August 2012

Addendum added:

May 2013

Compendium of Research Instruments for STEM Education

PART I: Teacher Practices, PCK and Content Knowledge

With Addendum

Community for Advancing Discovery Research in Education

Written for CADRE by:

Abt Associates
Daphne Minner
Alina Martinez
Brian Freeman

August 2012

Addendum added 2013

The original compendium included instruments that were identified through review of the projects funded in the first five cohorts of the National Science Foundation's Discovery Research K-12 (DR K-12) program. An Addendum has been added to this version of the compendium to include an additional ten instruments that were identified through review of the projects in the sixth DR K-12 cohort, which received initial funding in 2012. The two tables in the Addendum (beginning on p. 49) include similar information about these additional instruments as was presented for the original set of instruments contained in the main document. However, the information from the additional instruments contained in the Addendum is not incorporated in the body of this compendium, which remains substantively unchanged from the first release in August of 2012.

This project is funded by the National Science Foundation, grant # 0822241. Any opinions, findings, and conclusions or recommendations expressed in this materials are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Table of Contents

Introduction

President Obama's administration has brought a renewed interest and focus on science, technology, engineering, and mathematics (STEM) education and related workforce issues. For example, The America COMPETES Reauthorization Act of 2010 (P.L. 111-358) called for the National Science and Technology Council's (NSTC) Committee on STEM Education to create a 5-year Federal STEM education strategic plan. As an initial step in this strategic planning effort, the NSTC conducted a portfolio review of federal STEM education programs (NSTC, 2011). This report described how 13 Federal agencies utilized $3.4 billion in fiscal year 2010 to support STEM education, out of the $1.1 trillion in annual U.S. spending on education. An independent audit conducted by the Government Accounting Office (GAO, 2012) found that across these 13 agencies, 209 STEM education programs were administered in fiscal year 2010. The Departments of Education (ED) and Health and Human Services (HHS) along with the National Science Foundation (NSF) had the largest fiscal investments, with NSF making the greatest investment (GAO, 2012), even though HHS administered slightly more programs. "Eighty percent of the funding supported STEM education investments were made by NSF, ED, and HHS" (NSTC, 2012, p.6). Across the NSF's six education research and development programs, Discovery Research K-12 (DR-K12) has the largest budget (NSTC, 2011).

The DR-K12 program seeks to significantly enhance the learning and teaching of STEM. The funded research projects focus on the "development, testing, deployment, effectiveness, and/or scale-up of innovative resources, models and tools" (NSF, 2011, p.2). As such, it is particularly important for the projects within this portfolio to use the soundest methods for testing the efficacy and ultimately effectiveness of the developed educational interventions. This compendium of measures is Part I of a two part series to provide insight into the measurement tools available to generate efficacy and effectiveness evidence, as well as understand processes relevant to teaching and learning. Part II will look at student outcome assessments. This work was undertaken through the Community for Advancing Discovery Research in Education (CADRE) learning resources network, which is funded by NSF to support DR K-12 grantees, raise external audiences' awareness and understanding of the DR K-12 program, and build new knowledge.[1] To provide support to grantees, CADRE has developed a website with program and project information, conducted principal investigator meetings, initiated a fellowship program for new researchers, and facilitated several communities of practice. The communities of practice are producing useful products and tools to advance and inform wider fields of study. Some of these have been developed for the DR K-12 community but have implications for work beyond this portfolio; others are intended for external audiences. CADRES' thematic studies that look across the work of individual DR K-12 projects help to build knowledge across projects and extend the program's contributions to the field beyond those made by individual projects. These studies take a comprehensive look at the DR K-12 portfolio of funded projects in order to understand the role that the program has played in advancing research on K–12 student and teacher learning of STEM disciplines. This compendium series represents one of these thematic portfolio studies. Here we present information gathered through a review of five cohorts of DR-K12 funded-projects' proposed instruments. These projects were funded from 2008 to 2012. We focused on instruments designed to assess teacher practices, pedagogical content knowledge, and content knowledge. This collection of instruments represents commonly used tools for gathering information about educational innovations in the U.S. given that the DR-K12 portfolio is the nation's largest STEM educational intervention research and development fiscal investment.

The purpose of this compendium is to provide an overview on the current status of STEM instrumentation commonly being used in the U.S and to provide resources useful to research and evaluation professionals. The information contained within is heavily dependent on information available on existing websites. For each

[1] CADRE's partner organizations include the Education Development Center, Inc., Abt Associates, Inc., and Policy Studies Associates, Inc.

instrument we provide information on the constructs/variables that are measured, the target audience of the instrument, the subject domains assessed, information on obtaining the instrument and related documents about reliability and validity evidence when it could be located. While the information about reliability and validity evidence is provided, we highly recommend that anyone intending to use an instrument consult the Standards for Educational and Psychological Testing published jointly by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (1999), to ensure proper deployment, piloting and evidence accumulation for a particular research or evaluation application and study population. These standards are in the process of being revised as of August 2012.

Methods

The driving research question for this instrument review was: *What are the instruments, constructs, and methods being used to study teacher outcomes within the DR-K12 portfolio*? The research team decided to include only extant, named instruments as opposed to instruments being developed as part of a current grant proposal. This decision allows for the information in this review to reflect assessment tools currently accessible to researchers, and thus can contribute to knowledge building across studies of similar learning and teaching phenomena. Additionally, if an instrument is already in existence, it stands the most likelihood of having psychometric information generated across multiple settings, which is a fairer assessment of the technical quality of the tool. Three commonly assessed teacher outcomes were the target constructs for this review—teacher practices, pedagogical content knowledge (PCK), and content knowledge.

The review process was conducted in two phases. The first phase included reviewing all available proposals funded by the DR-K12 program since 2008. This netted 295 eligible projects. Additional materials such as annual reports, publications, products, etc., where available, were reviewed as well, to extract the name of proposed teacher instruments and the constructs being measured. Once this initial dataset was constructed, a second phase of data collection was conducted for instrument-specific information about reliability and validity evidence, development and piloting, accessibility of the instrument, administration, and constructs measured. This information was gathered through internet searches with the name of the instrument as the keyword. Information provided by the developer of an instrument was given preference over other sources if there was conflicting information. In some cases, the instrument was restricted use, so requests were made to the developer for the instrument-associated documentation. There were some instances where multiple versions of an instrument are available in which case the latest version was included in the dataset. All data was entered into Excel then coded into descriptive categories so frequency counts could be generated.

Limitations to this data collection effort primarily relate to the sources of data. Since CADRE is funded as a cooperative agreement rather than a contract, we do not have direct access to Fastlane files and thus relied on project materials that were provided directly by the project Principal Investigators. There were 36 projects where the project materials were not available for our review representing an 11% missing data rate. Often PIs do not know exactly what they will end up using in the project, until the project is funded. For convenience we use phrases like "projects used," but in fact we only know what they proposed to use or consider, not what they ended up using in their studies.

Though we made a reasonable effort to obtain information about each instrument identified in proposals, for eight (14%) of the 57 instruments measuring PCK or practices we could not obtain copies of the actual instrument, and for an additional six (11%) purchasing was required. However, in most cases we were able to obtain some information about the constructs assessed on these measures and sometimes the psychometric properties. All notations regarding how to access the instruments and links to the supportive documents we were able to locate are contained in Appendices H and I.

Cross-Cutting Comparisons

Seventy-five projects (25% of the overall DR-K12 portfolio) proposed to measure teacher practices, pedagogical content knowledge, or content knowledge as an outcome of the funded work. Seventy-one percent of these projects measured only one teacher outcome (32 projects measured practice, 14 PCK, and 7 content); twenty-four percent measured two outcomes (4 PCK and content; 8 PCK and practice; 6 practice and content); and only 5% (4) measured all three types of outcomes. Across these 75 studies, 82 extant instruments were identified. The three most common instruments used for each outcome are listed in Table 1. For practices, a total of 42 instruments were identified, for PCK 24 instruments were identified, and for content knowledge 27 instruments were identified. Some instruments were described by the Principal Investigators as measuring more than one type of outcome.

Table 1: Number of studies that used the most frequently named instruments by construct

Instrument Name	Constructs		
	Practices	PCK	Content knowledge
Reformed Teaching Observation Protocol (modified)	15	2	1
Inside the Classroom Observation and Analytic Protocol	8		
Surveys of Enacted Curriculum (modified)	5	1	
Mathematical Knowledge for Teaching[2]	1	14	3
Knowledge of Algebra for Teaching (modified)		2	
Science Teaching Efficacy Belief Instrument		2	
Views of Nature of Science Form C			3
National Assessment of Educational Progress (modified)			3
Praxis content tests/ Earth & physical science (modified)			2
TIMSS content tests (modified)			2

Information gathered during the second phase of data collection provided additional details for a more fine-grained analysis of the substance and the psychometric evidence of the measurement tools. The types of reliability indicators that we captured include: internal scale consistency alpha; interrater agreement as Kappa, percent agreement, or Spearman rank-order correlations. The type of validity evidence included relates to content, construct, predictive, concurrent and discriminant data. With this information we were able to assess the strengths and weaknesses in the measurement landscape for key educational constructs. The 54 PCK and Practice instruments were further differentiated into five categories of instruments: instructional practices, instruction plus one or two other constructs, instructional beliefs, system-wide reform focused, and discourse focused. The instruments in each of these categories are profiled in Appendices A-G. Most of the content knowledge instruments were developed for students and then adapted to assess teacher knowledge (Appendix F). As such, they were frequently developed by psychometricians for large-scale administration, and had undergone rigorous development and testing. These instruments often assess multiple content areas and without knowing exactly which version of a test, year of administration, or part of the test that was extracted for a project, it would not be possible to provide precise information about reliability and validity. Therefore, they were not included in this more detailed cross-cutting analysis, but access details are provided for researchers to obtain this information for their own purposes in Appendix I. There were also three instruments that were unique and therefore are profiled in Appendix G, but not included in this cross-comparison of key educational constructs.

[2] Learning Mathematics for Teaching (LMT) is the name of the project, not an instrument. Content Knowledge for teaching Mathematics (CKT-M) and Mathematical knowledge for teaching (MKT) are the same and the current name is MKT. Therefore any study mentioning LMT, CKT-M, or MKT were counted here.

Instructional Practices

There were eleven instruments that primarily assessed classroom instructional practices (see Appendix A). These instruments (seven observation protocols, three rubrics—educational products are scored, one survey) were predominantly designed for pre-kindergarten through middle school teachers (6, 55%). There were slightly more focused on science (5, 45%) than mathematics (3, 27%) or technology (2, 18%). The science observation protocols (3) capture variables ranging from the lesson's temporal flow and percentage of time students spend in different types of groupings, to the extent of opportunity for students to engage in the various phases of the investigation cycle. The two science scoring rubrics are intended to be applied to lesson artifacts and instructional materials that the teacher provides students. They contain codes for student grouping, structure of lessons, use of scientific resources, hands-on opportunities through investigation, cognitive depth of the materials, encouragement of the scientific discourse community, and opportunity for explanation/justification, and connections/applications to novel situations. The one math scoring rubric describes the quality of instruction using proportion of lesson time spent on six dimensions: core mathematical ideas; representations matched to algorithms; conceptual and temporal links; elicitation of student thinking and teacher responsiveness; amount of student work contributed; and the kind of student work in the lesson. The two mathematics observation protocols for PK and PK-6 classrooms assess general aspects instruction such as the type and depth of the mathematics. The two technology instruments (one survey and one observation protocol) assess how technology is being used within the classroom context. One observation protocol for post-secondary classrooms was also identified; it captures the use of various instructional strategies to foster high level thinking skills such as metacognition and divergent thinking. Across these eleven instruments, one had low reliability evidence, and four (36%) had acceptable or good evidence. For only two instruments was the team able to find validity evidence (see Table 2).

