November 2012

Compendium of Research Instruments for STEM Education

PART 2: Measuring Students' Content Knowledge, Reasoning Skills, and Psychological Attributes

Community for Advancing Discovery Research in Education

Written for CADRE by:

Abt Associates
Daphne Minner
Elisabeth Ericson
Sally Wu
Alina Martinez

November 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 that 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 II 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 I looks at teacher outcome assessments, and Part II looks 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 various 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

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

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 the instruments proposed by projects in five cohorts of DR-K12. These projects were initially funded from 2008 to 2012. In this compendium we focus on instruments designed to assess students' STEM content knowledge, reasoning skills, and psychological attributes such as attitudes, beliefs, emotional regulation, motivation, and career identity/aspirations. This collection of instruments likely represents commonly used tools for gathering information about outcomes of 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 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 November 2012.

Methods

The driving research question for this instrument review was: *What are the instruments, constructs, and methods being used to study student outcomes within the DR-K12 portfolio*? The research team coded information regarding all data sources to measure student outcomes that was contained in the proposal or supporting documents of the DR K-12 projects. However, this compendium focuses on extant, named instruments as opposed to instruments being developed as part of a current grant. Therefore, the information in this compendium reflects assessment tools currently accessible to researchers. Three commonly assessed student outcomes were the target constructs for this review—content knowledge, reasoning skills, and psychosocial attributes.

The review process was conducted in two phases. The first phase included reviewing all available proposals for projects 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 student instruments and the constructs being measured. Once this initial dataset was constructed, a second phase of data collection was conducted to obtain 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. 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, Abt Associates does 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 and the work is underway. 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. 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

One hundred ninety projects (64% of the overall DR-K12 portfolio) proposed to measure students' cognitive or psychosocial attributes as an outcome of the funded work. Across these 190 projects, 118 extant instruments were identified. The most common instruments used among the DR-K12 projects are listed in Table 1. The remaining 97 instruments were proposed to be used by one or two projects only.

Table 1. Number of projects using named instrument

Instrument Name	Number of Projects using
National Assessment of Educational Progress (NAEP)	21
Trends in International Mathematics and Science Study (TIMSS)	14
California Standards Test (CST)	6
Patterns of Adaptive Learning Scales (PALS)	6
Motivated Strategies for Learning Questionnaire (MSLQ)	5
New York Regents High School Examination	5
Colorado Student Assessment Program (CSAP)	4
Force Concept Inventory (FCI)	4
Massachusetts Comprehensive Assessment System (MCAS)	4
Measures of Academic Progress for Science	4
New York State Testing Program (NYSTP)	4
Pennsylvania System of School Assessment (PSSA)	4
Test of Science-Related Attitudes (TOSRA)	4
Attitude Toward Science in School Assessment (ATSSA)	3
Iowa Test of Basic Skills (ITBS)	3
New Mexico Standards Based Assessment (NMSBA)	3
Program for International Student Assessment (PISA)	3
Research-Based Early Maths Assessment (REMA)	3
Scholastic Aptitude Test (SAT)	3
Self-Efficacy in Technology and Science (SETS)	3
Texas Assessment of Knowledge and Skills (TAKS)	3

Figure 1 indicates that forty-six percent of projects (91) did not identify any specific, named instruments intended to be used in the funded DR-K12 research. However, they may have identified a class of instruments such as "state tests" for the states in which they intended to work, which would not be captured within the data in Figure 1. Twenty-four percent (46) of projects proposed to use one extant instrument to measure student outcomes. As can be seen in Figure 2, most commonly projects proposed to use four different data sources, including things like school grades, to round out their multi-method data collection approach. Fourteen percent (27) of projects proposed to measure student outcomes using one data source. Thirty-three percent of these single-data-source projects developed criterion-referenced tests for the

proposed study, 30 percent used extant criterion-referenced tests, and 37percent used a mix of interviews and student work.

Figure 1. Percent of projects proposing to use extant, named instruments

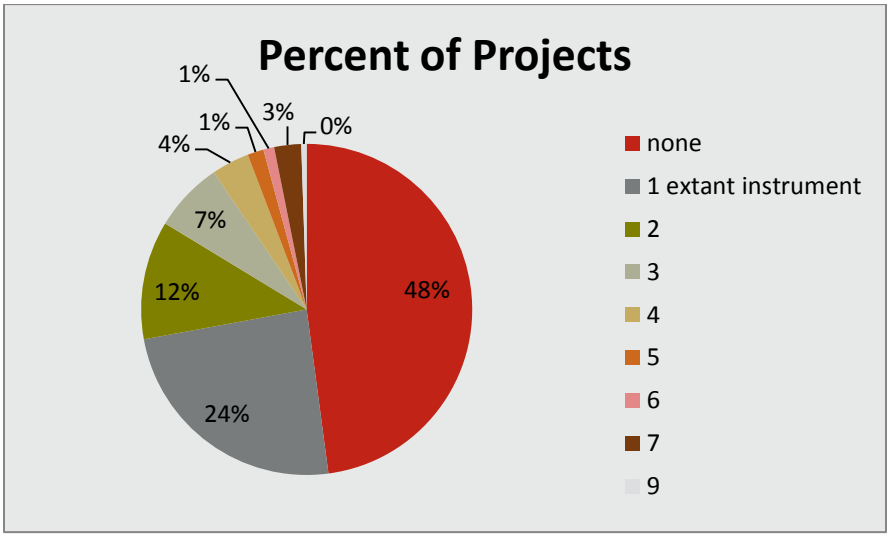
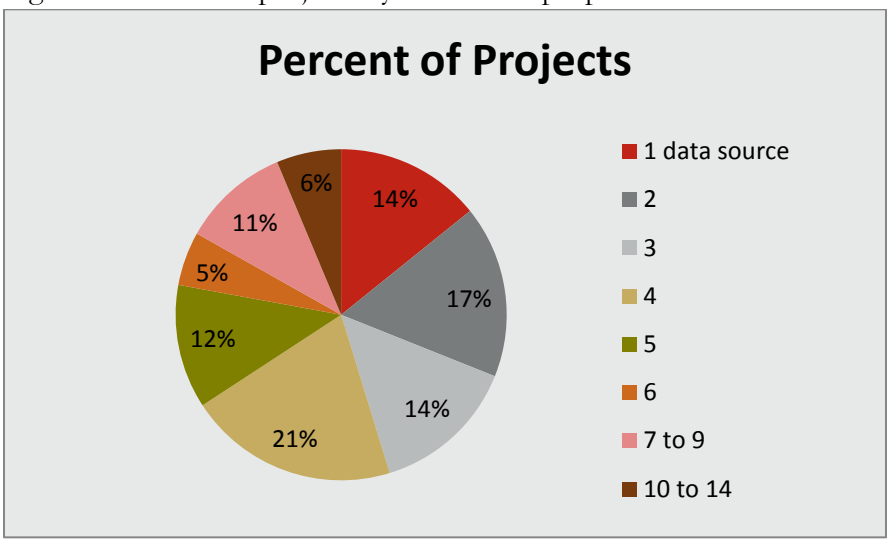


Figure 2. Percent of projects by number of proposed data sources



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 correlations. Validity evidence was related to content, construct, criterion-related/ predictive, concurrent/ concurrent, and discriminant/ divergent data. With this information we were able to assess the strengths and weaknesses in the measurement landscape for key educational constructs. The 118 student instruments fell within two outcome domains—cognitive and psychosocial attributes. Within each domain the instruments are further differentiated by the main constructs assessed, as noted in Table 2. The instruments in each of these categories are profiled in Appendices A-G.

Table 2. Number and percentage of unique instruments identified by construct

Outcome Domain	Student Constructs	N (%)*
Cognitive 82 (69%)	Content knowledge and reasoning skills	43 (36%)
	a) Science focus	23 (53%)
	b) Math focus	11 (26%)
	c) STEM focus large scale	9 (21%)
	Literacy and language skills	8 (7%)
	State academic achievement tests	31 (26%)
Psychosocial 36 (31%)	Attitudes	13 (11%)
	Emotional attributes	8 (7%)
	Motivational attributes	12 (10%)
	Career identity	3 (3%)

*denominator = 118 instruments

State academic achievement tests and other large-scale tests are developed by psychometricians for large-scale administration, and have 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, this information is not provided, but access details for technical reports are provided for researchers to obtain this information for their own purposes in Appendix I. For these types of instruments we present information regarding the subject domains assessed, the specific math and science areas that are measured, and the target audience of the instrument. For the assessments that are embedded within a specific curriculum, we do not provide reliability and validity information for these assessments because they often involve multiple assessments, therefore it would be incumbent on a researcher to identify that information for a particular embedded assessment they are intending to use. Again access information is provided in the appendices, but reliability and validity information is not. In the next section of the report, the instruments associates with each of the student constructs listed in Table 2 are presented and trends noted.

