Session Title:
From Vision Building to Implementation to Evaluation: The Many Uses of a Classroom Observation Protocol

MSP Project Name:
AIM: K-8 Science

Presenters:
Eric R. Banilower, Horizon Research
Courtney L. Nelson, Horizon Research

Authors:
Eric R. Banilower, Horizon Research (Lead)
Courtney L. Nelson, Horizon Research

Project Session

Strand 3

Summary:
AIM has developed a new classroom observation protocol based on the research presented in How People Learn for examining student opportunity to learn in science instruction. This session will introduce participants to the protocol and discuss how using such a protocol could help a project throughout its life cycle: from developing a common vision of effective instruction, to examining how teachers translate what they learn in professional development to the classroom (so mid-course adjustments to PD can be made), to examining impacts of the project on student opportunity to learn important science content.

Section 1: Questions framing the session:
How can projects develop and communicate a common vision of effective instruction? Which project stakeholders need to have this shared vision?
In what ways can conducting classroom observations using a learning-theory based observation protocol help a project?

Section 2: Conceptual framework:
The logic model that implicitly drives most professional development efforts, including those in the MSP program, is that professional development will lead to changes in teacher knowledge, beliefs, and skills, which will lead to improved classroom practice, and ultimately better student outcomes. (See Figure 1.)

Theory of Action for Professional Development (Big Picture)

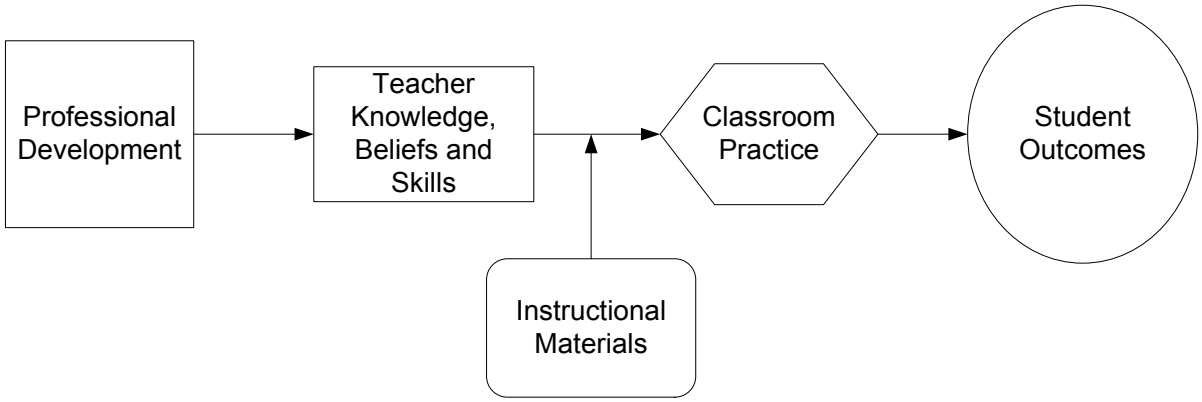

Figure 1

However, projects' efforts to study their efforts are often hampered by the scarcity of high-quality instruments. Even though a number of instruments have been developed in recent years to help address this need, these instruments cover only a small portion of the STEM content areas. Thus, many projects face the dilemma of having to cobble together instruments in an attempt to assess content that they target, or use established instruments that may only partially align with the topics and skills being addressed.

Furthermore, the push for accountability in schools may force projects to rely on data from state assessments. Although these assessments are important to many stakeholders, using them to evaluate an MSP project can be challenging. Even if project efforts align well with some topics on state assessments, most state assessments cover a much broader range of topics than projects address, raising the likelihood that the assessments would not be sensitive to impacts the project is having. In addition, the effects of a professional development program may not manifest themselves in student outcomes for several years.

For these reasons, and others, most projects also examine "upstream" outcomes such as teacher content knowledge and classroom practice to provide evidence of impact. Unfortunately, similar instrumentation issues exist for these outcomes as well.

Assessing the Impact of the MSPs: K–8 Science (AIM) was funded to help address these issues. As part of its work, AIM has developed and made available to the MSP community a number of carefully developed and tested instruments. These instruments include:

- Pairs of conceptually oriented teacher and student content assessments that use common misconceptions as distractors in each of four topics at two different grade levels (upper elementary and middle school);
- Professional development provider logs for documenting PD efforts; and
- A survey of teacher beliefs about effective science instruction.

Recently, AIM has developed a classroom observation protocol based on research in How People Learn (NRC, 2003). The protocol is structured around an instructional model that includes surfacing students' initial ideas, engaging students with evidence-providing phenomena, making and critiquing claims based on evidence, and making sense of the targeted idea (Banilower, Cohen, Pasley, & Weiss 2010).

Instruments like those developed by AIM can be used by MSPs for multiple purposes. For example, they can be used to study the impact of professional development on teacher content knowledge. They could also be used to examine the relationships among teacher content knowledge, other teacher attributes, and student learning gains.

However, just as importantly, these instruments can be used to facilitate implementation and monitoring of project efforts. For example, project personnel can use a classroom observation protocol to help develop a common vision for and language to communicate about effective instruction by co-observing lessons (either live lessons or ones captured on video) and using the observation protocol to structure discussion. Doing so provides a common language for describing common images of instruction.