Table 2. Number and percentage of instruments by focus, reliability, and validity indicators

Instrument Focus (n=54)	**Reliability Evidence Level* (%)**					**Validity Evidence Type****					
	Miss	Low	Accept	Good	N/A	Miss	Content	Const	Pred	Concur	Discr
Instructional Practices (11)	6 (55)	1 (9)	3 (27)	1 (9)	--	9 (82)	1 (9)	--	--	1 (9)	--
Instruction plus… (11)	6 (55)	--	2 (18)	3 (27)	--	7 (64)	2 (18)	2 (18)	--	1 (9)	--
Instructional Beliefs (9)	--	--	3 (33)	3 (33)	3 (33)	2 (22)	4 (44)	5 (56)	--	--	--
System-wide reform (10)	3 (33)	--	3 (33)	3 (33)	1 (11)	4 (44) 1 n/a	2 (22)	1 (11)	3 (33)	--	--
Discourse (13)	4 (31)	1 (8)	2 (15)	6 (46)	--	5 (38)	4 (31)	2 (15)	6 (46)	3 (23)	2 (15)

* Miss=missing; low ≤ 0.59; acceptable 0.60–0.79; good ≥ 0.80
**Miss=missing; Const=construct; Pred=predictive; Concur=concurrent; Discr=discriminant. An instrument may have generated evidence of more than one type of validity.

Instructional Practices plus Additional Constructs

There were eleven instruments that measured instructional practices in addition to one or two other constructs such as physical context, demographics, teacher content knowledge, or some aspect of classroom management (see Appendix B). This more comprehensive nature is also reflected in the subject domain—five of these observation instruments assess both mathematics and science; and one, more general teaching skills such as lesson planning and assessment. There are two each, mathematics-specific (survey, observation), and science-specific (surveys) instruments. There is one technology specific observation protocol. Five (45%) of these instruments had evidence of acceptable or good reliability and four provided validity evidence (36%) (see Table 2).

Instructional Beliefs

Nine instruments were identified as measuring instructional beliefs (eight surveys and one interview protocol) (see Appendix C). Six (67%) of these were developed for science, two for math and one was non-subject-specific. The science beliefs measured centered around self-efficacy in teaching in general, teaching science content, and teaching science investigation skills. The mathematics surveys similarly measured self-efficacy in various math domains and in teaching math. Sixty-seven percent of these instruments had evidence of either acceptable or good reliability, and 78% had demonstrated evidence of validity (see Table 2).

Multidimensional

This last set of 23 instruments that assess instructional practices or PCK, also assess a number of other constructs. There are two subsets of instruments— those that assess variables related to systemic reform efforts (see Table 3) and those that assess variables related to the discourse environment in the classroom (see Table 4).

System-wide Reform

The subset of ten instruments used to investigate the effect of education system reform efforts capture instructional practices, instructional beliefs, the administrative/policy context influencing instruction, student and teacher demographics, and the content being taught (see Table 3). Half of the instruments capture mathematics and science instruction, one is science-specific, three are mathematics-specific, and one is non-domain-specific (see Appendix D). In this set of instruments is where we saw English Language Arts-specific items included on two of the instruments. This set of instruments predominately employs survey administration, with 60% of the instruments using this approach. There are two observation protocols, one interview, and one rubric. Sixty-six percent of the instruments had acceptable or good reliability and 56% had validity evidence (see Table 2).

Table 3. Instruments assessing multiple dimensions related to system-wide reform efforts.

Acronym	Instrument Name	Constructs									
		Practices	Content	Beliefs	Management	Assessment	Social	Physical	Admin context	Demographic	planning
CIP	Inside the Classroom Teacher Interview Protocol			✓			✓	✓	✓		
COP	Inside the Classroom Observation and Analytic Protocol	✓	✓		✓		✓	✓	✓		
CTRI	Coaching /Teacher Reflection Impact Surveys	✓		✓	✓					✓	
FFT	Danielson's Framework for Teaching Domains	✓	✓		✓	✓	✓	✓			✓
--	Inside the Classroom Teacher Questionnaire: Math or Science version	✓	✓	✓	✓					✓	
LSC	LSC Core Evaluation Classroom Observation Protocol	✓	✓				✓				✓
SEC	Surveys of Enacted Curriculum	✓		✓		✓			✓	✓	
SII	Study of Instructional Improvement	✓	✓	✓					✓	✓	
TIMSS-R	TIMSS-R Science Teacher Questionnaire			✓					✓	✓	✓
TIMSS-R	TIMSS-R Mathematics Teacher Questionnaire			✓					✓	✓	✓

Classroom Discourse Community

The second subset of 13 instruments measuring multiple constructs looks at instructional practices as well as the social aspects of the classroom community, including classroom management (see Table 4). All of these instruments are observation protocols and four additionally have a survey or scoring rubric component (see Appendix E). Fourth-six percent of the instruments (6) are non-domain-specific and instead focus on assessing the teacher-student interaction, verbal discourse, and emotional support exhibited in the classroom. There are four instruments in this set that have English Language Arts-specific items. Three instruments measure mathematics-specific discourse, three measure science-specific discourse, and one measures both. Sixty-two percent of these instruments (8) have acceptable or good reliability and evidence of validity. Fifty-four percent of the instruments had demonstrated evidence of more than one type of validity—more than any other category of instruments (see Table 2).

Table 4. Instruments assessing multiple dimensions related to classroom discourse environment.

Acronym	Instrument Name	Constructs			
		Practices	Management	Social	Physical
CLASS	The Classroom Assessment Scoring System	✓	✓	✓	
--	Classroom Snapshot	✓	✓	✓	✓
CLO	Classroom Lesson Observation Instrument	✓	✓	✓	
COEMET	Classroom Observation of Early Mathematics Environment and Teaching	✓	✓	✓	
DAISI	The Dialogic Activity in Science Instruction	✓		✓	
EAS	The Emergent Academic Snapshot	✓	✓	✓	
ELLCO	The Early Language and Literacy Classroom Observation	✓	✓		✓
IQA	Instructional Quality Assessment	✓		✓	
ISIOP	Inquiring into Science Instruction Observation Protocol	✓	✓	✓	✓
--	Mathematics Classroom Observation Protocol	✓		✓	
RTOP	Reformed Teaching Observation Protocol	✓		✓	
--	Science Classroom Observation Guide (NCOSP)	✓		✓	
SPC	Standards performance continuum	✓		✓	

Reliability and Validity Evidence

In assessing this collection of 54 instruments for reliability and validity evidence overall, we found a rather alarming level of missing information. For reliability evidence 38% (19/50) of the eligible instruments (4 were n/a) have missing information (see Tables 2 & 3); for validity evidence 51% (27/53) have missing information (1 was n/a). In Table 5 we see that by method type, namely the low frequency of interview and rubric instruments, dramatically influences the percentage of missing information. Therefore, for users of these types of protocols, it is particularly important to obtain instrument-specific information prior to deployment and to pilot them with your own study participants. In comparing observation protocols to surveys, there was proportionally more missing information for the observation protocols.

Table 5. Number and percentage of instruments measuring PCK and practice constructs for each method type by reliability and validity indicators

Method type*	Reliability Evidence Level** (%)					Validity Evidence Type***						
	Miss	Low	Accept	Good	N/A	Miss	Cont	Const	Pred	Concur	Discr	N/A
Observation (n=29)	9 (31)	1 (3)	8 (30)	11 (38)	--	15 (52)	8 (28)	4 (14)	6 (21)	4 (14)	2 (7)	--
Interview (n=2)	1 (50)	--	--	--	1 (50)	2 (100)	--	--	--	--	--	--
Rubric (n=4)	3 (75)	1 (25)	--	--	--	3 (75)	--	--	--	1 (25)	--	--
Survey (n=18)	5 (28)	--	5 (28)	5 (28)	3 (17)	6 (33)	5 (28)	6 (33)	3 (17)	--	--	1 (6)

*one instrument unobtainable and unable to determine method type from existing descriptions.
** Miss=missing; low ≤ 0.59; acceptable 0.60–0.79; good ≥ 0.80
*** Miss=missing; Cont=content; Const=construct; Pred=predictive; Concur=concurrent; Discr=discriminant

The reliability and validity information that was available for this sample of instruments, indicates that there is a greater proportion of observation protocols with higher reliability levels (in the good range being 0.80 or higher) than surveys (38% vs. 28%). In looking at the balance of evidence by instrument foci, stronger evidence of multiple types of validity using a single instrument exists for protocols assessing discourse variables and instructional beliefs than the other three categories of foci. Overall, thirteen of the 50 eligible instruments (26%) had evidence of acceptable levels of reliability (range of 0.60-0.79) and sixteen (32%) had good levels. In terms of validity evidence across the fifty-three eligible instruments, thirteen (23%) had addressed content validity, ten (19%) construct validity, nine (17%) predictive validity, five (9%) concurrent validity, and two (4%) discriminant validity.

Conclusion

This review of the state of measurement tools for STEM educational interventions indicates that as a community of scholars, we need to be more cognizant about providing relevant psychometric information on the tools we develop and use. Without the basic information about what is needed to achieve an acceptable level of interrater reliability, users of these observation protocols, interview protocols, and scoring rubrics do not have the necessary information to make informed choices about the implementation of these tools in their own work. Information about survey scale coherence, as well as content and construct validity is essential to move the field forward in reaching a community consensus on operational definitions of key outcome variables. Policy makers need our tools to provide predictive, concurrent and discriminant validity evidence so that informed decisions about the efficacy and effectiveness of interventions can be made soundly. We unfortunately found that just over half of the instruments identified in this Compendium have evidence of acceptable or good levels of reliable implementation and scale consistency, and less than a third have associated validity evidence—indicating that there is a good deal of work yet to accomplish.

We hope that this Compendium can serve as an initial step towards the systematic assessment and improvement of our STEM research tools. We view this document as a first step rather than and ending point, and as such, if there is additional information that you are aware of that would be useful to include in this compendium, please send it to Daphne Minner for consideration in possible future revisions of this Compendium or other publications building on this work: Daphne_Minner@abtassoc.com. Any corrections to the information contained within this Compendium along with supporting documentation are also welcomed.

This Compendium can serve as a resource for colleagues to reflect and discuss issues and hurdles related to instrument development and deployment. One initial conversation took place at the DR-K12 Principal Investigator meeting in April of 2012 sponsored by CADRE. During this discussion, colleagues raised several unresolved issue that can serve as an initial starting point for dialogue.

With the nearing release of the Next Generation Science Standards, and with the Common Core Standards for Mathematics and the National Educational Technology Standards already in place, there is a need to realign many existing tools developed prior to, or on previous versions of Standards. It is particularly difficult to obtain funding for revision and realignment of existing instruments, yet this is essential if we are to have tools to determine the full complement of outcomes associated with Standards adoption. During this revision process, it would be an opportune time to reassess and release reliability and validity evidence needed to further build our instrument infrastructure in STEM.

Another issue related to alignment was voiced by colleagues at the meeting. They wanted to know more about the alignment between what is measured on the designated instrument, the intervention being studied, or a phenomenon under study. Typically issues of alignment are assessed during the peer review process, which is too late for making any adjustments. Hopefully, by providing the information in this Compendium about a range of instruments, colleagues who serve as project advisors can provide more timely input about instrument alignment and selection, even if they are not familiar with a particular tool.

In the next section we have provided reference to additional resources on the development, testing, and use of various research tools, other instrument libraries to consult, and references from the text of this Compendium.

Additional Resources

The resources listed in this section were accumulated during the second phase of work—the process of locating specific instruments and supporting documentation. We did not complete a comprehensive search for all relevant STEM instrument resources.

