New Instruments for Studying the Impacts of Science Teacher Professional Development

Peggy J. Trygstad
Eric R. Banilower
P. Sean Smith
Courtney L. Nelson

April 1, 2014

horizon
RESEARCH, INC.

Acknowledgement

This work was supported by funding from the National Science Foundation (Grant Nos. 0928177 & 0335328). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the presenters and do not necessarily reflect the views of the National Science Foundation.

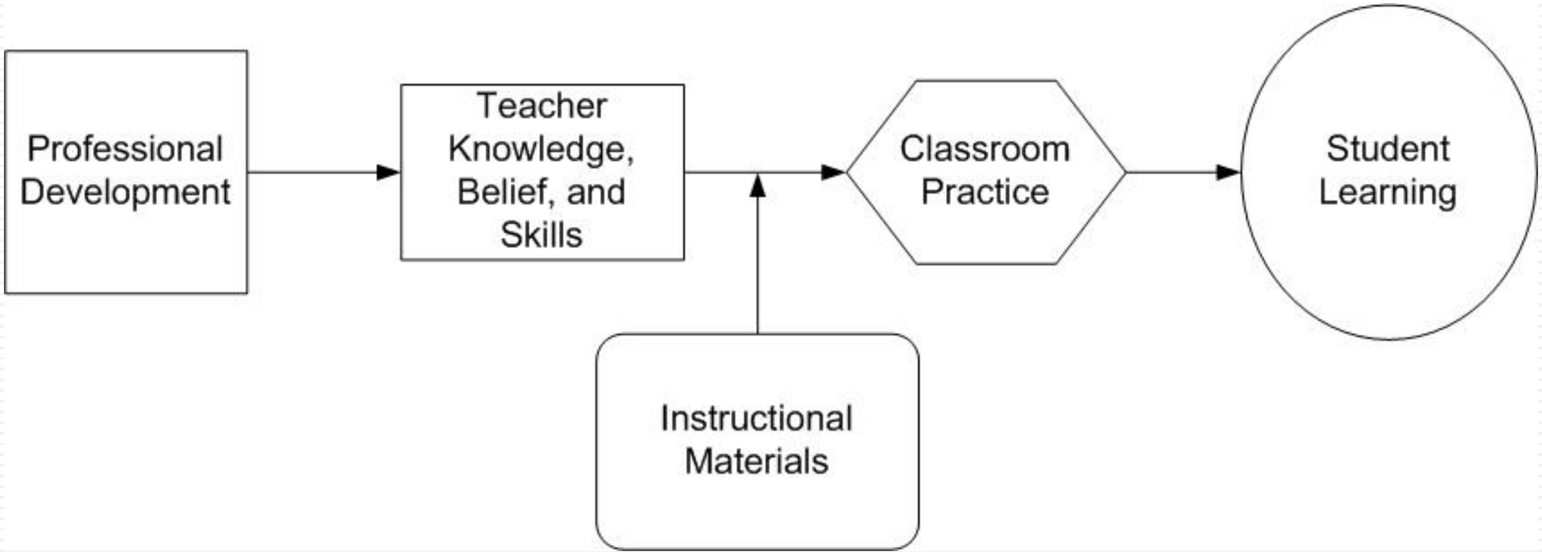
NSF

horizon
RESEARCH, INC.

Agenda

- Overview of the AIM project
- Teacher and Student Content Assessments
- Teacher Beliefs Questionnaire
- Classroom Observation Protocol
- Summary

Implicit Theory of Action for PD



What Makes PD Effective?

- Emerging Consensus: (Garet et al., 2001)
  - Focuses on content knowledge and how students learn content
  - Involves a substantial number of hours
  - Sustains focus over time
  - Models effective practice, including active learning experiences
  - Engages teachers in communities of learning
  - Involves active participation of school leaders

- Although there is a great deal of "wisdom of practice" about effective PD…
- There is little empirical research for many of these ideas.
- In addition, there is a lot that we don't know about what makes PD effective.

Why Don't We Know More from the Empirical Research?