Student Assessments for Outcomes within the Cognitive Domain

The first domain of student outcomes to be assessed in the DR-K12 projects was cognitive, capturing variables such as students' (a) content knowledge and reasoning skills within STEM—science focus, math focus, both; (b) literacy and language skills; and (c) academic achievement. There were 82 assessments identified within this domain comprising 69 percent of the 118 assessments identified across the portfolio. Among the 43 student assessments that measure content knowledge and reasoning skills within STEM, fifty-three percent (23) deal with science content, 26 percent (11) mathematics, and 21 percent (9) that measure both science and math, and sometimes other skills such as literacy.

Content Knowledge and Reasoning Skills within STEM

Science, Technology and Engineering Topics—This is the largest category of student content outcomes that are measured across the DR-K12 program, with 23 instruments named to assess these outcomes. Within the science specific instruments assessing content knowledge and reasoning skills seven were from existing curricula (see Table 3), while the remaining instruments were stand-alone assessments. There were five categories of topics assessed as the main focus of an instrument: life science (4 instruments), physical science (4), multiple domains (6), scientific inquiry (3), and technology/engineering (6). There were four instruments that specifically assessed misconceptions related to particular science topics. Of the 16 instruments for which we tried to obtain reliability and validity information, 56 percent (9) had evidence of reliability and 50 percent (8) had validity evidence (See Appendix A for additional details).

Mathematics Topics—Within the mathematics area, there were three named curriculum-based assessments (see Table 4, p.9). Across the eleven instruments there were four categories of topics assessed as the main focus of the instrument: algebra (2 instruments), geometry (3), math skills (5), and multiple areas (1). Of the 8 instruments for which we tried obtained reliability and validity information, 38 percent (3) had evidence of reliability and 38 percent (3) had validity evidence (See Appendix A for additional details). There was a mixture of scoring and data collection formats including criterion-referenced instruments (7), norm-referenced (1), interview (1), observation protocol (1), and coding rubric (1).

Large-scale STEM Assessments—There were nine large-scale assessments that capture various math and science topics, or both topics. These assessments typically have item banks, multiple years of data collection (often available to the public), and are used for national or international comparisons across locales, students, and years. Table 6 provides the general characteristics of these assessments. Six of them are norm-referenced and three are criterion-referenced; all measure both content knowledge and reasoning skills.

Literacy and Language Skills

There were eight instruments that measured language and literacy skills (5 and 3 respectively, see Appendix B for details). Two of these instruments were for preK learners and the others were for a wide range of ages PK-adult. The literacy tests assess comprehension in terms of reading and writing skills, whereas the language skills tests focus on oral skills—primarily vocabulary.

Academic Achievement

The final group of assessments that measure student content knowledge and reasoning skills is state achievement tests (see Appendix C). The DR-K12 projects identified 31 tests across 25 states as instruments to use for measuring student outcomes. The presence of state tests among measures of student outcomes is not a surprise since many states provide item banks and comparative data for educators and researchers. Additionally, there is a good deal of pressure to use these state tests to tie research findings into the assessment systems that policy-makers require.

We performed a content analysis based on the descriptions of assessment blueprints provided on the state websites to identify what is measured regarding mathematics and science. There is much more consistency within the mathematics topics assessed across tests, than there is within science. Most of the math topics assessed on the state tests identified by the DR-K12 projects correspond to the Common Core State Standards for Mathematics. Among the ten conceptual categories we identified from the assessment blueprints, five of them were addressed by 22 of the 26[2] state tests (85%) (see Table 6, p. 11). Whereas in the science topical areas, there were only two broad categories—life, earth/space, physical sciences; and inquiry/ scientific thinking—that were addressed in the majority of the 25 tests (see Table 7). This may change as the Next Generation Science Standards (NGSS) are adopted by states. Interestingly, there were only two states (MA, OR) that explicitly addressed engineering or design issues in their current state test blueprints. This will likely change as engineering plays a prominent role in NGSS.

[2] There were 31 state tests identified in the sample, but five of these did not have available assessment blueprints for review and are not included in this analysis.

Table 3. Science topic assessments

Acronym	Instrument Name	Instrument type	Grade Level 1=elem-middle 2=middle-high 3=K-12 4=high-post[1]	Content Knowledge and Reasoning Topic Areas	Reliability evidence	Validity evidence
SLA	Science Learning Assessment	criterion	1	life science, inquiry	x	x
	Cornell Scientific Inquiry Series	st. work*	4	environmental science inquiry	**	**
	EcoMUVE	criterion	2	environmental science	**	**
	Students' Use of Evidence in Written Scientific Explanations	st. work	4	natural selection, argumentation	x	
MBT	Mechanics Baseline Test	criterion	4	physics		x
FCI	Force Concept Inventory	criterion	4	physics belief system and **misconceptions**	x	
CCI	Chemistry Concept Inventory	criterion	4	chemistry and **misconceptions**	x	x
ACS Exams	American Chemical Society Exams	criterion	4	chemistry		
MAP	Measures of Academic Progress for Science	criterion	4	life, physical, Earth/space sciences	x	
MOSART	Misconceptions-Oriented Standards-Based Assessment Resources for Teachers	criterion	3	life, physical, Earth/space sciences and **misconceptions**		x
SASKS	Science Attitudes, Skills, & Knowledge Survey	survey	4	life, physical, Earth/space sciences		
FOSS	Full Option Science System	criterion	1	life, physical, Earth/space sciences	**	**
TELS	Technology Enhanced Learning in Science	st. work	2	life, physical, Earth/space sciences, simulation-based reasoning	**	**
	AAAS Item Bank	criterion	2	life, physical, Earth/space sciences, NOS, and **misconceptions**		x
SUSSI	Student Understanding of Science and Scientific Inquiry	survey	4	scientific inquiry, NOS (Nature of Science)	x	x
	Science Notebooks Assessment Tool Rubric	st. work	1	scientific inquiry	x	x
SEPUP	Science Education for Public Understanding Program	criterion	2	scientific inquiry	**	**
	TELS Knowledge Integration Scoring Rubric	st. work	2	simulation-based knowledge integration	**	**
EiE	Engineering is Elementary	criterion	1	engineering and technology attitudes	**	**
	4-H Robotics Concept Test	criterion	1	engineering, technology	x	x
	Scientific Inquiry and Engineering Design Scoring Guides	st. work	3	scientific inquiry, engineering		
	Engineering Design Process Knowledge	st. work	2	engineering design knowledge	x	
TAGLIT	Taking a Good Look at Instructional Technology	criterion	3	technology literacy		

[1]post refers to post-secondary and adult; *student work; **=curriculum assessment NA

Table 4. Mathematics topic assessments

Acronym	Instrument Name	Instrument type	Grade Level 1=elem-middle 2=middle-high 3=K-12 4=high-post	Content Knowledge and Reasoning Topic Areas	Reliability evidence	Validity evidence
	Acuity Algebra Assessment	criterion	2	algebra		x
EASA	Early Algebra Student Assessment	criterion	1	algebra	**	**
BBA	Building Blocks Assessment of Early Mathematics	interview	1	number, geometry	x	x
EG	Entering Geometry Test	criterion	4	geometry	x	
REMA	Research-Based Early Maths Assessment	criterion	1	geometry, math skills		
	Upper-Elementary Mathematics Assessment Modules	criterion	1	math skills		
TEMA-3	Test of Early Mathematics Ability-Third Edition	norm	1	math skills	x	x
IMP	Interactive Mathematics Program	criterion	4	math skills	**	**
	Mathematics Classroom Observation Protocol	observation	1	math skills		
MAJAC	Mathematical Argument as Joint Activity in the Classroom	coding rubric	unsure	math skills, argumentation		
CPMP	Core-Plus Mathematics Project	criterion	4	algebra, geometry, statistics, discrete math	**	**

**=curriculum assessment NA

Table 5. Large-scale mathematics **and** science topic assessments

Acronym	Instrument Name	Instrument type	Grade Level 1=elem-middle 2=middle-high 3=K-12 4=high-post	Content Knowledge and Reasoning Topic Areas
KeyMath-3 DA	KeyMath-3 Diagnostic Assessment	norm	3	math concepts, operations, and applications
SAT	Scholastic Aptitude Test	norm	4	reading, writing, math skills
ACT Science	ACT Test of Science Reasoning	norm	4	life, physical, Earth/space sciences, scientific reasoning
PISA	Program for International Student Assessment	criterion	4	science, math, literacy
TIMSS	Trends in International Mathematics and Science Study	criterion	1	science, math
NAEP	National Assessment of Educational Progress	criterion	3	math skills, science inquiry
Terra Nova	Terra Nova Algebra Assessment	norm	3	math skills, science inquiry
ITBS	Iowa Test of Basic Skills	norm	1	math, science, scientific inquiry, mathematical reasoning
Stanford 10	Stanford Achievement Test	norm	3	science, math content knowledge and reasoning skills

Table 6. Mathematics conceptual categories assessed in current state assessments identified in this review with assessment blueprints

