EVALUATION MATERIALS

Written by
REBECAH BUSSELLE

Terms

Instrument: general term for a measurement device (e.g., survey, test, questionnaire, etc.)

Validity: the extent to which an instrument is measuring what it is supposed to measure

Reliability: the extent to which an instrument consistently measures what it is intended to measure

Once an Evaluation Team has planned the design of a program evaluation for teacher professional development (PD), their next task is to identify or develop instruments for data collection. Ideally, an Evaluation Team would employ instruments that are specific to the professional development program being evaluated and the evaluation questions being asked. But this isn't always feasible.

> "In-house professional learning is hard to evaluate due to lack of high-quality tools. So we use simple ones."
>
> —Unconference participant

Often an evaluation design calls for an instrument that has been externally developed and/or validated. It can be challenging to find a publicly available instrument that fits the specific needs of an evaluation.

Developing new instruments for each evaluation solves this problem, but instrument development is costly and requires very specific expertise. This is particularly true if the instrument needs to be statistically valid and reliable. More often than not, instrument development is prohibitively costly for Evaluation Teams.

As a result, many PD evaluators take the adopt-and-adapt approach. They identify previously developed instruments that are closely aligned with their needs and adapt them for each new evaluation. These adapted instruments are often combined with newly developed instruments and externally developed instruments as part of a complete evaluation strategy. This is a particularly useful approach when some, but not all, evaluation questions require a statistically valid and reliable instrument.

Evaluation Instruments for K–8 Science PD

The tables on the following pages provide an overview of publicly availably instruments commonly used in evaluating PD for K–8 science teachers. The level of evaluation has been identified for each instrument to assist you in locating instruments to answer your specific evaluation questions. There are many additional resources that have been used in educational research available behind subscription paywalls that have not been listed here. The research compendiums from CADRE and the STELAR database (see table) are useful tools for identifying these additional resources.

Validation

- Validated Instrument(s)

Level of Evaluation

[1] **Quality of PD Design & implementation**
[2] **Teacher-level outcomes**
- [2a] Teacher knowledge and skills
- [2b] Teacher beliefs & attitudes
- [2c] Teacher self-efficacy

[3] **Classroom practices**
[4] **Student-level outcomes**
- [4a] Cognitive (Performance & Achievement)
- [4b] Affective (Attitudes & Dispositions)
- [4c] Psychomotor (Skills & Behaviors)

		Developer Description	[1] Quality of PD Design & Implementation	[2a] Teacher knowledge and skills	[2b] Teacher beliefs & attitudes	[2c] Teacher self-efficacy	[3] Classroom practices	[4a] Cognitive (Performance & Achievement)	[4b] Affective (Attitudes & Dispositions)	[4c] Psychomotor (Skills & Behaviors)	Other
DATABASE DEVELOPERS	Horizon Research, Inc.	As part of an NSF-funded project, Assessing the Impact of the MSPs: K-8 Science (AIM), Horizon Research, Inc. developed a set of tools to examine the impact of professional learning on teacher knowledge and practice as well as on student learning. The science content instruments focus on four contents areas included in the 2009 NAEP Science Framework: evolution and diversity, force and motion, interdependence, and properties of and changes in matter. AIM instruments are available at no charge. Source: http://www.horizon-research.com/aim/instruments/	•	•	•		•	•			
	STELAR	The STEM Learning and Research Center (STELAR) provides technical support to facilitate the NSF's Innovative Technology Experiences for Students and Teachers (ITEST) program. ITEST supports the research and development of K-12 STEMeducation. STELAR maintains a searchable database of instruments for use in research and evaluation. Source: http://stelar.edc.org/resources	•	•	•	•	•	•	•	•	•
	CADRE (DRK12)	The Discovery Research K-12 (DR K-12) Program of the NSF supports research and development (R&D) on innovative resources, models, and tools for use by students, teachers, administrators, and policy makers. Each project also has formative and summative evaluation to support and measure the effectiveness of those activities. To support these research and evaluation activities, CADRE has published two compendiums of research instruments for STEM education, one focused on Teacher Practices, PCK and Content Knowledge and the second focused on Students' Content Knowledge, Reasoning Skills, and Psychological Attributes. These resources include information about state content tests for student assessment and teacher accreditation. Source: http://cadrek12.org/resources		•	•	•	•	•	•	•	•
	PASS	Partnership for the Assessment of Standards-based Science (PASS) is a valid and reliable standards-based science assessment for elementary, middle, and secondary levels. PASS is aligned to the National Science Standards and the Benchmarks for Science Literacy. Source: https://www.wested.org/service/partnership-assessment-standards-based-science/						•			
	MSS	Student & Teacher Content Tests developed by Making Sense of SCIENCE (MSS) are intended to help gauge growth in knowledge about a specific science content area as a result of an intervention, such as professional learning, curriculum use, or mentoring. Tests are available in a range of topics for teachers and students. Source: https://we-mss.weebly.com/products.html		•				•			