- In applying standards of evidence, the MSP-KMD project often found vague or incomplete documentation of programs or interventions.

- Consequently, we know something worked, but we don't know a lot about what "it" was.

Why Don’t We Know More

- Studies tended to be more like program evaluations rather than research on particular strategies.

- We know the overall experience worked, but we don’t know how much particular interventions contributed to the gains.

Why Don't We Know More

- We often found serious limitations with study research designs, including:
  - Selection bias in samples and contexts
  - Lack of comparison groups or criteria
  - Idiosyncratic instrumentation, without evidence of validity/reliability/credibility

Why Don't We Know More

- There are too few studies of any one phenomenon to be able to have confidence in the robustness of the findings in any case.

- High quality research is expensive, which may explain why so many in-depth studies in the literature involved fewer than 5 teachers.

AIM: K–8 Science

- AIM was funded as a MSP RETA to develop instruments and collect data that single MSP projects typically do not have the resources to do.

Teacher and Student Content Assessments

Teacher and Student Content Assessments

- Coupled teacher and student content assessments in each of four topics at upper elementary and middle school levels:
  - Evolution and Diversity
  - Force and Motion
  - Populations and Ecosystems
  - Properties and States of Matter

Development Process

1. Define the content domain
2. Write multiple-choice items
3. Cognitive interviews
4. Pilot
5. Create final assessment

Closely mirrors a development process that has produced teacher and student science assessments with strong evidence of validity and reliability (Smith, 2010)

Defining the Content Domain

- Four topic areas were selected from the *Science Framework for the 2009 National Assessment of Educational Progress (NAEP)* (National Assessment Governing Board, 2008)

- Content in each area for each grade range unpacked into discrete, assessable statements:
  - Science concepts students should learn
  - Science concepts teachers need to know to teach the content well

Developing Multiple-Choice Items

- Developed collaboratively by teams of researchers organized by content expertise.

- All items are keyed to a specific idea.

Teacher Assessment Items

- Three types of items for assessing teacher content knowledge:
  - Level 1: Knowledge of content
  - Level 2: Using content knowledge to analyze student thinking
  - Level 3: Using content knowledge to make instructional decisions

Level 1 Teacher Item Features

- Requires knowledge of targeted science content.
- Only one answer choice is “content-correct.”

Sample Level 1 Teacher Item

A teacher asks her students how the elements are organized in the Periodic Table. Which of the following student responses is most accurate?

A. Alphabetically
B. According to similar properties
C. According to the number of neutrons
D. According to when it was discovered

Level 2 Teacher Item Features

- Address teachers' ability to *analyze student thinking using science content knowledge.*

- Only one answer choice is "content-correct" and relevant to the instructional context.

- Fairly high cognitive load.

Sample Level 2 Teacher Item

A teacher asks his students if plants have any competitors in an ecosystem. One student responds:

"Plants do not need to compete with each other, because they make their own food. They're not like animals who have to fight over food."

Based on this statement, which of the following ideas does the student seem to be missing?

A. Plants are producers.
B. Food is not the only resource for which organisms compete.
C. Animals compete with other animals for resources.
D. None. The student has an accurate understanding of competition.

Level 3 Teacher Item Features

- Address teachers' ability to *make instructional decisions using science content knowledge.*

- Only one answer choice is "content-correct" and relevant to the instructional context.

- High cognitive load.

Sample Level 3 Teacher Item

A teacher gives her students the following scenario: *"Three books are sitting on a table. Each has a different mass. If I push each book just as hard for the same amount of time, which book's motion will change the most?"*

Most students agree that all of the books will have the same change in motion because the same force is applied to all of the books. Which of the following would be the best next step to move these students forward in their understanding about the effect of forces on motion?

A. Drop all three books from the same height at the same time and see which book hits the ground first.
B. Push the books across different surfaces that have varying amounts of friction.
C. Show a video that illustrates how the strength of an applied force and the mass of an object affect an object's motion.
D. Have a class discussion about the difference between mass and weight.