State	Test Acronym	GEOMETRY	DATA ANALYSIS, STATISTICS, PROBABILITY	NUMBER & OPERATIONS	ALGEBRA	MEASUREMENT	PATTERNS/ RELATIONS/ FUNCTIONS	MATH REASONING	COMPUTATION	PROBLEM SOLVING	ESTIMATION
AK	SBA	X	X	X		X	X		X		X
AL	ARMT	X	X	X	X	X					
CA	CST	X	X	X	X	X		X			
CO	CSAP	X	X	X		X	X	X			
IL	ISAT	X	X	X	X	X					X
IN	ISTEP	X	X	X	X	X	X		X	X	
KY	K-PREP	X	X	X	X	X	X	X			
LA	iLEAP	X	X	X	X	X	X				
MA	MCAS	X	X	X	X	X	X				
MD	HSA		X		X		X			X	
MD	MSA	X	X	X	X	X	X		X		
MO	MAP	X	X	X	X	X	X				
MS	MCT2	X	X	X	X	X					
NC	EOG	X	X	X		X					
NJ	NJ ASK	X	X	X	X	X		X			
NM	NMSBA	X	X	X	X	X					
NY	Regents	X			X						
NY	NYSTP	X	X	X	X	X					
OH	OGT	X	X	X	X	X	X				
OH	OAA	X	X	X	X	X	X				
OR	OAKS	X	X	X	X	X					
PA	PSSA	X	X	X	X	X					
TX	TAKS	X	X	X	X			X		X	
UT	U-PASS	X			X						
VA	SOL		X	X	X	X	X		X		X
WA	MSP	X	X	X		X		X		X	
PERCENTAGE		92%	92%	88%	85%	85%	46%	23%	15%	15%	12%

Table 7. Science topics assessed in current state assessments identified in this review with assessment blueprints

State	Test Acronym	LIFE, EARTH, PHYSICAL	INQUIRY/ SCIENTIFIC THINKING	NATURE OF SCIENCE	UNIFYING IDEAS	ENVIRONMENT	SCIENCE, TECHNOLOGY, SOCIETY	TECHNOLOGY	INVESTIGATION/ EXPERIMENTATION	SKILLS/ PROCESSES	ENGINEERING / DESIGN
AK	SBA	X	X	X				X			
CA	CST	X							X		
CO	CSAP	X		X					X		
IL	ISAT		X				X				
IN	ISTEP	X	X	X			X	X			
KY	K-PREP	X			X						
LA	iLEAP	X	X			X					
MA	MCAS	X			X			X			X
MD	HSA	LIFE									
MD	MSA	X				X				X	
MO	MAP	X	X			X		X			
MS	MST2	X	X								
NC	EOG		X		X						
NJ	NJ ASK	X				X	X	X		X	
NM	NMSBA	X		X			X				
NY	Regents	X				X					
NY	NYSTP		X		X	X		X			
OH	OGT	X	X	X			X				
OH	OAA	X	X	X			X		X		
OR	OAKS	X	X		X					X	X
PA	PSSA	X		X							
TX	TAKS	X		X	X						
UT	U-PASS	X									
VA	SOL	X	X		X	X	X		X		
WA	MSP	X	X		X			X	X		
PERCENTAGE		88%	52%	32%	32%	28%	28%	28%	20%	12%	8%

Student Assessments for Outcomes within the Psychosocial Domain

The second domain of student outcomes proposed in the DR-K12 projects was psychosocial, capturing variables such as attitudes, emotional and motivational aspects, and career aspirations and identity. There were 36 assessments identified within this domain comprising 31 percent of the 118 total assessments identified.

Attitudes

There were 13 attitude assessments named. Only one was related to mathematics, two looked at engineering and /or computer science, and the remaining 5 involved science. In addition to attitudes toward STEM, there were five instruments that assess other constructs such as perceptions of science professionals or views of the nature of science as a discipline (see Table 8 and Appendix D for additional details).

Emotional and Motivational Attributes, and Career Identity

There were eight instruments identified as measuring variables such as temperament, social skills, goal orientation, self-concept and creativity (see Table 9 and Appendix E for details). None of these were specific to a STEM area. Two of the instruments were surveys and two were norm-referenced assessments for establishing typical early childhood behaviors. There were twelve instruments that measured some aspect of motivation such as self-efficacy (3 instruments—one each about science and technology, and one non-subject specific), metacognitive awareness (2), motivational profiles for different types of activities (4), and student engagement during instruction (3). All three of the student engagement instruments were observation protocols that capture various aspects of the teacher's instructional behaviors (see Table 10 and Appendix F for details). There were three instruments related to career interests and one that specifically focuses on engineering identity (see Table 11 and Appendix G for details).

Table 8. Student attitudes assessments

Acronym	Instrument Name	Instrument Type	Grade Level 1=elem-middle 2=middle-high 3=K-12 4=high-post	Attitudes	Reliability evidence	Validity evidence
ATSSA	Attitude Toward Science in School Assessment	survey	2	attitudes towards science as an academic subject	x	
mATSI	Modified Attitudes towards Science Inventory	survey	1	perception of science teacher, anxiety toward science, social value of science, science self-concept	x	x
TOSRA	Test of Science-Related Attitudes	survey	4	science engagement, attitudes towards science	x	x
SAI II	The Scientific Attitude Inventory: A revision	survey	2	science interest, engagement, competence, attitudes	x	
SPOCC	Student Perceptions of Classroom Climate	survey	1	attitudes towards and perceptions about science instruction		
	AWE Pre-College Recruiting Surveys	survey	2	attitudes of, and interest in engineering, science, computer science		
CSABA	Computer Science Attitudes and Beliefs Assessment	survey	4	confidence, interest, usefulness of computing skills	x	x
ATMI	Attitudes Toward Mathematics Inventory	survey	2	math self-confidence, beliefs about its usefulness, enjoyment and motivation to engage in math	x	x
DAET	Draw an Engineer Test	performance	1	elicits knowledge about engineering and perceptions of engineers	x	x
DAST	Draw-A-Scientist Test	student work	1	elicits conceptions about the image of a scientist		
VASS	Views about Science Survey	survey	4	attitudes towards science education, views of NOS	x	
VNOS	Views of Nature of Science Questionnaire	survey	3	views about various aspects of science such as empirical, creative, inferential, etc.		x
VOSI	Views of Science Inquiry	survey	3	ideas of what scientists do to produce valid science knowledge		x

Table 9. Student emotional attributes assessments

Acronym	Instrument Name	Instrument Type	Grade Level 1=elem-middle 2=middle-high 3=K-12 4=high-post	Emotional Attributes	Reliability evidence	Validity evidence
EATQ-R	Early Adolescent Temperament Questionnaire–Revised	survey	2	temperament	x	x
ERC	Emotion Regulation Checklist	survey	1	emotion regulation	x	x
SSIS-RS	Social Skills Improvement System-Rating Scales	norm	3	social skills	x	x
AGQ	Achievement Goals Questionnaire	survey	4	goals	x	x
GOALS-S	Goal Orientation and Learning Strategies Survey	survey	4	goals		x
PALS	Patterns of Adaptive Learning Scales	survey	3	goals	x	x
SDQII	Self-Description Questionnaire II	survey	2	self-concept	x	x
WKCT	Wallach-Kogan Creativity Test	norm	1	creativity	x	x

Table 10. Student motivational attributes assessments

Acronym	Instrument Name	Instrument Type	Grade Level 1=elem-middle 2=middle-high 3=K-12 4=high-post	Motivational Attributes	Reliability evidence	Validity evidence
SETS	Self-Efficacy in Technology and Science	survey	2	technology self-efficacy	x	x
SSSE	Sources of Science Self-Efficacy Scale	survey	2	science self-efficacy	x	
	Children's Self-Efficacy Scale	survey	3	self-efficacy	x	x
MSLQ	Motivated Strategies for Learning Questionnaire	survey	4	metacognition		x
Jr. MAI	Junior Metacognitive Awareness Inventory	survey	1	metacognition	x	x
IMI	Intrinsic Motivation Inventory	survey	3	motivation	x	x
IMMS	Instructional Material Motivational Survey	survey	4	motivation		
	Children's self & task perceptions during elem. school	survey	1	math, literacy, sports, music--motivation	x	x
NASA-TLX	NASA Task Load Index	performance	4	task demands--motivation	x	
IQA	Instructional Quality Assessment	observation	1 (teacher)*	literacy, math--student engagement	x	
SCE	Student Collective Engagement	observation	3 (teacher)	student engagement	x	
FFT	Framework for Teaching Evaluation Instrument	observation	3 (teacher)	student engagement		x

*(teacher) indicates that the observation instrument collects specific data about both students and teachers.