Useful Guides Related to Instrumentation

The resources in this section provide methodological guidance related to developing or using certain types of instruments. There are also some other synthesis reports related to specific areas of study such as evaluating teacher effectiveness, or comparisons between specific instruments.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. http://www.apa.org/science/programs/testing/standards.aspx

Derry, S. (Ed.) (2007). *Guidelines for Video Research in Education: Recommendations from an Expert Panel*. Retrieved from University of Chicago, Data Research and Development Center website: http://drdc.uchicago.edu/what/video-research-guidelines.pdf

Bill & Melinda Gates Foundation. (2012). *Gathering Feedback for Teaching: Combining high-quality observations with student surveys and achievement gains*. Retrieved from http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf

Meehan, M., Cowley, K., Finch, N., Chadwick, K., Ermolov, L., Joy, M., & Riffle, S. (2004). *Special strategies observation system-revised: A useful tool for educational research and evaluation*. Retrieved from AEL website: http://www.edvantia.org/products/pdf/04SSOS-R.pdf

Gallagher, C., Rabinowitz, S., & Yeagley, P., (2011). *Key considerations when measuring teacher effectiveness: a framework for validating teachers' professional practices*. Retrieved from Assessment and Accountability Comprehensive Center website: http://www.aacompcenter.org/cs/aacc/view/rs/26517

Hill, H. & Herlihy, C., (2011). *Prioritizing teaching quality in a new system of teacher evaluation*. American Enterprise Institute for Public Policy Research website: http://www.aei.org/article/education/k-12/teacher-policies/prioritizing-teaching-quality-in-a-new-system-of-teacher-evaluation/

Henry, M., Murray, K., & Phillips, K., (2007). *Meeting the challenge of STEM classroom observation in evaluating teacher development projects: A comparison of two widely used instruments*. Retrieved from: http://www.mahenryconsulting.com/pdf/RTOP%20ITC%20%20article%20Final%20120707.pdf

Existing Instrument Directories Relevant to STEM

This collection of instrument directories contains measures that assess more global constructs related to human development. These may be particularly useful for investigators and practitioners interested in developmental phenomenon as well as emotional and affective variables for students and teachers. There are also resources for specific kinds of educational or intervention contexts such as after-school settings and day care settings.

Halle, T., & Vick, J. E. (2007). *Quality in Early Childhood Care and Education Settings: A Compendium of Measures*. Washington, DC: Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Retrieved from: http://www.childtrends.org/Files/Child_Trends-2007_12_10_FR_CompleteCompendium.pdf

Halle, T., Vick Whittaker, J. E., & Anderson, R. (2010). *Quality in Early Childhood Care and Education Settings: A Compendium of Measures, Second Edition*. Washington, DC: Child Trends. Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Retrieved from: http://www.childtrends.org/Files/Child_Trends-2010_03_10_FR_QualityCompendium.pdf

This document, developed by The Forum for Youth Investment, is an excellent source for readers seeking information about measures that assess after-school and youth program quality. Yohalem, N. and Wilson-Ahlstrom, A. with Fischer, S. and Shinn, M. (2009, January). *Measuring Youth Program Quality: A Guide to Assessment Tools, Second Edition*. Washington, D.C.: The Forum for Youth Investment. Retrieved from: http://forumfyi.org/files/MeasuringYouthProgramQuality_2ndEd.pdf

The Program in Education, Afterschool & Resiliency (PEAR) located at McLean Hospital and Harvard Medical School has a searchable database of assessment tools for informal science programming can be retrieved from: http://www.pearweb.org/atis

The American Psychological Association provides useful information on existing directories of psychological tests for constructs such as attitudes, cognitive skills, personality traits, etc. This site is a good first stop if you are trying to measure intrapersonal constructs. The website is: http://www.apa.org/science/programs/testing/find-tests.aspx#

A collection of teacher efficacy scales provided by Anita Woolfolk Hoy at Ohio State University, with reliability and validity information provided. http://people.ehe.osu.edu/ahoy/research/instruments/#Sense

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., and Mooney, K. (2011). *Measuring student engagement in upper elementary through high school: a description of 21 instruments*. (Issues & Answers Report, REL 2011–No. 098). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs.This review's direct link: http://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL_2011098.pdf

Existing STEM Instrument Directories

This set of resources includes links to other collections of STEM instruments and related information.

The ITEST Learning Resource Center has compiled information on various instruments to help researchers, evaluators, and practitioners, identify and locate instruments used to assess learning and other related outcomes in STEM learning environments. This searchable database can be found at: http://itestlrc.edc.org/STEM_education_instruments

The Math and Science Partnership Network (MSPnet) is an electronic learning community for the NSF Math and Science Partnership Program. The resources, including instruments can be retrieved from: http://hub.mspnet.org/index.cfm/resources

The Mathematics Assessment Project (MAP) is working to design and develop well-engineered assessment tools to support US schools in implementing the Common Core State Standards for Mathematics (CCSSM). Their assessment items can be located at http://map.mathshell.org/materials/tests.php

The Database of Teachers' Mathematics/Science Content Knowledge is a searchable instrument database that provides researchers and practitioners with information about measures of teacher content knowledge used in empirical research, including research conducted by MSP projects. It currently only contains instruments that yield a numeric score. This database can be accessed at: http://www.mspkmd.net/instruments/index.php

The Field-tested Learning Assessment Guide (FLAG) for science, math, engineering and technology instructors is a website that provides discipline-specific instruments for college-level students. It is searchable by discipline and can be accessed at: http://www.flaguide.org/tools/tools.php

The tests available on this web site http://www.cfa.harvard.edu/smgphp/mosart/aboutmosart_2.html were developed by a team of researchers in the Science Education Department of the Harvard-Smithsonian Center for Astrophysics. The content of the questions is based on published studies of science misconceptions and the NRC National Science Education Standards. This web site was developed to make MOSART's assessment instruments available to individuals involved in science education. The tests are free and can be accessed after completion of four online tutorials that explain test design, use, scoring, and interpretation of results. Each MOSART assessment instrument comprises a set of multiple-choice items that are linked to the K–12 physical science and earth science content, and K–8 life science content in the NRC National Science Education Standards, as well as to the research literature documenting misconceptions concerning science concepts.

This is a collection of instruments across many domains of science. Somewhat dated, but provides references and is easy to locate information as a starting point.
http://cosmos.bgsu.edu/communities/research_community/MeasurementInst/pdfs/sample%20of%20classroom%20assessment%20instruments.pdf

References

GAO, *Science, Technology, Engineering, and Mathematics Education: Strategic planning needed to better manage overlapping programs across multiple agencies*, GAO-12-108 (Washington, D.C.: Jan. 20, 2012).

NSF (2011). Discovery Research K-12 (DR-K12): Program Solicitation 11-588.

NSTC (2011, December). *The Federal Science, Technology, Engineering, and Mathematics (STEM) Education Portfolio*. http://www.whitehouse.gov/sites/default/files/microsites/ostp/costem__federal_stem_education_portfolio_report.pdf

NSTC (2012, February). *Coordinating Federal Science, Technology, Engineering, and Mathematics (STEM) Education Investments: Progress Report*. http://www.whitehouse.gov/sites/default/files/microsites/ostp/nstc_federal_stem_education_coordination_report.pdf

Appendix A: Instruments to Determine Teacher Practices

Acronym	Name	Variables measured/ scales *Teacher Practices*	Type of tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
AFM	Assessment of the Facilitation of Mathematizing	Pedagogy, use of context, and knowledge of mathematics scales	Observation PK-6	Math	Interrater Kappa/acceptable
EMCO	Early Mathematics Classroom Observation	Mathematics instruction by preschool and kindergarten teachers	Observation PK	Math	
ETAP	EdTech Assessment Profile	Computer knowledge and skills (general knowledge and skills, internet skills, email skills, word processing skills, presentation software skills, spreadsheet software skills, database software skills), using technology in the classroom, using tech to support student learning, personal use, student use, staff development needs, technical support	Survey --	Technolog y	
ISCOP	Instructional Strategies Classroom Observation Protocol	Identifying sense of purpose; asking account of student ideas; engaging students with relevant phenomena; developing and using scientific ideas; promoting student thinking about phenomena, experiences, and knowledge	Observation MIDDLE	Science	Interrater % agreement /acceptable
LFCPO	Lesson Flow Classroom Observation Protocol	Lesson's temporal flow and student arrangement in terms of percent time in individual, small group, or large group settings	Observation MIDDLE	Science	
LoFTI	Looking for Technology Integration	Variety of broad areas of technology implementation and impact, in any setting where teaching and learning are taking place: the environment - arrangement, student grouping, and instructional collaborators; teaching and learning activities - content area, teacher, student, and assessment methods; student engagement; use of technology; hardware and software in use. Analysis of data provides a profile of technology use in the school as a whole, rather than for individual staff members.	Observation --	Technolog y	

Acronym	Name	Variables measured/ scales *Teacher Practices*	Type of tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
O-TOP	OCEPT-Classroom Observation Protocol	Reform-based strategies: habits of mind, metacognition, student discourse and collaboration, rigorously challenged ideas, student preconceptions and misconceptions, conceptual thinking, divergent thinking, interdisciplinary connections, PCK, multiple representations of concepts	Observation or planning tool POST-SEC	Science Math	Interrater % agreement /acceptable
—	The Quality of Instruction Measure	Describes quality of instruction using proportion of lesson time spent on six dimensions: core mathematical ideas; representations matched to algorithms; conceptual and temporal links; elicitation of student thinking and teacher responsiveness; amount of student work contributed; and the kind of student work in the lesson	Scoring rubric applied to teacher's ratings of video of instruction --	Math	
—	Scoop Notebook	Portfolio assessment that captures: grouping, structure of lessons, use of scientific resources, hands-on, inquiry, cognitive depth, scientific discourse community, explanation/justification, assessment, connections/applications	Artifact rubric MIDDLE	Science	Interrater % agreement /low CONCURRENT
STIR	Science Teacher Inquiry Rubric	Opportunity for learners to engage with a scientifically oriented question; teachers engage learners in planning investigations to gather evidence in response to questions; teachers help learners give priority to evidence which allows them to draw conclusions, develop and evaluate explanations; learners formulate conclusions and explanations from evidence; learners evaluate conclusions in light of alternative explanations; leaners communicate and justify their proposed conclusions	Observation ELEM	Science	Interrater % agreement /good CONTENT

Acronym	Name	Variables measured/ scales *Teacher Practices*	Type of tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
TIDES	Transforming Instruction by Design in Earth Science	Rubrics for analyzing the quality of teacher assignments focus on the opportunities to learn embedded in the materials selected. Raters characterize instructional materials with respect to five criteria: (1) opportunities for students to understand science as a dynamic body of knowledge that develops through investigation; (2) opportunities for students to engage in constructing new knowledge and applying content understandings to problem contexts different from the contexts in which they were introduced; (3) opportunities for students to obtain accurate content understandings; (4) opportunities for students to engage in multiple aspects of the inquiry process; and (5) opportunities for students to engage in scientific communication	Teacher assignment quality rubrics --	Science	

Appendix B: Instruments to Determine Teacher Practices Plus one or two Other Constructs

Acronym	Name	Variables measured/ scales *Instructional Practices Plus*	Type of Tool GRADE LEVEL	Subject Domain	Construct Assessed	Reliability type/ level VALIDITY EVIDENCE
CETP-COP	The Collaboratives for Excellence in Teacher Preparation core evaluation classroom observation protocol	Background information, classroom demographics, classroom context, class description and purpose (type of instruction, student engagement, cognitive activity), overall quality of instruction	Observation & interview K-16	Science Math	Instruction, physical context, demographic	Internal consistency alpha/ good Interrater Kappa/ acceptable
EQUIP	Electronic Quality of Inquiry Protocol	Descriptive information; time usage for: activity focus, organizational structure, student attention, cognitive levels displayed by students; inquiry instruction components: instructional strategies, order of instruction, teacher role, student role, knowledge acquisition; discourse factors: questioning level, complexity of questions, questioning ecology, communication pattern, classroom interactions; assessment: prior knowledge, conceptual development, student reflection, assessment type, role of assessing; curriculum factors: content depth, learner centrality, standards, organizing and recording information	Observation K-12	Science Math	Instruction, demographic	Internal consistency alpha/ good Interrater Kappa/ acceptable CONTENT CONSTRUCT