Student Item Features

- Only one answer choice is “content-correct” and relevant to the question being asked.

- Include common misconceptions as distractors.

Sample Student Item

The deepest parts of the ocean are dark and very cold. Why are some organisms able to survive even in this environment?

A. Some organisms are strong and fit, so they are able to survive in any environment.
B. Some organisms are able to survive in dark, cold ocean water because they prefer that environment.
C. Different organisms have characteristics that help them survive in different environments.
D. Different organisms can decide to change their bodies to help them survive in different environments.

Cognitive Interviews (Desimone & Le Floch, 2004)

- Purpose was to maximize item validity:
  - Determine whether teachers/students interpreted items as intended
  - Reveal whether teachers/students used knowledge of the content domain to answer the questions

- Collaboratively edited items based on feedback.

- Additional rounds of interviews as needed.

Piloting

- Each teacher assessment piloted with ~350-450 teachers.

- Each student assessment piloted with ~500 students.

Create Final Assessments

- Used classical and IRT analyses to select 20-30 items for each final assessment.

IRT Reliabilities for Teacher Assessments

Assessment	Number of Items	IRT Reliability
Elementary Evolution and Diversity	30	0.88
Elementary Force and Motion	30	0.86
Elementary Populations and Ecosystems	27	0.83
Elementary Properties and States of Matter	30	0.90
Middle School Evolution and Diversity	30	0.85
Middle School Force and Motion	30	0.95
Middle School Populations and Ecosystems	26	0.78
Middle School Properties and States of Matter	30	0.84

IRT Reliabilities for Student Assessments

Assessment	Number of Items	IRT Reliability
Elementary Diversity of Life	22	0.82
Elementary Force and Motion	25	0.81
Elementary Populations and Ecosystems	25	0.83
Elementary Properties and States of Matter	25	0.77
Middle School Evolution and Diversity	30	0.84
Middle School Force and Motion	30	0.66
Middle School Populations and Ecosystems	26	0.82
Middle School Properties and States of Matter	30	0.79

User Manuals

- Background on development process
- Measurement properties of the assessment (content coverage, validity, reliability)
- Answer key
- Description of use of the assessment (appropriate uses, administration guidelines, computation of scores)

Questions?

Measuring Teacher Beliefs about Science Instruction

Landscape of Beliefs Measures

- Beliefs about science
  - Views about the Nature of Science (V-NOS) (Lederman, Abd El-Khalik, Bell, & Schwartz, 2002)
    - Is there a difference between scientific knowledge and opinion? Give an example to illustrate your answer.
    - How are science and art similar? How are they different?
  - Thinking about Science Survey Instrument (TSSI) (Cobern & Loving, 2002)

- Beliefs about science teaching
  - Self-efficacy measures
    - Science Teaching Efficacy Beliefs Instrument (STEBI) (Riggs & Enochs, 1990)
    - Teaching Science as Inquiry (TSI) (Smolleck, Zembal-Saul, & Yoder, 2006)
  - Others
    - Context Belief About Science Teaching (CBAST) (Lumpe, Haney, & Czerniak, 2000)
    - Beliefs About Reformed Science Teaching and Learning (BARSTL) (Sampson & Benton, 2006)

The Gap

- No beliefs measure explicitly aligned with research on learning (NRC 2000, NRC 2005)
- Elements of effective instruction
  - Motivation
  - Surfacing prior knowledge
  - Engaging with phenomena
  - Using evidence
  - Making sense

horizon RESEARCH, INC.

Assessment Development Approach

- Specifying the content domain
  - Unpacking the elements of effective instruction
- For example:
  - Learning is enhanced with students have opportunities to **engage with phenomena**:
    - that provide data that are **relevant** to the targeted content.
    - that are **appropriate** in terms of the students' life experiences.
    - that yield data that are **sufficiently precise** to form the science concept.
    - for which **students can collect their own data**.