Table 11. Student career identity assessments

Acronym	Instrument Name	Instrument Type	Grade Level 1=elem-middle 2=middle-high 3=K-12 4=high-post	Career Identity Aspects	Reliability evidence	Validity evidence
EIDS	Engineering Identity Development Scale	survey	1	engineering identity	x	x
KCS	Kuder Career Search	survey	2	career interests	x	x
Strong	Strong Interest Inventory	survey	unsure	career interests	x	x

Reliability and Validity Evidence

As mentioned before, we did not collect reliability and validity evidence for the large-scale, state achievement, or embedded curriculum assessments. That left a total of 68 eligible assessment for which we did attempt to locate this information. Below you will find the number and percentage of studies by outcome domain (Table 12) and type of assessment (Table 13) that demonstrated acceptable or high levels of at least one type of reliability or validity. For additional details on the types of evidence see Appendices A through G. The majority of eligible instruments that assess the psychosocial domains had sufficient evidence of both reliability and validity associated with their use, whereas only the literacy and language instruments met this threshold for the domain-specific knowledge and reasoning assessments. However, this finding should be interpreted with caution because the large-scale and state achievement tests, that more researchers use, have this kind of evidence available. Therefore, we suggest either, greater care should be used in the development and testing of domain specific assessments or, researchers should use existing tools that have demonstrated reliability and validity evidence within a cognitive domain. Of the more commonly used types of instruments, surveys and norm-referenced tests had more reliability and validity evidence than did criterion-referenced tests in this sample of instruments (see Table 13).

Table 12. Number and percentage of instruments by reliability and validity evidence

Outcome Domains	Number of Eligible Instruments	Acceptable to High Reliability Evidence	Evidence of Validity
Cognitive			
Science	16	8 (50%)	8 (50%)
Math	8	3 (40%)	3 (40%)
Literacy/Language	8	4 (50%)	**5 (63%)**
Psychosocial			
Attitudes	13	**7 (54%)**	**7 (54%)**
Emotional	8	**6 (75%)**	**8 (100%)**
Motivation	12	**8 (67%)**	**7 (58%)**
Career	3	**3 (100%)**	**3 (100%)**

Table 13. Number and percentage of instruments by reliability and validity evidence

Instrument Type	Number of Eligible Instruments	Acceptable to High Reliability Evidence	Evidence of Validity
Survey	30	**21 (70%)**	**22 (73%)**
Criterion-referenced test	18	7 (39%)	8 (44%)
Norm-referenced test	5	**4 (80%)**	**5 (100%)**
Student work	5	2 (40%)	1 (20%)
Observation	3	1 (33%)	1 (33%)
Other/performance	3	**3 (100%)**	**2 (67%)**
Interview	2	1 (50%)	**2 (100%)**
Coding rubric	1	0	0

Conclusion

This compendium is intended to serve as an initial step towards the systematic assessment and improvement of tools available to STEM education developers and researchers interested in measuring student outcomes. It documents that there is a wealth of resources available to assess student outcomes and points to available information that can help guide the selection of appropriate and tested assessments. One consideration in selecting instruments is the alignment between what is measured on the designated instrument, the intervention being studied, or a phenomenon under study, a consideration naturally outside the scope of this compendium. However, we hope that 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.

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. It will serve as a focal point to bring together a group of DR K-12 PIs to dialogue about the use of existing student assessments in research and development projects. Discussion topics will include the availability of information about assessments, the features considered when selecting assessments for use, and whether common instruments across studies is feasible and desirable.

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.

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 student achievement and attitudes, 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

International Society for Technology in Education technology standards (NETS) http://www.iste.org/standards

Proceedings of the National STEM Assessment Conference (2006). http://www.openwatermedia.com/downloads/STEM(for-posting).pdf

Psychometric analyses of additional tools:
Owen SV, Toepperwein MA, Blalock CL, Liu Y, Pruski LA, Grimes K, Lichtenstein MJ. (2008). Finding pearls: Psychometric re-evaluation of the Simpson-Troost Attitude questionnaire (STAQ) (http://www.pearweb.org/atis/tools/21). *Science Education, 92*(6):1076-1095.

Blalock CL, Lichtenstein MJ, Owen SV, Pruski LA, Marshall CE, Toepperwein MA. (2008). In pursuit of validity: A comprehensive review of science attitude instruments 1935-2005. *International Journal of Science Education*, *30*(7):961-977.

Marshall CE, Blalock CL, Liu Y, Pruski LA, Toepperwein MA, Owen SV, Lichtenstein MJ.(2007). Psychometric re-evaluation of the Image of Science and Scientists Scale (ISSS) *School Science and Mathematics, 107*(4):149-160.

Owen SV, Toepperwein MA, Pruski LA, Blalock LA, Liu Y, Marshall CE, Lichtenstein MJ. (2007). Psychometric re-evaluation of the Women in Science Survey (WiSS) (http://www.pearweb.org/atis/tools/20). *Journal of Research in Science Teaching, 44*(10):1461-1478.

Special Issue on Assessment: *Journal of Research in Science Teaching* (2012). *49*(6).

Tyler-Wood, T., Knezek, G., & Christensen, R. (2010). Instruments for assessing interest in STEM content and careers. *Journal of Technology and Teacher Education. 18*(2), 341-363. http://www.iittl.unt.edu/IITTL/itest/msosw_web/pubs/STEMInstruments.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.

Morsy, L., Kieffer, M., & Snow, C. (2010). *Measure for Measure: A Critical Consumers' Guide to Reading Comprehension Assessments for Adolescents.* New York: NY: Carnegie Corporation of New York. http://carnegie.org/fileadmin/Media/Publications/PDF/tta_Morsy.pdf

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) is a searchable website of assessment tools for informal science learning and can be accessed from: http://www.pearweb.org/atis. The goal of this resource is to provide practitioners, evaluators, researchers and policy makers with the information to choose appropriate tools for assessing program quality and outcomes for children and youth. PEAR, supported by the Noyce Foundation and located at Mclean Hospital and Harvard Medical School, reviewed existing tools and published their findings in a report titled *Toward a Systematic Evidence-Base for Science in Out-of-*

School Time: The Role of Assessment. This website is based on the findings of that report and will be continuously updated in collaboration with the Youth Development Researchers at 4-H.

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#

Education Development Center, Inc. *Evidence of Understanding: An Introduction to Assessments in K–12 Science Curricula:* http://cse.edc.org/products/assessment/assesshome.asp
In this web guide, educators can learn about different types of assessments and view examples from different science instructional materials. Within each grade level section, assessment examples have been categorized into five groups: (1) observations, interviews, and discussions, (2) written assessments, (3) performance assessments, (4) graphics, and (5) self-assessments. Several examples of each type of assessment are provided for each grade level group.

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. Some selected instruments that were not identified in the DR-K12 portfolio, but may be of interest to investigators are included here.

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

Performance Assessment Links in Science (PALS) from SRI International, Center for Technology in Learning: http://pals.sri.com/ PALS is an on-line, standards-based, continually updated resource bank of science performance assessment tasks indexed via the National Science Education Standards (NSES) and various other standards frameworks. The tasks,

collected from numerous sources, include student directions and response forms, administration procedures, scoring rubrics, examples of student work, and technical quality data calculated from field testing.

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 at: http://www.cfa.harvard.edu/smgphp/mosart/aboutmosart_2.html were developed by 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 (NSES). 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 NSES, as well as to the research literature documenting misconceptions concerning science concepts.

The Northwest Ohio Center for Excellence in STEM Education compiled a collection of instruments, although not recently updated, across many domains of science that provides useful references and easy to locate information as a starting point. http://cosmos.bgsu.edu/communities/research_community/MeasurementInst/pdfs/sample%20of%20classroom%20assessment%20instruments.pdf

The Society for the Advancement of Biology Education Research (SABER) has compiled a list of conceptual assessments in biology for undergraduate students that is available at: http://saber-biologyeducationresearch.wikispaces.com/Concept+Assessments-Biology

AAAS Project 2061 Science Assessment Website item bank is appropriate for assessment of middle and early high school students' understanding of physical, earth, life sciences and the nature of science concepts can be found at: http://assessment.aaas.org/

Undergraduate Research Student Self-Assessment (URSSA) is an online survey instrument for use in evaluating student outcomes of undergraduate research experiences in the sciences. There are online tools that enable customization and administration to gather anonymous responses from students. Survey results are available in both summarized and spreadsheet formats. It can be accessed at: http://www.colorado.edu/eer/research/undergradtools.html

References

Common Core State Standards Initiative. *Common Core State Standards for Mathematics.* http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

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

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

National Science and Technology Council (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

National Science and Technology Council (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 Assess Students' Content Knowledge, Understanding, Reasoning and Process Skills (alphabetical by acronym)