Acronym	Name	Variables measured/ scales *Instructional Practices Plus*	Type of Tool GRADE LEVEL	Subject Domain	Construct Assessed	Reliability type/ level VALIDITY EVIDENCE
ICOT	ISTE Classroom Observation Tool	Description of the classroom setting, student groupings, teacher roles during observation, general learning activities observed during the period (creating presentations, research, information analysis, writing, test taking, drill and practice, simulations, hands-on skill training), how essential tech. was to teaching and learning activities, checklist of specific technologies used, NETS teacher standards checklist, time estimates for technology use.	Observation K-12	Technolog y	Instruction, technology use	Interrater other/ chi square used to determine differences between 12 teachers rated 3 times by 3 observers. No significant differences found except for 5 technologies on checklist
KAT	Knowledge of Algebra for Teaching	Knowledge of middle and high school algebra; knowledge of advanced math (calculus, abstract algebra); teaching knowledge (common misconceptions, curriculum trajectories)	Survey, Assessment MIDDLE, HIGH	Math	Content PCK	
MQI	Mathematical Quality of Instruction	Previous version of this instrument was called the Quality of Mathematics in Instruction (QMI). Measures: format of lesson segment, classroom work with content, mode of instruction, richness of mathematics, working with students and mathematics, errors and imprecision, student participation in meaning-making and reasoning mathematics instructional quality	Video observation ELEM, MIDDLE	Math	PCK	Internal consistency alpha/ acceptable Interrater Kappa/ low CONSTRUCT
—	OH Middle Level Mathematics and Science Education Bridging Study - Teacher Questionnaire	"How I Teach" , "What My Students Do," "Knowledge of Polar Regions," subscales on pre and post-test; "Evidence of impact on students" post-test only	Survey K-12	Science	Instruction, demographic , content knowledge	

Acronym	Name	Variables measured/ scales *Instructional Practices Plus*	Type of Tool GRADE LEVEL	Subject Domain	Construct Assessed	Reliability type/ level VALIDITY EVIDENCE
PRAXIS	Praxis Teaching Foundations: Science	Knowledge of teaching foundation, including human development, addressing learning differences and special needs, working with ELL, building reading skills, assessing student progress, classroom management techniques, and teaching methods in science	Survey and open response MIDDLE, HIGH	Science	Instruction, class mgmt., assessment	
PRISM	Preschool Rating Instrument for Science and Mathematics	Comprehensive, 16-item instrument designed to measure the presence of classroom materials and teaching interactions that support both mathematics and science learning. The science items focus on materials and teaching interactions that support explorations of biological and non-biological science; encourage reading about, writing about, and representing science; encourage investigations and discussions of scientific concepts; support observing, predicting, comparing, and contrasting; and encourage recording of scientific information in journals, graphs, and other representational formats. In addition, items on measurement and classification cross the math and science domains.	Observation PK	Science Math	Instruction, physical context	Internal consistency alpha/ acceptable CONCURRENT
SESAME	Self-Evaluation of Science and Math Education	This is a companion to the PRISM materials assessment tool. It is designed for use by administrators and teachers to inform practice by participating in “reflective coaching cycles.”	UNCLEAR PK	Science Math	Coaching	
SIOP	Sheltered Instruction Observation Protocol	Preparation (lesson planning process, including the language and content objectives, the use of supplementary materials, and the meaningfulness of the activities); ELL Instruction (building background, comprehensible input, strategies, interaction, practice/application, and lesson delivery); Review/Assessment (whether the teacher reviewed the key vocabulary and content concepts, assessed student learning, and provided feedback to students on their output)	Observation & rating tool to match implementation of lesson delivery to a model of instruction for ELL students. ELEM	General	Instruction, planning	

Acronym	Name	Variables measured/ scales *Instructional Practices Plus*	Type of Tool GRADE LEVEL	Subject Domain	Construct Assessed	Reliability type/ level VALIDITY EVIDENCE
TIMSS	Third International Mathematics and Science Video Study	QUESTIONNAIRE on 7 dimensions of the videotaped lesson: larger unit or sequence of typicality of the videotaped lesson; ideas that guide teaching; educational background, teaching background, and teaching load; school characteristics; and attitudes about teaching. VIDEO (Science lessons) coded for 11 dimension: lesson length and phase structure; public talk; social organization structure; activity structure; activity function; learning environment; types of independent activities; coverage codes (science knowledge developed during the lesson); occurrence codes (science ideas supported with data, phenomena, and/or visual representations); science content of lesson; features of practical activities. VIDEO (math lessons) coded for: purpose; classroom routine; actions of participants; content; classroom talk; and classroom climate.	Observation with Survey MIDDLE	Math Science	Instruction, social context, demographic	Interrater % agreement /good CONTENT

Appendix C: Instruments to Determine Teacher Instructional Beliefs

Acronym	Name	Variables measured/ scales *Instructional Beliefs*	Type of tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
IMBS	Indiana Mathematics Beliefs Scale	Measures six beliefs dimensions: effort can increase mathematical ability; understanding concepts is important in mathematics; word problems are important in mathematics; there are word problems that cannot be solved with simple, step-by-step procedures; I can solve time-consuming mathematics problems; mathematics is useful in daily life	Student survey MIDDLE, POST-SECONDARY	Math	Internal consistency alpha/ acceptable CONTENT CONSTRUCT
MTEBI	Mathematics Teaching Efficacy Belief Instrument	Mathematics efficacy beliefs in preservice elementary teachers: Personal Mathematics Teaching Efficacy and Mathematics Teaching Outcome Expectancy	Survey PRESERVICE	Math	Internal consistency alpha/ good CONSTRUCT
PSI-T	Principles of Scientific Inquiry-Teacher*	Measures teacher and student perceptions of the frequency of occurrence when students are responsible for: framing research questions, designing investigations, conducting investigations, collecting data, and drawing conclusions *PSI-T is categorized as beliefs because it is self-report of classroom instructional activities, i.e., perception of frequency of doing certain types of investigation related instruction rather than actual frequency recording.	Survey (teacher and student versions) HIGH	Science	Internal consistency alpha/ good CONTENT CONSTRUCT
SETAKIST	Self-Efficacy Teaching and Knowledge Instrument for Science Teachers	Teaching efficacy and science knowledge efficacy	Survey ELEM	Science	n/a CONSTRUCT
STEBI	Science Teaching Efficacy Belief Instrument	Personal Science Teaching Efficacy (PSTE) subscale, which reflect science teachers' confidence in their ability to teach science. Science Teaching Outcome Expectancy (STOE) subscale, which reflect science teachers' beliefs that student learning can be influenced by effective teaching	Survey ELEM	Science	Internal consistency alpha/ acceptable

Acronym	Name	Variables measured/ scales *Instructional Beliefs*	Type of tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
TBI	Teacher Belief Interview	(learning) 1. How do you maximize student learning in your classroom? 3. How do you know when your students understand? 6. How do your students learn science best? 7. How do you know when learning is occurring in your classroom? (knowledge) 2. How do you describe your role as a teacher? 4. In the school setting, how do you decide what to teach and what not to? 5. How do you decide when to move on to a new topic in your classroom?	Semi-structured interview HIGH	Science	n/a --
TSES	Teachers' Sense of Efficacy Scale	Efficacy in student engagement, instructional practices, classroom management	Survey PK-6	General	Internal consistency alpha/ good CONSTRUCT
TSI	Teaching Science as Inquiry	Assesses preservice teachers self-efficacy beliefs in relation to the teaching of science as inquiry	Survey ELEM	Science	Internal consistency alpha/ acceptable CONTENT
VNOS-C	Views of Nature of Science Form C	Students' views about several aspects of the nature of science: empirical; tentative; inferential; creative; theory-laden; social and cultural; myth of the "Scientific Method"	Survey—for students --	Science	n/a CONTENT

Appendix D: Instruments to Assess System-wide Reform Efforts

Acronym	Name	Variables measured/ scales *System-wide Reform Efforts*	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
CIP	Inside the Classroom: Teacher Interview Protocol	Teachers' perceptions of the factors that influenced the selection of lesson content and pedagogy. Influences considered include: standards documents, accountability systems, administrative support, teacher professional development, teacher beliefs about the content and about students, and the physical environment. Results of the interview are used to complete the second half the Observation and Analytic Protocol.	Interview K-12	Science Math	
—	Inside the Classroom Teacher Questionnaire	Teacher opinions on topic; teacher background and demographics; course/class information; student demographics and class composition; student objectives for courses, practices, class routines	Survey K-12	Math Science	
COP	Inside the Classroom: Observation & Analytic Protocol (Part 1)	Quality of an observed K-12 science or mathematics classroom lesson by examining the design, implementation, mathematics/science content, and culture of that lesson. The second half of the protocol allows researchers to rate various influences that shaped the selection of lesson content and pedagogy, as inferred from a post-observation interview with the teacher.	Observation K-12	Science Math	Internal consistency alpha/ acceptable
CTRI	Coach/Teacher Reflection Impact Surveys	Nature of math instructional coaching relationship between the coach (one instrument) and the teacher (another instrument). They both have the same parallel items: frequency/duration of coaching sessions, topics discussed (math content, math concept and inquiry, classroom environment/culture, reflection and planning, likely impact of coaching on instruction. Coaching instrument factors: student centeredness discussions, mathematics pedagogy discussions, coaching relationship, content discussions, impact of coaching. Teacher instrument factors: topics discussed, coaching relationship, coaching impact	Surveys K-8	Math	Internal consistency alpha/ good CONSTRUCT
FFT	Danielson's Framework for Teaching Domains	Planning and preparation; classroom environment; instruction; professional responsibilities. Rated on a scale of unsatisfactory, basic, proficient, distinguished	Scoring rubric K-12	General	

Acronym	Name	Variables measured/ scales *System-wide Reform Efforts*	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
LSC	Local Systemic Change Classroom Observation Protocol	Overall quality of the observed lesson; lesson design; lesson implementation; math/science content; classroom culture; likely impact on students' understanding	Observation K-12	Science Math	Internal consistency alpha/ good Interrater % agreement/ good CONTENT
SEC	Survey of Enacted Curriculum	School and class descriptions; instructional practices(classroom activities, problem-solving, hands-on, small group work, use of assessments, use of homework, influences on curriculum); subject content (topic coverage, level of cognitive demand); expectations for students; teacher characteristics (perceived readiness, professional development, school conditions, teacher beliefs)--generates maps that depict alignment between the state standards, student performance assessments, curriculum being taught	Survey K-12	Science Math ELA	Internal consistency alpha/ good CONTENT PREDICTIVE
SII	Study of Instructional Improvement: Teacher Questionnaire	Teachers' perspective of the school and its faculty; teaching practices and priorities; pedagogical content knowledge; their experiences with school improvement efforts, professional development opportunities, demographic information, and their professional background	Survey ELEM	Math ELA	n/a—depends on specific version used
TIMSS-R	TIMSS-R Science Teacher Questionnaire - Main Survey	Science teachers' academic and professional backgrounds, instructional practices, and attitudes towards teaching science	Survey ELEM, MIDDLE	Science	Internal consistency alpha/ acceptable PREDICTIVE
TIMSS-R	TIMSS-R Mathematics Teacher Questionnaire - Main Survey	Math teachers' academic and professional backgrounds, instructional practices, and attitudes towards teaching math	Survey ELEM, MIDDLE	Math	Internal consistency alpha/ acceptable PREDICTIVE

Appendix E: Instruments to Determine Multiple Constructs related to STEM Discourse