Developing Items

- Individual and collaborative item writing
- Item camps
- The “practical” problem

The Preamble

We recognize that teachers have to make many trade-offs when they are responsible for teaching many standards in one year. Teachers may not be able to emphasize the instructional strategies they believe are effective and still cover the entire curriculum. When you respond to the statements below, we ask that you put those trade-offs aside. Imagine that you have no constraints, including state/district standards, available time and resources, and feasibility. We want to know what you think effective instruction looks like, without all the constraints that limit what you can do in the classroom.

Cognitive Interviews

- Middle school science teachers nationally
- Purpose: ensure that items are interpreted as intended (validity)
- The “phenomena” problem

Piloting and Field Testing

- 100 items with 950 middle grades science teachers—two response scales (agreement and importance)
- 23 items with 250 middle grades science teachers—importance scale (6 pt)
- 25 items with 250 middle grades science teachers—agreement scale (6pt)
- 21 items with 600 middle grades science teachers randomly assigned to paper vs. web
- 21 items with 900 elementary, middle, and high school science teachers

Results

- Three factors
  - Learning-theory-aligned science instruction (11 items) Cronbach's alpha 0.713

    Students should rely on evidence from classroom activities, labs, or observations to form conclusions about the science concept they are studying.
  - Confirmatory science instruction (7 items) Cronbach's alpha 0.771

    Students should know what the results of an experiment are supposed to be before they carry it out.
  - All hands-on all the time (3 items) Cronbach's alpha 0.758

    Students should do hands-on or laboratory activities, even if they do not have opportunities to reflect on what they learned by doing the activities.

Correlations Among Factors

	Learning-Theory-Aligned Science Instruction	Confirmatory Science Instruction	All Hands-on All the Time
Learning-Theory-Aligned Science Instruction	1.00		
Confirmatory Science Instruction	-0.18	1.00	
All Hands-on All the Time	-0.07	0.45	1.00

Uses

- Measuring impact of PD
- Measuring changes throughout a pre-service program
- Covariate

For More Information

Smith, P. S., Smith, A. A., & Banilower, E. R. (in press). Situating Beliefs in the Theory of Planned Behavior: The Development of the Teacher Beliefs about Effective Science Instruction Questionnaire. In C. M. Czerniak, R. Evans, J. Luft, & C. Pea (Eds.), *The Role of Science Teachers' Beliefs in International Classrooms: From Teacher Actions to Student Learning*. Book in preparation.

horizon RESEARCH, INC.

Questions?

AIM Classroom Observation Protocol (COP)

History of COP Development

- HRI initially developed a widely-used COP for the evaluation of NSF's Local Systemic Change through Teacher Enhancement Initiative (LSC).

- The LSC COP was revised for use in the Inside the Classroom Study.

LSC / Inside the Classroom COPs

- Lessons were rated on:
  - Design
  - Implementation
  - Content
  - Classroom Culture

- But, observers often favored certain features of instruction, such as hands-on and higher order questioning, even if these features were not well aligned with the lesson's goals.

Other COPs

- Many others have developed protocols for observing science instruction with varying goals, including:
  - Reformed Teaching Observation Protocol (RTOP) (Sawada et al., 2002)
  - Science Teacher Inquiry Rubric (STIR) (Beerer & Bodzin, 2003)
  - Practices of Science Observation Protocol (PSOP) (Forbes et al., 2013)

Learning Theory

- The release of *How People Learn* (Bransford et al., 1999), and the issues we saw with how our previous COP was used, spurred us to redesign.
- AIM COP is explicitly aligned with learning theory.

Additionally…

- Other features of the AIM COP:
  - Targeted idea is central to all ratings
  - Neutral to pedagogy
  - Focus on conceptual understanding of science ideas
  - Aligned with the nature of science

AIM COP Structure

- Two main sections:
  1. Description of Instruction
  2. Ratings and Rationales

Ratings

- Students' Opportunity to Learn is rated on the following elements of effective science instruction:
  1. Opportunities to surface prior knowledge
  2. Engaging with examples/phenomena
  3. Using evidence to draw conclusions and make claims about the examples/phenomena
  4. Sense-making of the targeted idea

Ratings

- Three components:
    1. To what extent are key features of the element present within the observed instruction?
    2. To what extent is the instruction aligned to the targeted idea?
    3. To what extent is the instruction sufficient for learning the targeted idea?