Acronym	Name	Student knowledge & reasoning variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain	Reliability type (level) VALIDITY EVIDENCE
---	4-H Robotics Concept Test	Measures general levels of achievement in science, engineering, and technology for youth ages nine to eleven.	criterion ELEM	science engineering technology	internal consistency (high) CONTENT
---	AAAS Item Bank	Tests (1) student conceptual understanding in the earth, life, physical sciences, and the nature of science and (2) common misconception and alternative ideas along with correct ideas.	criterion MIDDLE EARLY HIGH	science	CONTENT
ACS Exams	American Chemical Society Exams	Measures general chemistry, organic chemistry, analytical chemistry, physical chemistry, inorganic chemistry, biochemistry, polymer chemistry, and high school chemistry	Criterion ADAPTED FOR HIGH SCHOOL	science	
ACT Science	ACT Test of Science Reasoning	Emphasizes scientific reasoning skills (interpretation, analysis, evaluation, reasoning, and problem solving) over recall of scientific content, which includes biology, chemistry, physics, and the Earth/space sciences (for example, geology, astronomy, and meteorology). Data Representation (38%): graphic and tabular material similar to that found in science journals and texts. Measures skills such as graph reading, interpretation of scatterplots, and interpretation of information presented in tables, diagrams, and figures. Research Summaries (45%): provides descriptions of experiments. These questions focus on the design of experiments and the interpretation of experimental results. Conflicting Viewpoints (17%): presents expressions of several hypotheses or views that, being based on differing premises or on incomplete data, are inconsistent with one another. These questions focus on the understanding, analysis, and comparison of alternative viewpoints or hypotheses.	large/ norm HIGH SCHOOL	science	interrater CONVERGENT

Acronym	Name	Student knowledge & reasoning variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain	Reliability type (level) VALIDITY EVIDENCE
---	Acuity Algebra Assessment	Measures Algebra I proficiency by the following strands: number sense, computation, and estimation; quadratic and exponential functions; variables and expressions; radical expressions and equations; inequalities and equations; rational expressions and equations; graphic inequalities and linear equations; geometry; functions; statistics and probability	criterion 6-12	math	CONVERGENT CRITERION
BBA	Building Blocks Assessment of Early Mathematics, PreK-K	Assesses developmental progression for math topics: number (verbal counting, object counting, number recognition, number comparison, number sequencing, numerals, number composition, adding and subtracting, place value); and geometry (shape identification, shape composition, congruence, construction of shape, turns, measurement, patterning)	interview PK-K	math	internal consistency (high) interrater/percent agreement (high) CONTENT CONVERGENT
CCI	Chemistry Concept Inventory	Assesses chemistry related misconceptions related to bonding, intermolecular forces, electrochemistry, equilibrium, thermochemistry and acids and bases.	criterion POST*	science	internal consistency (accept) CONVERGENT DISCRIMINANT CONTENT
---	Cornell Scientific Inquiry Series CURRICULUM TEST	A series of manuals aimed at enabling high school students to carry out authentic environmental science research. Each of the manuals includes a student edition and a teacher's guide.	student work HIGH	science	NA—CURRICULUM TEST
CPMP	Core-Plus Mathematics Project assessments CURRICULUM TEST	Centered around four strands: algebra/functions, geometry/trigonometry, statistics/probability, and discrete mathematics.	criterion HIGH	math	NA—CURRICULUM TEST
EASA	Early Algebra Student Assessment	early algebra--COULDN'T FIND INSTRUMENT	criterion 1-6	math	COULDN'T FIND SPECIFIC INFORMATION

* post-secondary or adult

Acronym	Name	Student knowledge & reasoning variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain	Reliability type (level) VALIDITY EVIDENCE
---	EcoMUVE assessments CURRICULUM TEST	Computer-based modules that help students develop a deeper understanding of ecosystems and causal patterns using Multi-User Virtual Environments (MUVEs).	criterion MIDDLE	science	NA—CURRICULUM TEST
EG	Entering Geometry Test	Geometry facts and concepts: area, perimeter, and angle relations.	criterion HIGH	math	internal consistency (accept)
EiE	Engineering is Elementary Research Instruments CURRICULUM TEST	Measures science and engineering topics addressed in specific EiE curricular units as well as general assessments that examine knowledge of technology and general engineering concepts and attitudes towards science and engineering.	criterion (and survey) 1-5	engineering technology science	NA—CURRICULUM TEST
---	Engineering Design Process Knowledge	Assesses the design knowledge into more or less sophisticated knowledge use: Remembering (recite information from memory without necessarily understanding it), Understanding (explain material within its own domain), Applying (use a concept to solve a particular Problem), Analyzing (parse something into its parts), Evaluating (judge different concepts and determine their value), Creating (combining concepts to create something new)	student work 6	engineering	internal consistency (Not Reported) alternate form (Not Reported)
FCI	Force Concept Inventory	Assess student understanding of the most basic concepts in Newtonian physics: kinematics, Newton's First, Second, and Third Laws, the superposition principle, and types of forces (such as gravitation, friction).	criterion HIGH POST ADULT	science	test-retest (high) internal consistency (high) alternate form (accept)
FOSS	Full Option Science System assessments CURRICULUM TEST	Assessments for 41 modules and courses available for grades K–8, organized under the strands of Life Science, Physical Science, Earth Science, and Scientific Reasoning and Technology.	criterion K-8	science	NA—CURRICULUM TEST

Acronym	Name	Student knowledge & reasoning variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain	Reliability type (level) VALIDITY EVIDENCE
IMP	Interactive Mathematics Program assessments CURRICULUM TEST	The curriculum integrates traditional material with additional topics recommended by the NCTM Standards, such as statistics, probability, curve fitting, and matrix algebra. Units are generally structured around a complex central problem.	criterion HIGH	math	NA—CURRICULUM TEST
ITBS	Iowa Test of Basic Skills	Multiple domains assessed. Math and science constructs below. Math: Levels 5 and 6 assess students' knowledge of beginning math concepts, focusing on numeration, geometry, measurement, and problem solving using addition and subtraction. Science: 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)	large/ norm ELEM MIDDLE	general	internal consistency (high) CRITERION
KeyMath -3 DA	KeyMath-3 Diagnostic Assessment	Measure of essential mathematical concepts and skills. Basic concepts (numeration, algebra, geometry, measurement, data analysis and probability); Operations (mental computation and estimation, written computation--addition and subtraction; multiplication and division); Applications (foundations of problem solving, applied problem solving)	large/ norm K-12	math	test retest (high) internal consistency (high) alternate form (high) CONTENT CONVERGENT CONSTRUCT
MAJAC	Mathematical Argument as Joint Activity in the Classroom	Argumentative discourse is examined turn by turn with respect to teaching moves in support of argumentation (e.g., open-ended and closed-ended questions, moves to orchestrate participation), amount of student argumentative talk (e.g., number of student statements across episodes), and cognitive complexity of student contributions (e.g., providing recalled facts versus explaining reasoning).	coding rubric UNSURE	math	

Acronym	Name	Student knowledge & reasoning variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain	Reliability type (level) VALIDITY EVIDENCE
MAP for Science	Measures of Academic Progress for Science	Measures two critical areas: (1) General Science covers specific science concepts within the three major domains of science: Life sciences, earth and space sciences, and physical sciences. (2) Concepts and Processes measures a student's performance in both the processes used in science and the major themes underlying the science disciplines.	Criterion PRIOR TO SPECIALIZED SCIENCE CURRICULUM IN UPPER HIGH SCHOOL	science	test retest (Not Reported) internal consistency (Not Reported)
---	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	obs ELEM MIDDLE	math	
MBT	Mechanics Baseline Test	Probe concepts and principles that cannot be grasped without formal knowledge about mechanics: Kinematics (linear and curvilinear motion), basic principles (Newton's First, Second, and Third Laws, superposition principle, energy conservation, impulse-momentum, and work) and special forces (gravity and friction). This is an advanced companion to the Force Concept Inventory (FCI).	criterion	science	High school and undergraduate
MOSART	Misconceptions-Oriented Standards-Based Assessment Resources for Teachers	Student (or teacher) understanding of K-12 physical, earth, life science content	criterion K-12	science	CONTENT CONSTRUCT
NAEP	National Assessment of Educational Progress	Arts, civics, geography, reading, science, U.S. history, writing, mathematics (number properties and operations, measurement, geometry, data analysis and probability, algebra; ability: conceptual understanding, procedural knowledge, problem solving), science (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)	large/ criterion 4, 8, 12	general	interrater kappa (high) interrater % agree (high) ICC (high) CONTENT DISCRIMINANT CONSTRUCT

Acronym	Name	Student knowledge & reasoning variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain	Reliability type (level) VALIDITY EVIDENCE
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 focuses on mathematics.	large/ criterion AGE 15	science math literacy	Internal consistency (high) alternate form (low) interrater: correlation (high)
REMA	Research-Based Early Maths Assessment	Number: recognition of number, verbal counting and object counting, number comparison, number sequencing, number composition and decomposition, adding and subtracting, place value, and multiplication and division. Geometry: shape recognition, shape composition and decomposition, congruence, construction of shapes, spatial imagery (turns), geometric measurement, and patterning using geometric shapes.	criterion AGES 3-7	math	
SASKS	Science Attitudes, Skills, & Knowledge Survey	Demographics, science content in life, physical, Earth/space, biological sciences	survey POST	science	
SAT	Scholastic Aptitude Test	Reading (reading passages and sentence completions), writing (a short essay and multiple-choice questions on identifying errors and improving grammar and usage), and math (arithmetic operations, algebra, geometry, statistics and probability). Subject Tests available for: multiple foreign languages, literature, U.S. history, world history, math levels 1 and 2, biology, chemistry, physics.	large/ norm HIGH	general	
---	Science notebooks as assessment tool Rubric	13 general entry categories: Defining, Exemplifying, Applying Concepts, Predicting/Hypothesizing, Reporting Results, Interpreting Results and/or Concluding, Reporting & Interpreting Results and/or Concluding, Reporting Procedures, Reporting Experiments, Designing Experiments, Content Questions/Short Answers, Quick Writes (e.g. reflections), and Assessments.	student work K-8	science	interrater % (high) CONTENT CONVERGENT CRITERION