Acronym	Name	Variables measured/ scales *STEM Discourse*	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
CLASS	The Classroom Assessment Scoring System	Teacher's sensitivity, quality of instruction across all academic areas, and classroom management. It assesses 10 domains of teacher-child interaction that form three subscales: (1) emotional support: (a) positive climate, (b) negative climate, (c) teacher sensitivity, (d) regard for children's perspectives); (2) classroom organization: (a) behavior management (proactive, non-disruptive, reinforcing positive behavior), (b) productivity (efficient use of time), (c) instructional learning formats (teacher enabling of children's experience, exploration and manipulation of materials); and (3) instructional support: (a) concept development, (b) quality of feedback, (c) language modeling). A tenth domain is child engagement.	Observation K-12	General	Internal consistency alpha/ good Interrater % agreement/ good CONTENT CONCURRENT PREDICTIVE
—	Classroom Snapshot	Active instruction (reading aloud, instruction/demonstration/lecture, discussion, practice/drill, projects/kinesthetic); passive instruction (silent seatwork, copying); organization and management (verbal instructions, other); off task activities for student and teacher (social interaction, student uninvolved, being disciplined, classroom management, teacher social interaction/teacher uninvolved, teacher management, teacher out of the room); materials present in classroom is also recorded	Observation K-12	General	
CLO	Classroom Lesson Observation Instrument	Development of learning objectives; selection and use of instructional materials; educational climate for learning; variety of instructional activities; preparation for class session; instructional methods; opportunity for student participation; individualization of instruction; responsiveness to student feedback; learning difficulties accommodated	Observation POST-SECONDARY	General	

Acronym	Name	Variables measured/ scales *STEM Discourse*	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
COEMET	Classroom Observation of Early Mathematics Environment and Teaching	Classroom elements, classroom culture (environment and interaction; personal attributes of the teacher), specific math activities (math focus; organization, teaching approaches, interactions; expectations; eliciting children's solution methods; supporting children's conceptual understanding; extending children's math thinking; assessment and instruction adjustment)	Observation PK-6	Math	Internal consistency alpha/ good Interrater % agreement/ good CONTENT PREDICTIVE
DAISI	The Dialogic Activity in Science Instruction	The DAISI rubric includes 14 sub-themes: Joint Productive Activity (collaboration, authority, production); language and literacy (authentic science literacy, tool for learning, science discourse, science vocabulary, primary language); contextualization (personal-home-community experiences, local ecological environment); challenging activities (complexity of concepts, feedback & inquiry); instructional conversation (initiation & questioning, uptake & follow-up)	Observation ELEM	Science ELA	Interrater % agreement/ good
EAS	The Emergent Academic Snapshot)	Children's activity setting; children's engagement with academic activities; peer play scale; teacher engagement of the children, including codes for seven kinds of instructional strategies	Observation PK-6	General	Interrater Kappa/ acceptable CONCURRENT PREDICTIVE
ELLCO	The Early Language and Literacy Classroom Observation	The functional environment: organization of the classroom, contents of the classroom, presence and use of technology, opportunities for child choice and initiative; The interactive environment: classroom management strategies, classroom climate; language and literacy facilitation: opportunities for oral language use, presence of books, book reading practices, reading instruction, approaches to children's writing, writing opportunities and instruction; broad support for literacy: approaches to curriculum, recognizing diversity in the classroom, facilitation home support for language & literacy, approaches to assessment	Observation & interview PK-6	Literacy	Internal consistency alpha/ acceptable Interrater % agreement/ good DISCRIMINANT PREDICTIVE

Acronym	Name	Variables measured/ scales *STEM Discourse*	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
IQA	Instructional Quality Assessment	Accountable talk (accountability to learning community, to knowledge, to rigorous thinking); academic rigor for reading comprehension, math; clear expectations/self-management of learning (clarity and detail of expectations, access to expectations, understanding of expectations, judging work based on expectations, revising work based on expectations, rigor of expectations)	Observation & assignment rubrics ELEM	Math ELA	Interrater % agreement/ low DISCRIMINANT
ISIOP	Inquiring into Science Instruction Observation Protocol	Teacher's verbal practices (teacher sense-making, student sense-making, content storyline); lesson structure; lesson science content; investigation experiences; classroom management (general teaching style, teacher support for self-directed learning, lesson organization, dealing with distractions)	Observation & survey MIDDLE, HIGH	Science	Interrater % agreement/ acceptable Interrater Kappa/ acceptable CONTENT CONSTRUCT
—	Mathematics Classroom Observation Protocol	Intellectual support; depth of knowledge and student understanding; mathematical analysis; mathematics discourse and communication; student engagement; academic language support for ELLS; funds of knowledge/culture/community support; use of critical knowledge/social justice	Observation ELEM, MIDDLE	Math ELA	
RTOP	Reformed Teaching Observation Protocol	Inquiry instruction: lesson design and implementation, content (propositional and procedural knowledge), classroom culture (communicative interactions, student/teacher relationships)	Observation K-16	Science Math	Best-fit linear regression of one set of observation on another/ good CONTENT CONSTRUCT PREDICTIVE

Acronym	Name	Variables measured/ scales *STEM Discourse*	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
—	Science Classroom Observation Guide (by NCOSP)	Components, elements, and indicators of effective science teaching that can be used by administrators and teachers to develop a shared understanding of quality science classrooms and to collaboratively identify targets for growth. Specifically, classroom culture, science content, and instructional strategies are included.	Observation checklist for instructional dialogue K-12	Science	
SPC	Standards Performance Continuum	Collaboration, language use, connected learning, cognitive complexity, student-teacher dialogue	Observation ELEM, MIDDLE	General	Interrater Spearman rank-order correlations/ good CONCURRENT PREDICTIVE

Appendix F: Survey Instruments to Determine Teacher Content Knowledge

Acronym	Name	Variables measured/ scales *Content Knowledge*	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
MKT	Mathematical Knowledge for Teaching	Problems on this instrument reflect real mathematics tasks teachers face in classrooms - for instance, assessing student work, representing numbers and operations, and explaining common mathematical rules or procedures. Subscales include: number and operations; geometry; patterns, functions, and algebra. Previous version of this instrument was called Content Knowledge for Teaching Mathematics (CKT-M) that was generated out of the Learning Mathematics for Teaching project (LMT).	Survey ELEM, MIDDLE	Math	Internal consistency alpha/ acceptable CONTENT
MOSART	Misconceptions-Oriented Standards-Based Assessment Resources for Teachers	Student (or teacher) understanding of K-12 physical, earth, life science content	Survey K-12	Science	Item Response Theory used to construct the tests CONTENT CONSTRUCT
M-SCAN	The Mathematics Scan	Use of tasks (dimensions of structure of the lesson, cognitive demand, problem solving, and connections and applications); discourse (explanation and justification and mathematical discourse community); representations (multiple representations and students' use of mathematical tools); demonstrated knowledge(mathematical accuracy)	Video observation ELEM	Math	Interrater % agreement/ good DISCRIMINANT PREDICTIVE

Acronym	Name	Variables measured/ scales *Criterion-referenced Content Knowledge*	Type of Tool **GRADE LEVEL**	Subject Domain
—	American Chemical Society Division of Chemical Education Examinations Institute	Includes more than 50 exams covering general chemistry, organic chemistry, analytical chemistry, physical chemistry, inorganic chemistry, biochemistry, polymer chemistry, and high school chemistry	Teacher test POSTSECONDARY	Science
ACT	American College Testing	The ACT (No Writing) consists of four multiple-choice tests: English, Mathematics, Reading, and Science. The ACT Plus Writing includes the four multiple-choice tests and a writing test. In the Mathematics Test, three sub-scores are based on six content areas: pre-algebra, elementary algebra, intermediate algebra, coordinate geometry, plane geometry, and trigonometry. The content of the Science Test includes biology, chemistry, physics, and the Earth/space sciences (for example, geology, astronomy, and meteorology). The test emphasizes scientific reasoning skills over recall of scientific content, skill in mathematics, or reading ability. The scientific information is conveyed in one of three different formats: • **Data Representation.** This format presents graphic and tabular material similar to that found in science journals and texts. The questions associated with this format measure: skills such as graph reading, interpretation of scatterplots, and interpretation of information presented in tables, diagrams, and figures. • **Research Summaries.** This format provides descriptions of one or more related experiments. The questions focus on the design of experiments and the interpretation of experimental results. • **Conflicting Viewpoints.** This format presents expressions of several hypotheses or views that, being based on differing premises or on incomplete data, are inconsistent with one another. The questions focus on the understanding, analysis, and comparison of alternative viewpoints or hypotheses.	Student test POSTSECONDARY	General
AP	Advanced Placement	Multiple courses and tests. Science related tests: biology, physics, chemistry, environmental science. Mathematics related tests: calculus, statistics.	Student test HIGH	General
—	Classroom Test of Scientific Reasoning (Lawson)	The test is designed to examine a small set of dimensions including (1) conservation of matter and volume, (2) proportional reasoning, (3) control of variables, (4) probability reasoning, (5) correlation reasoning, and (6) hypothetical-deductive reasoning. These skills are important concrete components of the broadly defined scientific reasoning ability. Scores on test classify students into three groups: empirical/inductive reasoners, transitional reasoners, and hypothetical/deductive reasoners.	Student test --	Science

Acronym	Name	Variables measured/ scales *Criterion-referenced Content Knowledge*	Type of Tool GRADE LEVEL	Subject Domain
CST	Content Specialty Test Earth Science for New York Teacher Certification	Foundations of scientific inquiry; space systems; atmospheric systems; geological systems; water systems; geological systems: constructed-response assignment	Teacher test K-12 levels	Science
DTAMS -science	Diagnostic Science Assessments for Middle School Teachers	Three different assessments: physical; earth/space; life science. Each content assessment measures four types of knowledge: declarative knowledge, scientific inquiry and procedures, schematic knowledge, pedagogical content knowledge.	Teacher test MIDDLE level	Science
DTAMS -math	Diagnostic Mathematics Assessments for Middle School Teachers	Four different assessments: whole number/computation; rational number/computation; geometry/measurement; probability/statistics/algebra. Each content assessment measures four types of knowledge: memorized knowledge, conceptual understanding, problem-solving/reasoning, pedagogical content knowledge.	Teacher test MIDDLE level	Math
FACETS	Diagnoser Tools	Force and motion, waves, energy, properties of matter, chemistry, biology	Student test and teacher tools MIDDLE, HIGH	Science
FCI	Force Concept Inventory assessment	The Force Concept Inventory (FCI) instrument is designed to assess student understanding of the most basic concepts in Newtonian physics. This forced-choice instrument has 30 questions and looks at six areas of understanding: kinematics, Newton's First, Second, and Third Laws, the superposition principle, and types of forces (such as gravitation, friction). Each question offers only one correct Newtonian solution, with common-sense distractors (incorrect possible answers) that are based upon student's misconceptions about that topic, gained from interviews. The Inventory is not a test of intelligence; it is a probe of belief systems.	Student test HIGH, POSTSECONDARY	Science

Acronym	Name	Variables measured/ scales *Criterion-referenced Content Knowledge*	Type of Tool **GRADE LEVEL**	Subject Domain
GRE	Graduate Record Exam	**General test includes:** The Verbal Reasoning section measures ability to: (1) analyze and draw conclusions from discourse; reason from incomplete data; identify author's assumptions and/or perspective; understand multiple levels of meaning, such as literal, figurative and author's intent; (2) select important points; distinguish major from minor or relevant points; summarize text; understand the structure of a text; (3) understand the meanings of words, sentences and entire texts; understand relationships among words and among concepts. The Quantitative Reasoning section measures ability to: understand quantitative information; interpret and analyze quantitative information; solve problems using mathematical models; apply basic mathematical skills and elementary mathematical concepts of arithmetic, algebra, geometry, probability and statistics. The Analytical Writing section measures ability to: articulate complex ideas clearly and effectively; support ideas with relevant reasons and examples; examine claims and accompanying evidence; sustain a well-focused, coherent discussion; control the elements of standard written English. **Content test** are available for: biochemistry, cell & molecular biology, biology, chemistry, physics, computer science, and mathematics	Teacher test POSTSECONDARY	General
—	IL Certification Testing System Study Guide-Science: Biology	The test objectives for each of the science fields (i.e., biology, chemistry, Earth and space science, environmental science, and physics) contain a set of common objectives in addition to objectives unique to the specialty field. The set of common objectives measures the candidate's core knowledge across all science fields.	Teacher test K-12 levels	Science