Science Content

- Observers rate the extent to which the science content in the instruction was:
  - Accurate
  - Developmentally appropriate

Opportunities to Surface Prior Knowledge

- Key Features
  - Students are made aware of their own prior knowledge
  - Students are asked to provide reasons for how they are thinking
  - Students record and/or make public aspects of their prior knowledge
  - Students’ ideas are surfaced without judgment

For example…

- If we had the following targeted idea:

A force is a push or pull exerted on one object by another object when they interact with one another.

Possible Activities for Surfacing Students' Prior Knowledge

A. What are some examples of forces that you saw on your way to school this morning?

Vs.

B. Imagine a soccer player kicks a ball, which flies toward the goal where the goalie catches it. When did the force of the kick stop acting on the ball?

Engaging with Examples/Phenomena

- Key Features
  - Examples/phenomena are accessible to students
  - Students are focused on the relevant aspects of the examples/phenomena
  - Students describe and/or keep record of the processes they use/data they generate

As a reminder

- Our targeted idea is:

A force is a push or pull exerted on one object by another object when they interact with one another.

Possible Activities for Engaging with Examples/Phenomena

Students are given a toy car with a piece of soft foam attached to the top.

A. Students measure how long it takes for the car to move a specified distance when they push on the foam softly. In a second trial, they push harder on the foam and time how long it takes to cover the same distance.

Vs.

B. Students record their observations of the shape of the foam when pushed and not pushed, doing so while the car is at rest and while it is moving.

Using Evidence to Draw Conclusions and Make Claims

- Key Features
  - Helps students understand what the data represent
  - Facilitates students' interpretation/analysis of the data
  - Students use evidence to support their claims
  - Students use evidence to critique claims

Possible Activities for Drawing Conclusions/Making Claims

A. The teacher asks students to share something from their observations

Vs.

B. The teacher asks students, “When was a force acting on the car and how do you know?”

Sense-Making

- Key Features
  - Students connect what they did in the instruction to the targeted idea
  - Students use evidence from multiple phenomena to support/critique claims about the targeted idea
  - Students compare their emerging understanding of the targeted idea to their prior ideas and other science ideas they already know

Possible Activities for Sense-Making

A. Write an entry in your journal about something you learned today.

Vs.

B. Let's revisit the soccer ball example. How has your thinking changed about when the force of the kick stops acting? Why?

Piloting the AIM COP

- We piloted the AIM COP during the 2011–2012 school year.

- Eight researchers observed the instruction of 28 teachers (over 500 lessons) during their implementation of a unit on force and motion, focusing on eight targeted ideas.

Training to Use the AIM COP

- Before beginning the observations, researchers trained together in order to develop a common understanding of the COP.

Completing AIM COPs

- Researchers wrote a lesson summary, based on field notes, for each observed lesson.

- After observing all instruction for a targeted idea, an AIM COP was completed using all of the lesson summaries related to that targeted idea.

Completing AIM COPs

- Initially, researchers who observed instruction on the same targeted idea by the same teacher collaborated on the completion of the AIM COP.

- Later, the researchers worked independently so that inter-rater reliability (IRR) could be assessed.

Findings

- IRR was examined for the five ratings using percent agreement and the intraclass correlation coefficient (ICC).

- Overall, the researchers agreed exactly on 77% of their ratings and ICC was 0.86.

- Both measures are above the minimum standard (Graham, Milanowski, & Miller, 2012), indicating sufficient IRR among researchers.

Questions about the AIM COP?

Summary

Ways Instruments Have Been Used

- AIM has been conducting studies examining:
  - Relationships among elements of PD and teacher learning;
  - Relationships among teacher content knowledge, beliefs about effective science instruction, classroom practices, and student learning; and
  - Impact of learning theory-based curriculum on classroom practices and student learning.