Acronym	Name	Student knowledge & reasoning variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain	Reliability type (level) VALIDITY EVIDENCE
---	Scientific Inquiry and Engineering Design Scoring Guides	Rubrics are separated by grades (3, 4-5, 6-8, and 9-12). *Scientific Inquiry* (grades 3-12): Forming a question or hypothesis, Defining an investigation, Collecting and presenting data, Analyzing and interpreting results *Engineering Design* (grades 4-12): Identifying and defining a problem to be solved, Generating possible solution, Testing solution(s) and collecting data, Analyzing and interpreting results.	student work 3-12	science engineering	
SEPUP	Science Education for Public Understanding Program assessments CURRICULUM TEST	Scoring guides assess ability to: design and conduct an investigation, analyze data, understand concepts, evaluate evidence and identify tradeoffs, communicate scientific information, work cooperatively in a group	criterion MIDDLE HIGH	science	NA—CURRICULUM TEST
SLA	Science Learning Assessment	Scientific Inquiry Processes: young children's functional understanding of the nature and processes of scientific inquiry. Life Science Concepts: children's understanding of specific science concepts related to living things and the physical world.	criterion K	science	internal consistency (accept) CONVERGENT DISCRIMINANT CONSTRUCT

Acronym	Name	Student knowledge & reasoning variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain	Reliability type (level) VALIDITY EVIDENCE
Stanford 10	Stanford Achievement Test	Reading, lexile measure, mathematics, language, spelling, listening, science, social science, science, math Science: life science, Earth science, physical science, and the nature of science are represented with questions that elicit problem solving and inquiry. Skills include estimating, making simple calculations, seeking patterns, making observations, recognizing cause and effect, reading standard instruments, and drawing conclusions. Math: number sense and operations; patterns, relationships, and algebra; geometry and measurement; and data, statistics, and probability. Questions assess processes in communication and representation; estimation; mathematical connections; and reasoning and problem solving. Mathematics Problem Solving measures the skills and knowledge necessary to solve problems in mathematics. Mathematics Procedures measures the ability to apply the rules and methods of arithmetic to problems that require arithmetic solutions.	large/ norm K-12	general	
---	Students' Use of Evidence in Written Scientific Explanations	A structural scheme that includes four major components of an explanation by natural selection: (a) the (changed) factor in the environment exerting selective pressure on individuals, (b) the effect that pressure had on individuals, (c) the trait that differentiated survivors from fatalities, and (d) the selective advantage of the trait (differential fitness). Assessed (1) the conceptual quality of students' explanations in terms of the extent to which their explanations articulated and provided warrants for each of the four components (2) the sufficiency of the data students cited as evidence for their claims.	student work HIGH	science	internal (accept) interrater % (high)

Acronym	Name	Student knowledge & reasoning variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain	Reliability type (level) VALIDITY EVIDENCE
SUSSI	Student Understanding of Science and Scientific Inquiry	Targets six themes: Observations and Inferences, Tentative Nature of Scientific Theories, Scientific Laws vs. Theories, Social and Cultural Influence on Science, Imagination and Creativity in Scientific Investigations, and Methodology in Scientific Investigations. Each theme consists of items that represent both most common naïve ideas and informed views consistent with the standards documents and current nature of science literature.	survey POST	science	internal (accept) interrater % (high) CONTENT
TAGLIT	Taking a Good Look at Instructional Technology	Collects information about student technology literacy regarding resources, skills, knowledge, use and application in the classroom. Topics: basic concepts & operations; social, ethical and human issues; technology productivity tools; technology communications tools; technology research tools (middle/high school version), technology problem-solving and decision-making tools (middle/high version)	criterion K-16 ADULT	technology	
TELS	Technology Enhanced Learning in Science (TELS) assessment activities CURRICULUM TEST	Measures how successfully students evaluate and change their perceptions about science in light of new information. Students link and connect ideas and give explanations for their conjectures, consistent with the knowledge integration framework that guides design of TELS curriculum. TELS also created six benchmark tests measuring knowledge integration. The tests cover middle school physical science, life science, and earth science, as well as high school biology, chemistry, and physics.	student work 6-12	science	NA—CURRICULUM TEST
---	TELS Knowledge Integration Scoring Rubric CURRICULUM TEST	Measures how successfully students evaluate and change their perceptions about science in light of new information. The scoring rubric for captures progressively more sophisticated levels of knowledge integration (KI) in student responses.	student work 6-12	science	NA—CURRICULUM TEST

Acronym	Name	Student knowledge & reasoning variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain	Reliability type (level) VALIDITY EVIDENCE
TEMA-3	Test of Early Mathematics Ability-Third Edition	Measures concepts and skills in the following domains: numbering skills, number-comparison facility, numeral literacy; mastery of number facts, calculation skills, understanding of concepts	norm 3-8	math	test-retest (high) internal consistency (high) alternate form (high) CONTENT CRITERION CONSTRUCT
---	Terra Nova Algebra Assessment	Reading, language, social studies, math and science. Math: number and number relations; computation and numerical estimation; operation concepts; measurement; geometry and spatial sense; data analysis, statistics, and probability; patterns, functions, algebra; problem solving and reasoning; communication Science: science inquiry; physical science; life science; Earth and space science; science and technology; personal and social perspectives in science; history and nature of science	large/ norm K-12	general	

Acronym	Name	Student knowledge & reasoning variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain	Reliability type (level) VALIDITY EVIDENCE
TIMSS	Trends in International Mathematics and Science Study	Grade 4 science and mathematics test items- Science 1995: earth science; life science; physical science; and environmental issues and the nature of 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. 2003: patterns and relationships; data; geometry; measurement; number Grade 8 science and mathematics test items- Science 1999: earth science; life science; physics; chemistry; environmental and resource issues; and scientific inquiry and the nature of 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 2003: algebra; data; geometry; measurement; and number	large/ criterion 4, 8	science math	
---	Upper-Elementary Mathematics Assessment Modules	Measure fourth and fifth grade students' mathematical knowledge (e.g., reasons, fundamental principles).	criterion 4-5	math	

Appendix B: Instruments to Assess Students' Literacy and Language Skills (alphabetical by acronym)

Acronym	Name	Literacy and language skill variables measured/ scales	Type of Tool GRADE LEVEL	Construct Assessed	Reliability type/ level VALIDITY EVIDENCE
CELDT	California English Language Development Test Reading Scale	Assesses the English language development of pupils whose primary language is a language other than English, specifically listening, speaking, reading and writing skills.	criterion K-12	language ELL	
EECA	Early Expository Comprehension Assessment	Includes a compare/contrast passage, manipulatives to represent the information in the paragraph, and three response tasks (Retelling, Mapping, and Comparing)	criterion PK	literacy	alternate form (Not Reported)
EVT-2	Expressive Vocabulary Test-2	Measures expressive vocabulary and word retrieval of the spoken word in standard American English and thus assesses vocabulary acquisition.	norm PK-ADULT	language	test-retest (high) internal consistency (high) alternate form (high) CONTENT CONVERGENT CONSTRUCT
GMRT	Gates-MacGinitie Reading Test	Reading comprehension and vocabulary	criterion 8	literacy	internal consistency (high) CRITERION
LAS Links	Language Assessment Scales Links	Measure language proficiency within five grade spans: K-1, 2-3, 4-5, 6-8, and 9-12. Measures speaking, listening, reading, writing, and comprehension skills in academic and social English, derived from the listening and reading domains. LAS Links blends the assessment of English proficiency with content area knowledge, using vocabulary that is needed for success in school.	criterion K-12	language/ literacy	