Acronym	Name	Variables measured/ scales *Criterion-referenced Content Knowledge*	Type of Tool **GRADE LEVEL**	Subject Domain
ITBS	Iowa Test of Basic Skills	**Multiple domains assessed. Only math and science are articulated below.** Math—The tests emphasize the ability to do quantitative reasoning and to think mathematically in a wide variety of contexts. The tests at Levels 5 and 6 assess students' knowledge of beginning math concepts, focusing on numeration, geometry, measurement, and problem solving using addition and subtraction. All questions are presented orally; responses are pictures or numerals. At Levels 7 through 14, there are three separate tests. • The first is called Math Concepts at Levels 7 and 8 and Math Concepts and Estimation at Levels 9 through 14. This test requires students to demonstrate their understanding of fundamental ideas in the areas of number properties and operations, geometry, measurement, algebra, probability and statistics, and estimation. At Levels 9 through 14, the separately timed Estimation section tests mental arithmetic, number sense, and various estimation skills such as rounding. • The second test, called Math Problems at Levels 7 and 8 and Problem Solving and Data Interpretation at Levels 9 through 14, includes word problems that require one or more steps to solve. In many cases, students select an appropriate method or approach, rather than compute an answer. At Levels 9 through 14, several real-world "stories" form the basis for sets of three to four problems, each requiring somewhat different skills to solve. Levels 7 through 14 also include data displays such as tables and graphs. Students use them to obtain information, compare quantities, and determine trends or relationships. • Each problem in the third test, Math Computation, requires one arithmetic operation — addition, subtraction, multiplication, or division. The problems require operations with whole numbers, fractions, decimals, and various combinations of these, depending on the test level. Science—The tests at all levels assess not only students' knowledge of scientific principles and information but also the methods and processes of scientific inquiry. At Levels 7 and 8, all questions are presented orally and response choices are pictures. Included in Levels 9 through 14 are scenarios presenting real-life science investigations with questions emphasizing the thought processes used in designing and conducting research and in analyzing data. The four major content areas covered in the Science tests are: • Scientific inquiry — Methods of science; analysis and interpretation of data • Life science — Structures and life cycles of living things; environmental interactions • Earth and space science — Earth's composition, structure, and its changes; the universe • Physical science — Forces and motion; energy; properties of and changes in matter	Student test ELEM, MIDDLE	General

Acronym	Name	Variables measured/ scales *Criterion-referenced Content Knowledge*	Type of Tool GRADE LEVEL	Subject Domain
MAP	Missouri Assessment Program	**Grade-Level Assessments** are augmented norm-referenced tests that are delivered annually each spring in communication arts and mathematics for grades 3-8, and science for grades 5 and 8. **End-of-Course (EOC) Assessments** are criterion-referenced tests that are delivered to middle and high school students when the course-level expectations for a particular course have been covered. English II, algebra I, biology, and government are required EOC assessments for all students to satisfy the requirements of No Child Left Behind and the Missouri State Board of Education. Four other EOC assessments are optional: English I, algebra II, geometry, and American history.	Student test ELEM, MIDDLE	General
NAEP	National Assessment of Educational Progress	Public released items available for arts, civics, geography, reading, science, U.S. history, writing, **Mathematics**— Content: number properties and operations, measurement, geometry, data analysis and probability, algebra Ability: conceptual understanding, procedural knowledge, problem solving **Science**— Content: physical, earth and space, life science Practices: identifying and using science principles, using scientific inquiry, using technological design Knowing and doing science: scientific investigations, practical reasoning, conceptual understanding	Student test K-12	General
PISA	Program for International Student Assessment	Key subjects: reading, mathematics and science, with focus given to one subject in each year of assessment. In 2000 the focus of the assessment was reading, in 2003 mathematics and problem solving, in 2006 science, and in 2009 reading again. The 2012 data collection focusing on mathematics. See links in Appendix I to each year's assessment framework with additional details on what was measured.	Student test 15 year olds	General
PRAXIS	content tests	*Praxis II*® Subject Assessments measure knowledge of specific subjects that K–12 educators will teach, as well as general and subject-specific teaching skills and knowledge. There are many subject areas to choose from but Earth & physical science were named by a DR-K12 PI.	Teacher test K-12 levels	General

Acronym	Name	Variables measured/ scales *Criterion-referenced Content Knowledge*	Type of Tool GRADE LEVEL	Subject Domain
Regents	New York State Regents exam	Elementary (grades 3-5) and Intermediate (grades 6-8): English language arts, science, math, social studies. High school grades: English language arts, other languages, social studies, science (chemistry, earth science, living environment, physics), math (integrated algebra, geometry, algebra 2/trigonometry)	Student test K-12	General
SAT	Stanford Achievement Test	The SAT doesn't test logic or abstract reasoning. It tests the reading, writing and math skills learned in school. The **critical reading** section includes reading passages and sentence completions. The **writing** section includes a short essay and multiple-choice questions on identifying errors and improving grammar and usage. The **mathematics** section includes questions on arithmetic operations, algebra, geometry, statistics and probability. **Subject Tests** available for: multiple languages, literature, U.S. history, world history, math levels 1 and 2, biology, chemistry, physics.	Student test HIGH	General
TAGLIT	Taking a Good Look at Instructional Technology	Suite of online assessment tools designed to provide educational institutions effective data to evaluate technology use and integration in the teaching and learning environment. The suite includes assessments for school leaders, teachers, and students. The Teacher Basic TAGLIT assessment is divided into 10 sections: teacher information; your technology skill; your technology use in teaching and learning; technology and the way your classroom works; your school's technology resources – hardware; your school's technology resources - software and electronic/online references; your school's technology resources - technology and instructional support; your technology professional development; your school's technology plan; in your own words	Teacher test POSTSECONDARY	Technology

Acronym	Name	Variables measured/ scales *Criterion-referenced Content Knowledge*	Type of Tool GRADE LEVEL	Subject Domain
TIMSS	content tests	Grade 4 science and mathematics test items available. Science 1995: earth science; life science; physical science; and environmental issues and the nature of science Science 2003: earth science, life science Math 1995: fractions and proportionality; measurement, estimation and number sense; data representation, analysis, and probability; geometry; and patterns, relations, and functions. Math 2003: patterns and relationships; data; geometry; measurement; number Grade 8 science and mathematics test items available. Science 1999: earth science; life science; physics; chemistry; environmental and resource issues; and scientific inquiry and the nature of science Science 2003: chemistry; earth science; environmental science; life science; and physics Math 1995: fractions and number sense; algebra, measurement; geometry; and data representation, analysis, and probability Math 2003: algebra; data; geometry; measurement; and number	Student test ELEM, MIDDLE	Science Math
WEST-E	Washington Educator Skills Test-Endorsements	Multiple subject tests available. STEM related ones listed here. Science— Grades 4-9: middle level science. Grades 5-12: chemistry, earth/space science, environmental and sustainability education, physics, science, agricultural education. Math—Grades4-9: middle level math. Grades 5-12: mathematics Technology—Grades 5-12: technology education	Teacher test K-12 level	General
WESTEST	WESTEST	This is the West Virginia state student test that assesses reading and English, math, science, and social studies.	Student test GRADES 4-9	General

Appendix G: Unclassified Instruments

Acronym	Name	Variables measured/ scales *Unclassified*	Type of Tool GRADE LEVEL	Subject Domain	Construct Assessed	Reliability type/ level VALIDITY EVIDENCE
ASW	Analysis of Student Work	Measures a different facet of a teacher's analytic comparison of the student videos: (a) Sophistication of the teacher's analysis of the problem representation and solutions; (b) Teacher's inferences about students' ability and understanding of math concepts and algebraic thinking; (c) Teacher's inferences about students' mathematical development trajectory as a basis for pedagogical decision making; (d) Teacher's meta-cognitive reflection.	Teachers provide written analyses comparing a standardized set of video cases of student problem solving. Responses to ASW videos scored using rubrics of teachers' ability MIDDLE	Math	Assessment	Interrater Spearman rank-order correlations/ good CONTENT DISCRIMINANT
LoU	Levels of Use Interviews	Method for determining via interview how much and how well a change is actually being implemented in the classroom. Different implementation levels focus on a person's behaviors and skills with respect to their use of the innovation/program: nonuse, orientation, preparation, mechanical use, routine, refinement, integration, and renewal.	Interview --	General	Reform adoption	
SEPUP	Group Interaction and Communication of Scientific Information Rubrics	Group Interaction (time management, role performance/participation, shared opportunity); Communication (organizational; technical aspects)	Student work assessment MIDDLE	Science	n/a	

Appendix H: Alphabetical Listing (by acronym) of Practice or PCK Instruments and Access Details

Instrument Name	LINK	Access details
Assessment of the Facilitation of Mathematizing (AFM)	http://resources.curriculum.org/LNS/coaching/files/pdf/Mathematics_City.pdf	the scales are described on p.10-13 of the article.
Analysis of Student Work (ASW)	Instrument and findings from use: Derry, S. J., Wilsman, M. J., & Hackbarth, A. J. (2007). Using contrasting case activities to deepen teacher understanding of algebraic thinking, student learning and teaching. Mathematical Thinking and Learning, 9(3), 305-329. Reliability and validity information: https://mywebspace.wisc.edu/ajhackbarth/web/Papers/Hackbarth.Derry.Wilsman.ICLS%20Final%20Paper.06.pdf	
The Collaboratives for Excellence in Teacher Preparation core evaluation classroom observation protocol (CETP-COP)	http://www.cehd.umn.edu/carei/cetp/Handbooks/COPHandbook.pdf	A video guide is also available
Inside the Classroom Teacher Interview Protocol (CIP)	http://www.horizon-research.com/insidetheclassroom/instruments/ti.php	free to download on website
Inside the Classroom Teacher Questionnaire	http://www.horizon-research.com/insidetheclassroom/instruments/tq.php	Math and Science are separate protocols
The Classroom Assessment Scoring System (CLASS)	Information on constructs: http://www.acf.hhs.gov/programs/opre/hs/national_academy/reports/early_child_assess/erly_chld_p2_6.html The implementation guide from Teachstone is here: http://www.teachstone.org/wp-content/uploads/2010/06/CLASSImplementationGuide.pdf	K-3 version is available, 4-6 isn't available through the website publisher yet, but it is developed. COST
Classroom Snapshot	http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2009/02/11/000333037_20090211004247/Rendered/PDF/473280WP0Box33101OFFICIAL0USE0ONLY1.pdf	

Instrument Name	LINK	Access details
Classroom Lesson Observation Instrument (CLO)	http://www1.umn.edu/ohr/prod/groups/ohr/@pub/@ohr/documents/asset/ohr_46461.pdf	there are several other instruments on the Univ. Minnesota site http://www1.umn.edu/ohr/teachlearn/resources/peer/instruments/index.html
Classroom Observation of Early Mathematics Environment and Teaching (COEMET)	Need to obtain instrument from author—Doug Clements	
Inside the Classroom Observation and Analytic Protocol (COP)	http://www.horizon-research.com/insidetheclassroom/instruments/obs.php	free to download on website
Coach /Teacher Reflection Impact Surveys (CTRI)	Need to obtain instrument from author—James Burroughs	
The Dialogic Activity in Science Instruction (DAISI)	Constructs and reliability information available at: http://gse.berkeley.edu/research/crede/09%20Conference/TrishStoddart.pdf	Unable to obtain instrument
The Emergent Academic Snapshot (EAS)	Need to obtain instrument from author—Carollee Howes	
The Early Language and Literacy Classroom Observation (ELLCO)	https://secure.edc.org/publications/prodview.asp?1710	COST
Early Mathematics Classroom Observation (EMCO)	Unable to obtain background information in any one document.	Unable to obtain instrument
Electronic Quality of Inquiry Protocol (EQUIP)	http://iim-web.clemson.edu/wp-content/uploads/2009/02/equip-2009.pdf Information on reliability and validity: http://iim-web.clemson.edu/wp-content/uploads/2009/04/narst-2009-equip-proceedings-paper.pdf	additional publications: http://iim-web.clemson.edu/?page_id=168
EdTech Assessment Profile (ETAP)	http://www.edtechprofile.org/training/TechnologyAssessmentProfile.doc	