- Others have been using the instruments:
  - The 16 content assessments have been used by a number of NSF- and state-funded Math Science Partnership projects to look at the impacts of their PD on teacher and student learning;
  - The TBEST has been used to study impacts on teachers' beliefs resulting from an eight-day summer workshop focused on kit-based instruction;
  - TBEST has also been used to study how pre-service teacher beliefs about science instruction change over the course of their preparation program;
  - The classroom observation protocol has been used as a basis for practicum observations and post-observation conferences with pre-service teachers.

- All of the instruments developed by AIM are, or will soon be, available through our website:

  http://www.horizon-research.com/aim/

- User Manuals are being posted to provide support for the instruments.

Questions?

- What questions and/or suggestions do you have about:
  - The instruments?
  - How they've been used?
  - How they could be used in the future?

References

Banilower, E., Cohen, K., Pasley, J., & Weiss, I. (2008). *Effective science instruction: What does research tell us.* Portsmouth. NH: Center for Instruction.

Beerer, K. M. & Bodzin, A. M. (2003). Promoting inquiry-based science instruction: The validation of the science teacher inquiry rubric (STIR). *The Journal of Elementary Science Education, 15*(2), 39–49.

Bransford, J., Brown, A. L., & Cocking, R. R. (1999). *How people learn: brain, mind, experience, and school*. Washington, DC: National Academy Press.

Cobern, W. W. & Loving, C. C. (2002). Investigation of preservice elementary teachers' thinking about science. *Journal of Research in Science Teaching*, *39*(10), 1016–1031.

Desimone, L. M. & Le Floch, K. C. (2004). Are We Asking the Right Questions? Using Cognitive Interviews to Improve Surveys in Education Research. *Educational Evaluation and Policy Analysis*, *26*(1), 1–22. doi:10.3102/01623737026001001

Forbes, C.T., Biggers, M., & Zangori, L. (2013). Investigating Essential Characteristics of Scientific Practices in Elementary Science Learning Environments: The Practices of Science Observation Protocol (P-SOP). *School Science and Mathematics,* 113(4), 180-190.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Education Research Journal, 38*(4), 915-945.

Graham, M., Milanowski, A., & Miller, J. (2012). Measuring and promoting inter-rater agreement of teacher and principal performance ratings. Washington, DC: Center for Educator Compensation Reform.

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, *39*(6), 497–521. doi:10.1002/tea.10034

Lumpe, A. T., Haney, J. J., & Czerniak, C. M. (2000). Assessing teachers' beliefs about their science teaching context. *Journal of Research in Science Teaching*, *37*(3), 275–292.

National Assessment Governing Board. (2008). Science Framework for the 2009 National Assessment of Educational Progress. Washington, DC: U.S. Department of Education.

References

National Research Council. (2000). *How people learn: Brain, mind, experience, and school (Expanded Edition).* J. D. Bransford, A. L. Brown, & R. R. Cocking (Eds.). Washington, DC: National Academy Press.

National Research Council. (2005). *How students learn: Science in the classroom.* M. S. Donovan & J. D. Bransford, (Eds.) Washington, DC: National Academy Press.

Riggs, I. M. & Enochs, L. G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education*, *74*(6), 625–637.

Sawada, D., Pilburn, M., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics, 102*(6), 245–253.

Sampson, V. & Benton, A. (2006). Development and validation of the beliefs about reformed science teaching and learning (BARSTL) questionnaire. *Annual Conference of the Association of Science Teacher Education (ASTE). Portland, Oregon. Retrieved February* (Vol. 16, p. 2008).

Smith, P. S. (2010, March). *New tools for investigating the relationship between teacher cotent knowledge and student learning.* Paper presented at the National Association for Research in Science Teaching, 2010 Annual International Conference, Philadelphia, PA.

Smolleck, L. D., Zembal-Saul, C., & Yoder, E. P. (2006). The development and validation of an instrument to measure preservice teachers' self-efficacy in regard to the teaching of science as inquiry. *Journal of Science Teacher Education*, *17*(2), 137–163.