Acronym	Name	Literacy and language skill variables measured/ scales	Type of Tool GRADE LEVEL	Construct Assessed	Reliability type/ level VALIDITY EVIDENCE
OPI	Oral Proficiency Interview	A global assessment of functional speaking ability. Determines how well a person speaks a language by comparing his or her performance of specific communication tasks with the criteria for each of ten proficiency levels. The topics that are discussed during the interview are based on the interests and experiences of the test candidate. Through a series of personalized questions, the interviewer elicits from the test candidate examples of his or her ability to handle the communication tasks specified for each level of proficiency in order to establish a clear 'floor' and 'ceiling' of consistent functional ability.	interview ALL	language	internal consistency (Not Reported) CONSTRUCT
PPVT-4	Peabody Picture Vocabulary Test-Fourth Edition	Measures understanding of the spoken word in standard American English--vocabulary acquisition.	norm PK-ADULT	language	test-retest (high) internal consistency (high) alternate form (high) CONTENT CONVERGENT CONSTRUCT
RBS-NA	The Renfrew Bus Story, North American Edition	Measures: sentence length (mean of 5 longest utterances), complexity (number of complex utterances), information (vocabulary/key words), independence (inverse of amount of prompting needed for retell), and behavior (observation score for listening and attention behavior). Standard scores and percentiles available for Sentence length, Complexity, and Information subscales. Additional qualitative scales provide a profile of a child's language, learning, and work style.	other PK	language	Test-retest (accept) interrater: correlation (accept) CRITERION

Appendix C: State instruments to Assess Students' Academic Achievement (alphabetical by state)

State	Acronym	Name	Academic achievement variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain
AK	SBA	Alaska Standards Based Assessment	Math: numeration, measurement, estimation & computation, functions & relationships, geometry, statistics/probability Science: Inquiry, Technology, and Nature of Science; earthy, physical, life science	criterion 3-10	science math literacy general
AL	ARMT	Alabama Reading and Mathematics Test	Grade 3-8: reading, language, math (number & operations, algebra, geometry, measurement, data analysis and probability) Grade 5 and 7: science Grade 6: social studies	criterion 3-8	science math literacy social studies
CA	CAT	California Achievement Test	Assesses reading, language, spelling, and mathematics in grades 3 through 7.	norm 3-7	science math literacy
CA	CST	California Standards Test	English-language arts, history-social science, mathematics (number sense; algebra & functions; measurement & geometry; statistics, data analysis & probability; mathematical reasoning), science (physical, life, earth sciences; investigation and experimentation)	criterion 2-11	science math literacy general social studies
CO	CSAP	Colorado Student Assessment Program	Tests students in reading, writing and math in third grade through tenth grade. Students in 5th, 8th and 10th grade also take a science test.	criterion 3-10	science math literacy general

State	Acronym	Name	Academic achievement variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain
IL	ISAT	Illinois Standards Achievement Test	3rd through 8th grade take reading and math (number sense, estimation and measurement, algebra and analytical methods, geometry, data analysis and probability). 4th and 7th grader take science (inquiry and design; concepts and principle; science, technology and society). 3rd, 5th, 6th and 8th graders take a writing test. In addition, schools can choose to administer ISAT tests in physical development and health, social science and fine arts.	criterion 3-8	science math literacy general
IN	ISTEP	Indiana Statewide Testing for Educational Progress	3rd through 8th grade students take English language arts and math (number sense, computation, algebra and functions, geometry, measurement, data analysis and probability, problem solving); 4th and 6th grade students take science (nature of science and technology; scientific thinking, the physical setting, the living environment, the mathematical world, historical perspectives and common themes); 5th and 7th grade students take social studies	criterion 3-8	science math
KY	KCCT	Kentucky Core Content Test	Beginning with the 2011-12 school year, Kentucky enacted a new state assessment accountability system known as the Kentucky Performance Rating for Education Progress (K-PREP). K-PREP math: numbers and algebraic thinking; number and operations in base ten; number and operations--fractions; measurement and data, geometry; ratios and proportional relationships; the number system; expressions of equations; geometry; statistics and probability; K-PREP science: life, earth/space, physical sciences, unifying ideas	criterion 3-8, 10-12	science math literacy general social studies
LA	iLEAP	Integrated Louisiana Educational Assessment Program	Students in 3rd, 5th, 6th and 7th grade take English language arts, math, science and social studies. Students in 9th grade take English language arts and math only. Math: number and number relations; algebra; measurement; geometry; data analysis, probability and discrete math; patterns, relations and functions. Science: science as inquiry; physical science; life science; earth and space science; science and the environment	criterion 3, 5-7	science math literacy general social studies

State	Acronym	Name	Academic achievement variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain
MA	MCAS	Massachusetts Comprehensive Assessment System	3rd through 8th grade and 10th grade take English language arts and math (number sense and operations; patterns, relations, and algebra; geometry; measurement; data analysis, statistics, and probability). Students in 5th and 8th grade also take a science test (Earth and space, life, physical, technology/engineering). There is a high school requirement to pass one of the following for graduation: introductory physics (motion and forces; heat and heat transfer; waves and radiation; electromagnetism), chemistry (properties of matter and thermochemistry; atomic structure and periodicity; bonding and reactions; solutions, equilibrium, and acid-base theory), biology (biochemistry and cell biology; genetics; anatomy and physiology; ecology; evolution and biodiversity), engineering (engineering design; construction and manufacturing; fluid and thermal systems; electrical and communication systems)	criterion 3-8,10	science math literacy general
MD	HSA	Maryland High School Assessments	Measure school and individual student progress toward Maryland's High School Core Learning Goals in in English, Algebra/Data Analysis (analyzing pattern and functions; modeling real-world situations; collecting, organizing, and analyzing data; using data to make predictions) and Biology (skills and processes of biology; structure and function of biologic molecules; structure and functions of cells and organisms; inheritance of traits; mechanisms of evolutionary change; interdependence of organisms in the biosphere).	criterion HIGH	science math literacy general government
MD	MSA	Maryland School Assessment	Science: skills and processes; earth/space science; life science; chemistry, physics, environmental Math: algebra/patterns, geometry/measurement, statistics/probability; number concepts/computation	criterion 3-8	science math literacy general
MI	MEAP	Michigan Educational Assessment Program	Tests students in third grade through eighth grade in reading, writing, English language arts and math. Students in 5th grade and 8th grade also take a science test, while students in 6th grade and 9th grade take a social studies test.	criterion 3-9	science math literacy social studies

State	Acronym	Name	Academic achievement variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain
MO	MAP	Missouri Assessment Program	MAPs were last used in 2008. Math: algebraic relationships, data and probability, geometric and spatial relationship, measurement, number and operations. Science: characteristics of living organisms, Earth's processes, force and motion, interactions of organisms, matter and energy, scientific inquiry, technology and the environment	criterion 3-8	science math literacy social studies
MS	MCT2	Mississippi Curriculum Test	Students in the 3rd through 8th grade take language arts and math (numbers and operations, algebra, geometry, measurement, data analysis and probability)	criterion 3-8	math literacy
MS	MST2	Mississippi Science Test	Students in 5th and 8th grade take content tests in inquiry, physical, life, Earth/space science	criterion 5,8	science
NC	EOG	North Carolina End-of-Grade Tests	The reading and math tests are given to students in 3rd, 4th, 5th, 6th, 7th, and 8th grades. Fifth graders and eighth graders also take the science test. Mathematics: five strands or goals: number and operations, measurement, geometry, data analysis and probability, and algebra. Science: demonstrate knowledge of important principles and concepts, understand and interpret laboratory activities, and relate scientific information to everyday situations	criterion 3-8	science math literacy general
NE	CRT	Omaha Public Schools Criterion-Referenced Tests	This test is no longer in use. Administered to students in grades 1-6 in the areas of reading/language arts, mathematics and science.	criterion 1-6	science math literacy
NJ	GEPA	Grade Eight Proficiency Assessment	This test is no longer used in NJ since it was replaced by the ASK 8. GEPA is a primary indicator for identifying eighth-grade students who may need instructional intervention in: language arts literacy, social studies, mathematics and science.	criterion 8	science math literacy general

State	Acronym	Name	Academic achievement variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain
NJ	NJ ASK	New Jersey Assessment of Skills and Knowledge	Measures students' critical thinking skills in the major content areas. Students in 3rd - 8th grades take tests in math and language arts literacy (reading, writing, speaking, listening, and viewing). Science exams are given to 4th and 8th grade. Math: number and numerical operations; geometry and measurement; geometry and algebra; data analysis, probability and discrete mathematics; mathematical processes Science: scientific processes; science and society; nature and process of technology; mathematical applications; life science; physical science--chemistry, physics; Earth science; astronomy and space science; environmental studies	criterion 3-8	science math literacy general
NM	NMSBA	New Mexico Standards Based Assessment	Students in 3rd, 4th, 5th, 6th, 7th, and 8th grades take tests in reading, math, writing, and science. Math: numbers and operations, algebra, geometry, measurement, data analysis and probability science: nature of science, physical, life, earth science, science and society	criterion 3-8	science math literacy general
NY	Regents	New York Regents High School Examination	High school students in New York State take Regents Exams to assess their mastery of New York State Learning Standards. In order to graduate from high school, students must pass exams in five subjects: English, mathematics, science, global history, and U.S. history and government. Algebra 2/Trigonometry, Comprehensive English, Geometry, Global History & Geography, Integrated Algebra, Living Environment, Physical Setting/Chemistry, Physical Setting/Earth Science, Physical Setting/Physics, U.S. History & Government, Global Studies, Mathematics, Reading, Science, and Writing.	criterion HIGH	science math literacy general foreign language
NY	NYSTP	New York State Testing Program	Students in 3rd - 8th grades take English language arts, including writing, and math; 4th and 8th grade take science and social studies. science: scientific inquiry, information systems, the living environment, interconnectedness, interdisciplinary problem solving math: number sense and operations, algebra, geometry, measurement, statistics and probability	criterion 3-8	science math literacy general social studies