Instrument Name	LINK	Access details
Danielson's Framework for Teaching Domains (FFT)	http://www.andrews.edu/~rjo/Artifacts/Danielson's%20Framework%20for%20Professional%20Practice%20web.pdf	Additional information from the MET project: http://metproject.org/resources/Danielson%20FFT_10_29_10.pdf
ISTE Classroom Observation Tool (ICOT)	http://scottmcleod.org/2008%20-%20ISTE%20-%20ICOT%20Blank%20Form.pdf	Free online tool. Have to create an account and download software to get the tool and training documents.
Indiana Mathematics Beliefs Scale (IMBS)	This is a student instrument, that was administered at pre-post to preservice teachers in the DR-K12 study. Kloosterman, P. & Stage, F. (1992). Measuring beliefs about mathematical problem solving. *School Science and Mathematics*, *92*, 109-115.	
Instructional Quality Assessment (IQA)	http://www.cse.ucla.edu/products/reports/r671.pdf	
Instructional Strategies Classroom Observation Protocol (ISCOP)	Study with background information: http://www.gwu.edu/~scale-up/documents/AERA_2007_%20FOI_Symposium.pdf Table 1 in this document has the ISCOP items: http://www2.gwu.edu/~scale-up/documents/Seasons_Paper_APA_Fnl_Draft_033106.pdf Dissertation with instrument: http://proquest.umi.com/pqdlink?Ver=1&Exp=07-26-2017&FMT=7&DID=1400963841&RQT=309&attempt=1&cfc=1	Unable to obtain instrument
Inquiring into Science Instruction Observation Protocol (ISIOP)	http://isiop.edc.org/	to access videos for training you need to request an account
Knowledge of Algebra for Teaching (KAT)	http://www.educ.msu.edu/kat/	
Lesson Flow Classroom Observation Protocol (LFCPO)	http://www.gwu.edu/~scale-up/documents/AERA_2007_%20FOI_Symposium.pdf	Unable to obtain instrument

Instrument Name	LINK	Access details
Looking for Technology Integration (LoFTI)	http://www.fi.ncsu.edu/assets/LoFTI.pdf Daily tally sheet: http://www.serve.org/uploads/docs/LoFTIpaperpencilAnalysis.pdf	
Levels of Use Interviews(LoU)	Book that contains the instrument: Hall, G.E., & Hord, S.M. (2001). *Implementing change: Patterns, principles, and potholes*. Boston: Allyn & Bacon.	Need to go through training to be a certified interviewer.
LSC Core Evaluation classroom observation protocol (LSC)	Instrument: www.horizon-research.com/instruments/hri_instrument.php?inst_id=14 Reliability and validity tech report: http://www.horizon-research.com/LSC/news/cop_validity_2000.pdf Predictive validity tech report www.horizon-research.com/LSC/news/cop_validity_2005.pdf	This instrument is the precursor to HRI COP and CIP available on Horizon's website.
Mathematics Classroom Observation Protocol	http://mathconnect.hs.iastate.edu/documents/Aguirre.pdf	
Mathematical Quality of Instruction (MQI)	http://isites.harvard.edu/icb/icb.do?keyword=mqi_training "This is the same as QMI, but the current name is MQI"—Heather Hill	Have to request to get onto their training site.
Mathematics Teaching Efficacy Belief Instrument (MTEBI)	More Information about this instrument can be found at: http://onlinelibrary.wiley.com/doi/10.1111/j.1949-8594.2000.tb17256.x/abstract	
Ohio Middle Level Mathematics and Science Education Bridging Study - Teacher Questionnaire	Need to obtain instrument from author—Elizabeth Walker	
OCEPT-Classroom Observation Protocol (O-TOP)	http://ret.fsu.edu/Files/Tools/Appendix.C.pdf	
Praxis Teaching Foundations: Science (PRAXIS)	http://www.ets.org/Media/Tests/PRAXIS/pdf/0438.pdf	COST
Preschool Rating Instrument for Science and Mathematics (PRISM)	http://ecrp.uiuc.edu/v13n1/brenneman.html	COST

Instrument Name	LINK	Access details
Principles of Scientific Inquiry-Teacher (PSI-T)	Information on constructs is from: http://itec.macam.ac.il/portal/ArticlePage.aspx?id=1881&referer=useJsHistoryBack Campbell, T., Abd-Hamid, N., & Chapman, H. (2010). Development of instruments to assess teacher and student perceptions of inquiry experiences in science classrooms. *Journal of Science Teacher Education*, *21*(1), 13–30.	
The Quality of Instruction Measure	Kersting, N., Givvin, K. B., & Stigler, J. W. (2009). Capturing Teacher Knowledge: Exploring the Classroom Video-Analysis (CVA) Measure's Relationship to Teaching Quality and Student Learning. Paper presented at the Fourth Annual IES Research Conference. Kersting, N. (2008). Using Video Clips as Item Prompts to Measure Teachers' Knowledge of Teaching Mathematics. *Educational and Psychological Measurement*, 68:845-886.	Unable to obtain instrument
Reformed Teaching Observation Protocol (RTOP)	http://mathed.asu.edu/instruments/RTOP/index.shtml Instrument: http://mathed.asu.edu/instruments/RTOP/RTOPform_IN001.pdf Reference manual: http://www.public.asu.edu/~anton1/AssessArticles/Assessments/Biology%20Assessments/RTOP%20Reference%20Manual.pdf	
Science Classroom Observation Guide	http://www.ncosp.wwu.edu/Tools/index.php?toolID=4	
Scoop Notebook	Source: Martinez, J., Borko, H., & Stecher, B. (2012). Measuring instructional practices in science using classroom artifacts: Lessons learned from two validation studies. *Journal of Research in Science Teaching*, *49*(1), 38-67.	
Surveys of Enacted Curriculum (SEC)	http://seconline.wceruw.org/secWebReference.htm	COST
Group Interaction and Communication of Scientific Information Rubrics (SEPUP)	www.lhs.berkeley.edu/sepup/assess.html	

Instrument Name	LINK	Access details
Self-Evaluation of Science and Math Education (SESAME)	Information: http://ecrp.uiuc.edu/v13n1/brenneman.html Frede, Ellen; Stevenson-Garcia, J., & Brenneman, Kimberly. (2010). *Self-evaluation for science and math education (SESAME)*. New Brunswick, NJ: Author.	Unable to obtain instrument
Self-Efficacy Teaching and Knowledge Instrument for Science Teachers (SETAKIST)	http://www.eric.ed.gov/PDFS/ED448208.pdf	
Study of Instructional Improvement: several are available for language and math (SII)	Instruments: http://www.sii.soe.umich.edu/instruments/ User's guide: http://www.sii.soe.umich.edu/documents/SII%20Data%20User%27s%20Guide-1.pdf#page=7	
Sheltered Instruction Observation Protocol (SIOP)	http://gse.berkeley.edu/research/credearchive/tools/research/siop/1.3doc2.shtml	COST
Standards Performance Continuum (SPC)	http://gse.berkeley.edu/research/credearchive/tools/research/standards/spac.shtml	
Science Teaching Efficacy Belief Instrument (STEBI)	http://people.ehe.osu.edu/ahoy/files/2009/02/science-te.pdf	
Science Teacher Inquiry Rubric (STIR)	http://www.theibsc.org/uploaded/IBSC/Conference_and_workshops/2011/2011_Workshops/Handouts/Cathcart_Handout_Binder.pdf	
Teacher Belief Interview (TBI)	http://ret.fsu.edu/Files/Tools/Beliefs-EJSE[1].pdf	
Transforming Instruction by Design in Earth Science--Teacher assignment quality rubrics (TIDES)	http://ctl.sri.com/projects/displayProject.jsp?Nick=tides	Unable to obtain instrument
Third International Mathematics and Science Video Study (TIMSS)	http://nces.ed.gov/pubs2011/2011049.pdf (science); http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2003012 (math) Appendix D has the 200 pg codebook for the videos. There are also teacher questionnaires for 2007 on the site.	
TIMSS-R Science Teacher Questionnaire - Main Survey	http://timss.bc.edu/timss1999i/pdf/BM2_TeacherS.pdf; http://timssandpirls.bc.edu/methods/t-instrument.html	

Instrument Name	LINK	Access details
TIMSS-R Mathematics Teacher Questionnaire - Main Survey	http://timss.bc.edu/timss1999i/pdf/BM2_TeacherM.pdf	
Teachers' Sense of Efficacy Scale (TSES)	http://people.ehe.osu.edu/ahoy/files/2009/02/tses.pdf	
Teaching Science as Inquiry (TSI)	Background information: http://ret.fsu.edu/Files/Tools/TSI_article.pdf	Unable to obtain instrument
Views of Nature of Science Form C (VNOS-C)	Instrument: http://ret.fsu.edu/Files/Tools/VNOS(C)[1].pdf Background information: http://www.flaguide.org/tools/diagnostic/views_of_nature_questionnaire.php	

Appendix I: Alphabetical Listing (by acronym) of Content Knowledge Instruments and Access Details

Instrument Name	LINK	Access details
Diagnostic Science Assessments for Middle School Teachers (DTAMS)	Instrument: http://www.azed.gov/wp-content/uploads/PDF/DTAMSScienceInfo.pdf Information: http://louisville.edu/education/centers/crmstd/diag-sci-assess-middle	COST for scoring
Diagnostic Mathematics Assessments for Middle School Teachers (DTAMS)	http://louisville.edu/education/centers/crmstd/diag-math-assess-middle	COST for scoring
Mathematical Knowledge for Teaching (MKT)	http://sitemaker.umich.edu/lmt/files/LMT_sample_items.pdf Project site: http://sitemaker.umich.edu/lmt/home	Have to go through their training to get access to the instruments and they offer it only twice a year.
Misconceptions-Oriented Standards-Based Assessment Resources for Teachers (MOSART)	http://www.cfa.harvard.edu/smgphp/mosart/	free instruments that can be accessed after completion of four online tutorials that explain test design, use, scoring and interpretation of results. Videos case studies of student interviews included.
The Mathematics Scan (M-SCAN)	http://www.vsup.org/M-Scan%20Workshop%20Article%202%2009272011.pdf	Training required.