Publicly Available Instruments Commonly Used in K–8 Science PD Evaluation

	INSTRUMENT	TYPE	Data on Reliability	Evidence of Validity	GRADE LEVELS K–2	3–5	6–8
[2b]Teacher beliefs & attitudes	CIP: Inside the Classroom: Teacher Interview Protocol	Interview			•	•	•
	Inside the Classroom Teacher Questionnaire	Survey			•	•	•
	TIMSS-R Science Teacher Questionnaire — Main Survey	Survey	•	•	•	•	•
[2a]Teacher Self Efficacy	SETAKIST: Self-Efficacy Teaching and Knowledge Instrument for Science Teachers	Survey		•	•	•	
	STEBI: Science Teaching Efficacy Belief Instrument	Survey	•		•	•	
	TSES: Teachers' Sense of Efficacy Scale	Survey	•	•	•	•	•
[3]Classroom Practices	CETP-COP: The Collaboratives for Excellence in Teacher Preparation core evaluation classroom observation protocol	Observation & Interview	•		•	•	•
	Classroom Snapshot	Observation			•	•	•
	COP: Inside the Classroom: Observation & Analytic Protocol (Part I)	Observation	•		•	•	•
	EQUIP: Electronic Quality of Inquiry Protocol	Observation	•	•	•	•	•
	FFT: Danielson's Framework for Teaching Domains	Rubric			•	•	•
	LSC: Local Systemic Change Classroom Observation Protocol	Observation	•	•	•	•	•
	RTOP: Reformed Teaching Observation Protocol	Observation	•	•	•	•	•
	SEC: Survey of Enacted Curriculum	Survey	•	•	•	•	•
	STIR: Science Teacher Inquiry Rubric	Observation	•	•	•	•	
[4a]Cognitive (Performance & Achievement)	DAST: Draw a Scientist Test	Performance Assessment	•	•	•	•	•
	Children's Self-Efficacy Scale	Survey	•	•	•	•	•
	mATSI: Modified Attitudes Towards Science Inventory	Survey	•	•		•	
	Patterns of Adaptive Learning Scales	Survey	•	•	•	•	•
	SAI II: The Scientific Attitude Inventory: A revision	Survey	•	•			•
	VNOS-C: Views of Nature of Science Form C	Survey	•	•	•	•	•
[4b]Affective (Attitudes & Dispositions)	MOSART: Misconceptions-Oriented Standards-Based Assessment Resources for Teachers	Survey	•	•	•	•	•
	NAEP: National Assessment of Educational Progress	Test	•	•		•	•
	Scientific Inquiry and Engineering Design Scoring Guides	Rubric				•	•
	SLA: Science Learning Assessment	Test	•	•	•		
	TIMSS: Trends in International Mathematics and Science Study	Test	•	•		•	•