In addition, developing a common vision among project staff allows projects to monitor how teachers translate their professional development experiences to the classroom. Previous large-scale professional development initiatives have found wide variation in teacher implementation of what they've learned, leading project directors to rethink their professional development programs (Pasley, 2002). Projects might also involve evaluators and researchers in these observations/discussions to help ensure alignment between those aspects of the work and the project's vision of effective instruction as well. Some projects have even used classroom observation protocols directly in professional development with principals and/or teachers, again to help develop a common vision of effective instruction. (Banilower, Boyd, Pasley, & Weiss, 2006).

Section 3: Explanatory framework:

The AIM Classroom Observation Protocol (COP) seeks to assess student opportunity to learn science concepts. Observers are asked to rate five components of instruction:

1. Appropriateness of science content;
2. Opportunities to surface prior knowledge;
3. Engaging with examples/phenomena;
4. Using evidence to draw conclusions and make claims about the examples/phenomena; and
5. Sense-making of the targeted ideas.

In each of these sections, observers first rate the extent to which several key features were present in instruction. Next, observers are asked to rate the extent to which these features of instruction aligned with the targeted science idea. Finally, observers combine this information to make a holistic rating of the extent to which the opportunities for students in that domain were likely to be sufficient for their learning of the targeted idea.

In each rating they make, observers are asked to consider what proportion of students were engaged in the instruction. For example, in the surfacing-prior-knowledge section, an observer rates the extent to which instruction was likely to make students aware of their own prior knowledge as well as their reasoning for those initial ideas, whether this information was recorded and/or shared publicly, and whether student ideas were surfaced without judgment by the teacher. Observers then rate the alignment of the surfacing opportunities with the targeted idea and then provide an overall judgment of the extent to which instruction provided sufficient opportunities for the students to surface their prior knowledge. Observers are asked to support each of their ratings with evidence from their observations.

AIM piloted the use of this protocol during the 2011-12 school year. In this pilot, the science instruction of 28 teachers during their units on force and motion was observed. Most lessons were observed by one of the nine researchers working on the project, though a few were observed by pairs of researchers. Researchers took field notes while in the classroom and wrote up lesson summaries afterwards. After all of the instruction on a targeted idea was complete, a researcher used all lesson summaries related to a targeted idea (some of which were written by other researchers who also observed the teacher teaching that idea) to complete an observation protocol for that idea. Early in the study, researchers collaborated on completing observation protocols and were given feedback by the lead researchers. Later, researchers worked independently so that interrater reliability (IRR) could be assessed.

Using the five synthesis ratings (one for each section of the protocol), IRR was examined using three methods: percent agreement; Cohen's kappa; and the interclass correlation (ICC). Overall, researchers agreed on 77 percent of their ratings. Cohen's kappa was 0.67 and the ICC was 0.86. Each of these measures is above the minimum standard described in the literature, indicating sufficient IRR among our researchers. Still, the measures could be higher, indicating that our researchers may have benefited from additional training.

In reviewing the completed observation protocols used in the IRR study, we found that differences in ratings were often due to some researchers including different parts of instruction than others did. This issue was particularly prevalent in the surfacing section. For example, because of a common curriculum being used by many of the teachers in the study, most lessons began with a surfacing activity to elicit student ideas. However, in addition to rating this initial activity, some researchers included events from later in the lesson when the teacher might have asked students to make a quick prediction before conducting an experiment. Because these quick predictions typically did not rate highly on all of the indicators in the surfacing section of the protocol (e.g., the teacher did not ask students for reasoning to support their prediction), researchers who included these instances tended to have lower ratings for surfacing.

This example points to a larger issue with using the AIM COP, and likely many other observation protocols. The AIM COP asks researchers to rate the extent to which instruction likely provided sufficient opportunities to learn. Judging "likely sufficiency"

is difficult and requires a great deal of expertise in the teaching and learning of a particular topic, as well as knowledge about the actual learners in the classroom.

We found that discussions among researchers about actual classroom instruction were extremely valuable for developing a shared understanding of not only the COP, but also of effective instruction. After researchers read the report on which the COP is based (Banilower, Cohen, Pasley, & Weiss 2010) and were using a common vocabulary to describe instruction, they watched and discussed videos of classroom instruction. These initial discussions made it clear that a shared vocabulary did not mean a common vision of effective instruction, as the researchers reached very different conclusions about student opportunity to learn. For example, one researcher may have given more weight in the sense making section to students making connections between the instructional activities and the targeted science idea while another research may have given more weight to students considering how their thinking had evolved over instruction. As the research team watched and discussed more instruction, we found that agreement increased.

Section 4: Discussion:

The findings from our work developing and testing this new COP highlight the importance of both having a common vocabulary for describing effective instruction and applying that vocabulary to shared images of instruction. MSPs that are working to improve classroom instruction would likely benefit by similar vision building among project staff, professional development providers, researchers, evaluators, and even district and school personnel. Using a classroom observation protocol to facilitate this process helps ensure that key stakeholders are using similar lenses for thinking about effective instruction. The AIM COP, which is based on learning theory research, could serve as a useful tool for assisting projects in building a common vision.

Section 5: How will you structure this session? What is your plan for participant interaction?

We envision beginning the session by surfacing participants' ideas about how a project might use classroom observations in its work. We then plan to briefly introduce the AIM COP and share our findings from piloting it, highlighting the power of co-observing instruction for developing a common vision of effective instruction. We then plan to ask participants to reflect on and discuss how they might use the AIM COP, or another observation protocol, in their own work.