Instrument Name	LINK
American Chemical Society Division of Chemical Education Examinations Institute	http://chemexams.chem.iastate.edu/materials/exams.cfm
American College Testing (ACT)	http://www.act.org/
Advanced Placement (AP)	http://www.collegeboard.com/student/testing/ap/about.html

Instrument Name	LINK
Classroom Test of Scientific Reasoning (Lawson)--for students	http://www.public.asu.edu/~anton1/AssessArticles/Assessments/Mathematics%20Assessments/Scientific%20Reasoning%20Test.pdf Other content tests available at Lawson's website: http://www.public.asu.edu/~anton1/LawsonAssessments.htm
Content Specialty Test Earth Science for New York Teacher Certification (CST)	http://www.nystce.nesinc.com/PDFs/NY_fld008_prepguide.pdf http://www.nystce.nesinc.com/NY_viewSG_opener.asp
Diagnoser Tools (FACETS)	http://www.diagnoser.com/ Note: need to provide your school information to receive the password in order to get these assessments
Force Concept Inventory assessment (FCI)	http://modeling.asu.edu/r&e/fci.pdf Note: need to provide your school information to receive the password in order to get this assessment--it is free of charge and can be downloaded from link
Graduate Record Exam (GRE)	http://www.ets.org/gre/
IL Certification Testing System Study Guide-Science: Biology	http://www.icts.nesinc.com/PDFs/IL_field105_SG.pdf General information: http://www.icts.nesinc.com/PDFs/IL_SG_Generic_Front.pdf
Iowa Test of Basic Skills (ITBS)	http://www.riversidepublishing.com/products/itbs/
Missouri Assessment Program (MAP)	http://dese.mo.gov/divimprove/assess/tech/
National Assessment of Educational Progress (NAEP)	http://nces.ed.gov/nationsreportcard/itmrlsx/default.aspx

Instrument Name	LINK
Program for International Student Assessment (PISA)	http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html 2003 assessment framework: http://www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/33694881.pdf 2006 assessment framework: http://www.oecd.org/pisa/pisaproducts/pisa2006/assessingscientificreadingandmathematicalliteracy.htm 2009 assessment framework: http://www.oecd.org/pisa/pisaproducts/pisa2009assessmentframework-keycompetenciesinreadingmathematicsandscience.htm 2012 DRAFT assessment framework: http://www.oecd.org/pisa/pisaproducts/pisa2012draftframeworks-mathematicsproblemsolvingandfinancialliteracy.htm
PRAXIS content tests/ Earth & physical science (modified)	http://www.ets.org/praxis/about
New York State Regents exam (REGENTS)	http://www.nysedregents.org/
Stanford Achievement Test (SAT)	http://sat.collegeboard.org/home
Taking a Good Look at Instructional Technology (TAGLIT)	http://www.testkids.com/taglit/
TIMSS content tests	http://nces.ed.gov/timss/ Public released item bank: http://nces.ed.gov/timss/educators.asp
Washington Educator Skills Test-Endorsements (WEST-E)	http://program.pesb.wa.gov/add-new/endorsement/list http://www.west.nesinc.com/WA_testobjectives.asp
Science WESTEST	http://wvde.state.wv.us/oaa/westest_index.html Science public released items: http://wvde.state.wv.us/teach21/westest-science.doc Math public released items: http://wvde.state.wv.us/teach21/westest-math.doc

Addendum. Additional Instruments Identified from Review of Most Recent Cohort

The following Addendum was added in May 2013 to include additional instruments identified through a review of the projects funded in the most recent cohort of the DR-K12 portfolio. The information contained in these two tables is NOT reflected in the report statistics above.

Addendum Table 1: Information on Additional Instruments Identified during Review of Cohort 6 Projects

Acronym	Name	Variables measured/ scales *Unclassified*	Type of Tool GRADE LEVEL	Subject Domain	Construct Assessed	Reliability type/ level VALIDITY EVIDENCE
—	Attitudes and Beliefs about the Nature of and the Teaching of Mathematics and Science	Measures teacher candidates' attitudes and beliefs about mathematics and science and the teaching of those subjects. There are 5 scales: beliefs about mathematics and science; attitudes toward mathematics and science; beliefs about teaching mathematics and science; attitudes toward using technology to teach mathematics and science; attitudes toward teaching mathematics and science.	Survey POST—preservice teachers	Science Math	Beliefs and attitudes	Internal alpha /acceptable CONTENT CRITERION-RELATED
CAOS	Comprehensive Assessment of Outcomes in a first Statistics Course	The CAOS test was designed to provide an instrument that would assess undergraduate students' statistical reasoning after any first course in statistics. Rather than focus on computation and procedures, the CAOS test focuses on statistical literacy and conceptual understanding, with a focus on reasoning about variability. There are 11 scales, consisting of 7-15 multiple-choice items, to be administered online. The test can be used to assess students' reasoning about a particular topic, and can also be used as a review. Instructors can retrieve test reports that can be helpful in identifying successfully learned concepts and where students are having difficulty reasoning about a particular topic. The topics for the scales are: data collection; data representation; measures of center; measures of spread; normal distribution; probability; bivariate quantitative data; bivariate categorical data; sampling distributions; confidence intervals; significance tests.	Assessment POST—undergraduate students	Statistics	Content knowledge	Internal alpha /acceptable CONTENT

Acronym	Name	Variables measured/ scales *Unclassified*	Type of Tool GRADE LEVEL	Subject Domain	Construct Assessed	Reliability type/ level VALIDITY EVIDENCE
DET	Teaching Design, Engineering and Technology Survey	Factors that are assessed include: importance of design, engineering, and technology (DET); familiarity with DET; stereotypical characteristics of engineers; and barriers in integrating DET	Survey K-12 teachers	Engineering Technology	Beliefs	Internal alpha /good CONSTRUCT
DiISC	Discourse in Inquiry Science Classrooms	The instrument was developed to measure teachers' use of strategies in their classrooms to foster a science classroom discourse community as a way of furthering achievement in science. The DiISC is an instrument for observing teachers, not students. It describes what teachers do and focuses on five sets of instructional strategies that form the following scales: inquiry, oral discourse, writing, academic language development and learning principles.	observation MIDDLE, HIGH teachers	Science	Instruction	Intraclass correlations/ good CONTENT CONCURRENT CONSTRUCT
ECI	Energy Concept Inventory	Probes typical student difficulties concerning energy and various dimensions of the energy concept. Includes items that deal with energy-related phenomena that are common in school science. For ease of use it has a multiple-choice format that forces discrimination of scientific aspects of energy storage and transfer from their common sense alternatives.	Assessment MIDDLE to POST students	Science	Content knowledge	-- --
IMAP	Integrating Mathematics and Pedagogy Beliefs Survey	Prospective teachers watch videos and write open-ended responses based on their observations that can then be scores using a quantitative rubric. Beliefs measured include those about mathematics; learning or knowing mathematics; and students' learning and doing mathematics.	Scoring rubric POST—preservice teachers	Math	Beliefs	-- CONTENT

Acronym	Name	Variables measured/ scales *Unclassified*	Type of Tool GRADE LEVEL	Subject Domain	Construct Assessed	Reliability type/ level VALIDITY EVIDENCE
P-TABS	Preschool Teacher Attitudes and Beliefs Towards Science	Can be used to gain a clearer picture of the ideas that teachers have about science and to assess the effects of professional development on these ideas. Three factors are assessed: teacher comfort, child benefit, and challenges	Survey POST—preschool teachers	Science	Beliefs	Internal alpha /-- CONCURRENT
SoCQ	Stages of Concern Questionnaire (SoCQ)	The Concerns-Based Adoption Model is a conceptual framework that describes, explains, and predicts probable educator concerns and behaviors as new initiatives are implemented. The following are the three principal diagnostic dimensions of the model: Stages of Concern: Seven different stages of feelings and perceptions that educators experience when they are implementing a new program or practice. Levels of Use: Eight behavioral profiles that describe actions and behaviors of educators and groups as they use an innovation. Innovation Configurations: Different ways an innovation may be implemented, shown along a continuum from ideal implementation or practice to least desirable practice.	Survey POST—in-service teachers where an innovation is being implemented	General	Admin. and policy context	-- --
—	Teacher Acceptability Index	Measures the acceptability of an intervention for children. Dimensions of intervention acceptability include: risk to the target child, amount of teacher time required, effects of the intervention on other children, and amount of teacher skill required. Intervention Rating Profile (IRP-15) is a more recent version of this instrument.	Survey K-12 teachers	General	Beliefs Admin. and policy context	Internal alpha/ good CONSTRUCT

Acronym	Name	Variables measured/ scales *Unclassified*	Type of Tool GRADE LEVEL	Subject Domain	Construct Assessed	Reliability type/ level VALIDITY EVIDENCE
—	TExES Bilingual Education Supplemental 4-8 Representative Exam	This instrument is a certification examination developed by the Education Testing Service (ETS) and the Texas Examinations of Educator Standards (TExES) of the State Board for Educator Certification. It serves as a proxy for the knowledge of students as English Language Learners. This instrument is designed to determine whether or not an individual has the skills and knowledge necessary to be a teacher in a bilingual education program at the middle school level. The exam covers the following areas: a) methods and procedures to create an effective bilingual learning environment, b) promoting language development in first and second language, and c) methods to effectively teach material in two languages for a variety of subjects, including mathematics.	Assessment POST—certification exam	Literacy	Instruction	-- --

Addendum Table 2: Alphabetical Listing (by acronym) and Access Details of Additional Instruments Identified in Review of Cohort 6 Projects

Instrument Name	LINK	Access details
Attitudes and Beliefs about the Nature of and the Teaching of Mathematics and Science	Instrument and additional psychometric information available in: McGinnis, J. R., Kramer, S., Shama, G., Graeber, A. O., Parker, C. A., & Watanabe, T. (2002). Undergraduates' attitudes and beliefs about subject matter and pedagogy measured periodically in a reform-based mathematics and science teacher preparation program. *Journal of Research in Science Teaching, 39*(8), 713–737.	
Comprehensive Assessment of Outcomes in a first Statistics Course (CAOS)	An online forced-choice testing format allows for quick summaries of the results to be sent to instructors via e-mail, as well as a spreadsheet of student records. A test may be given as an assessment of student learning, for research purposes or for feedback to the instructor. It may also be used as an out of class review for students. Website with additional information: https://apps3.cehd.umn.edu/artist/caos.html For additional information: http://www.stat.auckland.ac.nz/~iase/serj/SERJ6(2)_delMas.pdf	To administer the CAOS test register at: https://apps3.cehd.umn.edu/artist/tests/index.html
Teaching Design, Engineering and Technology Survey (DET)	Additional information available in: Yasar, S., Baker, D., Robinson-Kurpius, S., Krause, S., & Roberts, C. (2006). Development of a survey to assess K-12 teachers' perceptions of engineers and familiarity with teaching design, engineering, and technology. *Journal of Engineering Education, 95*(3), 205-216. Hong, T., Purzer, Ş., & Cardella, M. (2011). A re-evaluation of the design, engineering and technology (DET) instrument. *Journal of Engineering Education, 100*(4), 800-818.	Need to obtain instrument from author— Senay Purzer at Purdue University
Discourse in Inquiry Science Classrooms (DiISC)	Reference manual: Technical Report No. 001, The Communication in Science Inquiry Project (CISIP), Arizona State University: http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1120&context=teachlearnfacpub	

Instrument Name	LINK	Access details
Energy Concept Inventory (ECI)	http://energyeducation.eku.edu/sites/energyeducation.eku.edu/files/EnergyConceptInventory.pdf For additional information: Swackhamer, G., & Dukerich. (2003). An energy concept inventory. Proceedings from 127th American Association of Physics Teachers National Meeting. Madison, WI.	
Integrating Mathematics and Pedagogy Beliefs Survey (IMAP)	The user manual for the online survey can be downloaded from: http://www.sci.sdsu.edu/CRMSE/IMAP/pubs.html Article with additional information: http://www.sci.sdsu.edu/CRMSE/IMAP/pubs/Assessing.pdf	
Preschool Teacher Attitudes and Beliefs Towards Science (P-TABS)	Maier, Greenfield, & Bulotsky-Shearer. (Fall 2011). Preschool teacher attitudes and beliefs towards science: Development and validation of a questionnaire. Paper presented at the annual conference of the Society for Research on Educational Effectiveness. http://www.eric.ed.gov/PDFS/ED528501.pdf	
Stages of Concern Questionnaire (SoCQ)	Additional information about the instrument is on this website: http://www.sedl.org/pubs/catalog/items/cbam21.html	Unable to obtain instrument COST

Instrument Name	LINK	Access details
Teacher Acceptability Index	Witt, J. C. and Martens, B. K. (1983), Assessing the acceptability of behavioral interventions used in classrooms. *Psychology in the Schools, 20*: 510–517. Intervention Rating Profile (IRP-15) is a more recent version of this instrument. Martens, B. K., Witt, J. C., Elliottt, S. N., & Darveaux, D. X. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology: Research and Practice, 16,* 191–198. For a review of these instruments see: Carter, S. (2007). Review of recent treatment acceptability research. *Education and Training in Developmental Disabilities, 42*(3), 301-316.	Unable to obtain instrument
TExES Bilingual Education Supplemental 4-8 Representative Exam	A cost is associated with this test at Educational Testing Service (ETS)	Unable to obtain instrument COST