the classroom observation tool. Countless hours were spent in front of a computer screen, viewing DVDs minute-by-minute by the two pilot coders, eight hired coders and lead instrument developer; time and resources allocated to 10 individuals. With multiple coders and significant variations in observed teaching patterns between the 19 teacher participants who were videotaped, the lens through which the instrument had been viewed changed from theoretical to pragmatic almost immediately following the first trials in using the ISOCS. Although we tended to include every possible significantly relevant and codable activity in the early stages of developing our observation instrument, it became evident that not all coded activities were essential features for implementing FAST, resulting in major modifications of the instrument. A reliable and cost-effective observation system should focus on examining a narrow part of a curriculum in depth, rather than concentrating on breadth. Because classrooms are complex systems, it is important to be clear and concise about those aspects of the teaching and learning environment that are going to be measured. Preliminary examination of rater agreement suggests that coding science teachers on their questioning strategies, using the ISOCS, is reliable, although further analyses will be conducted.

In addition to the content validity evidence provided here, which demonstrates how the research team narrowed the focus of the ISOCS to address the core features of inquiry, we plan to collect additional validity evidence using FAST expert observers who will measure teachers' quality through holistic ratings. These results will be compared with our earlier findings from the ISOCS, serving as a validity check of the instrument.

With high demands on teachers for excellence in student performance, ongoing professional development and formative assessment are crucial to teachers' professional growth. The Inquiry Science Observation Code Sheet has been designed to measure the quality of questioning strategies used in the FAST middle school science classroom. The results of our study will be made available to those teachers who participated and are interested in these results, as well as extend to the broader scientific educational community. Furthermore, the reflections offered here will help researchers decide whether observations are feasible for their studies and will provide insights about the many iterations required for developing and administering an observations system.

References

Building Science and Engineering Talent. (2004). *What it takes: Pre-K-12 design principles to broaden participation in science, technology, engineering and mathematics.* San Diego, CA: Author.

Curriculum Research & Development Group. (1996). *Alignment of Foundational Approaches in Science Teaching (FAST) with the national science education standards grades 5–8*. Honolulu: Author.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*, 237–256.

Evertson, C. M. &. Green, J. L. (1986). Observations as inquiry and method. *Handbook of research on teaching* (pp. 162–213). New York: Macmillan.

Keys, C. W. & Bryan, L. A. (2001, December). Co-constructing inquiry-based science with teachers: Essential research for lasting reform. *Journal of Research in Science Teaching*, 38, 631–645.

Killion, J. (1999). *What works in the middle: Results-based staff development.* Oxford, OH: National Staff Development Council.

Loucks-Horsley, S., & Bybee, R. (1998). Implementing the National Science Education Standards. *Science Teacher, 65*(6), 22–26.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13–103). New York: American Council on Education/Macmillan.

National Research Council (2000). *The national science education standards.* Washington, DC: National Academy Press.

Northwest Regional Educational Laboratory. (1998). *Catalog of school reform models.* Portland, OR: Author.

Office of Educational Research and Improvement. (1990). *Science education programs that work.* Washington, DC: U.S. Department of Education.

Office of Educational Research and Improvement. (1994). *Science and mathematics education programs that work.* Washington, DC: U.S. Department of Education.

Pottenger, F. M. (2005). *Inquiry in the Foundational Approaches in Science Teaching program*. Honolulu: University of Hawaiʻi at Mānoa, Curriculum Research & Development Group.

Rogg, S. & Kahle, J.B. (1997). *Middle level standards-based inventory.* Oxford, OH: University of Ohio.

She, H., & Fisher, D. (1999, January). The development of a questionnaire to describe science teacher communication behavior in Taiwan and Australia. *Science Education (84)*, 706–726.

Taum, A. H. K. (2004). *Foundational approaches to science teaching and the five standards for effective pedagogy*. Unpublished manuscript.

Taum, A. H. K. & Brandon, P. B. (2005, April). *Coding teachers in inquiry science classrooms using the Inquiry Science Observation Guide*. Paper presented at the meeting of the American Educational Research Association, Montreal, Canada.

U.S. Department of Education Mathematics and Science Education Expert Panel. (2001). *Exemplary & promising science programs 2001.* Washington, DC: U.S. Department of Education Office of Educational Research and Improvement.

Vygotsky, L. S. (1978). *Mind in society* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.

Winters, T., & Hollweg, K. *Inquiry in the Standards-Based Science Classroom*. *ENC Focus 8*(2), 20–22.

Young, D., & Pottenger, F. (1983). *Instructional Guide: FAST, Foundational Approaches in Science Teaching*. Honolulu, HI. University of Hawaiʻi at Manoa, Curriculum Research and Development Group.