State	Acronym	Name	Academic achievement variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain
OH	OAA	Ohio Achievement Assessments	Reading and math, grades 3 to 8; writing, grades 4 and 7; science, grades 5 and 8; social studies, grades 5 and 8. Mathematics, grades 3 to 8: number, number sense, and operations, measurement, geometry and spatial sense, patterns, functions, and algebra, data analysis and probability. Science, grades 5 and 8: science and technology, scientific inquiry, scientific ways of knowing, Earth and space sciences, life sciences, physical sciences.	criterion 3-8	science math literacy social studies writing
OH	OGT	Ohio Graduation Tests	Math: number, number sense and operations; measurement; geometry and spatial sense; patterns, functions, and algebra; data analysis and probability Science: science and technology, scientific inquiry and scientific ways of knowing; Earth and space sciences; life sciences; physical sciences	criterion 10	science math literacy social studies
OR	OAKS	Oregon Assessment of Knowledge and Skills	grades 3-8 and 10: reading and math (number and operations; algebra; data analysis; geometry; measurement; probability; statistics) grades 4, 7 and 10: writing grades 5, 8 and 10: science (structure and function; interaction and change; science processes--scientific inquiry, engineering design; physical science; life science; Earth and space science)	criterion 3-8,10	science math literacy
PA	PSSA	Pennsylvania System of School Assessment	3rd - 8th, 11th Grades: reading and math (numbers and operations, measurement, geometry, algebraic thinking, data analysis and probability) 4th, 8th, 11th Grades: science (nature of science, biological sciences, physical sciences, and Earth and space sciences) 5th, 8th, 11th Grades: writing	criterion 3-8,11	science math literacy general

State	Acronym	Name	Academic achievement variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain
TX	TAKS	Texas Assessment of Knowledge and Skills	TAKS tests are aligned to the Texas Essential Knowledge and Skills (TEKS) learning standards. Reading: 3rd, 4th, 5th, 6th, 7th, 8th, and 9th grades Math (3rd, 4th, 5th, 6th, 7th, 8th, 9th, 10th, and 11th grades): numbers, operations and quantitative reasoning; patterns, relationships and algebraic reasoning; algebra; geometry and spatial reasoning; measurement; probability and statistics; mathematical processes and tools Writing: 4th and 7th grades English Language Arts: 10th and 11th grades Science (5th, 8th, 10th, and 11th grades): nature of science; organization of living systems; interdependence of organisms; structures and properties of matter; motion, forces, and energy; life, Earth, physical, biological sciences Social Studies: 8th, 10th, and 11th grades	criterion 3-11	science math literacy general social studies
UT	U-PASS	Utah Performance Assessment System for Students	English Language Arts: 2nd - 11th grades. Math: 2nd - 7th grades and course specific (Algebra 1 and 2, geometry, and pre-algebra.) Science: 4th - 8th grades and course specific (Earth systems, biology, physics, and chemistry) Direct Writing Assessment: 6th and 9th grades. Utah Basic Skills Competency Test (UBSCT): Tenth grade students take in reading, writing, and math, and must pass in order to graduate.	criterion 2-11	science math literacy general
VA	SOL	Virginia Standards of Learning Assessments	Reading, Math: 3rd - 8th grades Writing, History/Social Studies: 5th and 8th grades Science: 5th, and 8th grades Math: number, number sense, computation and estimation; measurement and geometry; probability, statistics, patterns, functions, and algebra 5th grade science: scientific investigation, reasoning, and logic; force, motion, energy, and matter; life processes and living systems; ecosystems; Earth and space systems and cycles; chemistry; biology 8th grade science: systems (inputs, outputs, boundaries and flows), inquiry (questioning and investigation), application (science, technology and problem solving), physical science (balanced and unbalanced forces, atoms and molecules, interactions of energy and matter), Earth and space science (the solar system, cycles in Earth systems, evidence of change), life science (from cells to organisms; flow of energy through ecosystems; inheritance, variation and adaptation)	criterion 3-8	science math literacy general social studies

State	Acronym	Name	Academic achievement variables measured/ scales	Type of tool GRADE LEVEL	Subject Domain
WA	MSP	Washington Measurements of Student Progress	Reading and Math: 3rd-8th grades Math: numbers, operations, algebra; geometry/measurement; data/statistics/probability; reasoning, problem solving, communication Writing: 4th and 7th grades Science: 5th and 8th grades: complex systems, planning investigations, application of technology, physical science (measurement of force and motion; states of matter; heat, light, sound and electricity), Earth and space science (Earth in space, formation of earth materials, focus on fossils), life science (structures and behaviors, food webs, heredity and adaptation)	criterion 3-8	science math literacy general

Appendix D: Instruments to Assess Students' Attitudes (alphabetical by acronym)

Acronym	Name	Attitudinal variables measured/ scales	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
ATMI	Attitudes Toward Mathematics Inventory	Attitudes toward mathematics: self-confidence (students' confidence and self-concept of their performance in mathematics); value (students' beliefs on the usefulness, relevance and worth of mathematics in their life now and in the future); enjoyment (degree to which students enjoy working mathematics and mathematics classes); and motivation (interest in mathematics and desire to pursue studies in mathematics).	survey MIDDLE HIGH	math	test-retest (high) internal consistency (high) CONTENT
ATSSA	Attitude Toward Science in School Assessment	General attitude toward science as an academic subject.	survey 7-8	science	internal consistency (high)
	AWE Pre-College Recruiting Surveys	*Lower and Upper Elementary Surveys*: Attitude toward science and/or engineering; interest and awareness of science and/or engineering *Middle and High School Surveys* (Engineering, Science, or Computer Science): Course-taking plans for high school, whether participant intends to study science, engineering, or computer science; what participant knows about what engineers, scientists, or computer scientists do; what factors (if any) about being an engineer, scientist, or computer scientist appeal to participant; events or persons that influenced participants' study plans; participant skill and confidence level in areas that are important for successfully completing a science, engineering, or computer degree; where participants plan to study science, engineering, or computer science/engineering; satisfaction with the quality of the activity in which s/he has participated.	survey MIDDLE HIGH	engineering science computer science	
CSABA	Computer Science Attitudes and Beliefs Assessment	Confidence (student assurance in their own ability to learn computing skills), interest (student interest in computing), gender (students' perceptions of computing as a male-dominated field), usefulness (beliefs about the value of learning computing), and professional (student beliefs about professionals in computing).	survey HIGH	technology	internal alpha (high) CONTENT

Acronym	Name	Attitudinal variables measured/ scales	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
DAET	Draw an Engineer Test	Questionnaire contains the following five questions on one page: In your own words, what is engineering?, What does an engineer do?, Draw a picture of an engineer at work. Do you know any engineers? If yes, who are they? The coding system has seven major classifications: Humans, human-engineered objects, system, environment, vibe, engineering field portrayed, and engineering understanding.	performance assessment with coding rubric and interview ELEM MIDDLE	engineering	interrater % agree (high) CONTENT CONCURRENT
DAST	Draw-A-Scientist Test	Indicators of a standard image of a scientist: lab coat (usually but not necessarily white); eyeglasses; facial growth of hair (including beards, mustaches, or abnormally long sideburns); symbols of research: scientific instruments and laboratory equipment of any kind; symbols of knowledge: principally books and filing cabinets technology; the products of science; relevant captions: formulae, taxonomic classification, the eureka! syndrome, etc. Drawings are analyzed and given a score to indicate the extent to which the standard image was present.	student work K-5	science	
mATSI	Modified Attitudes towards Science Inventory	Measures multiply dimensions of student attitudes towards science: Perception of the Science Teacher, Anxiety toward Science, Value of Science in Society, Self-concept of Science, and Desire to do science.	survey 5	science	internal consistency (accept) CONTENT CONSTRUCT
SAI II	The Scientific Attitude Inventory: A revision	Assesses students interest in science, their attitudes toward science, their views of scientists, and their desire to become scientists: Engagement, Attitude, Competence	survey 6,9,12	science	internal consistency (accept) interrater split-half (high)
SPOCC	Student Perceptions of Classroom Climate	Attitudes about and perceptions of science instruction	survey ELEM	science	

Acronym	Name	Attitudinal variables measured/ scales	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
TOSRA	Test of Science-Related Attitudes	Assesses science-related attitudes and engagement: social implications of science, normality of scientists, attitude toward scientific inquiry, adoption of scientific attitudes, enjoyment of science lessons, leisure interest in science, & career interest in science.	survey POST	science	test retest (accept) internal consistency (high) DISCRIMINANT
VASS	Views about Science Survey	Students' learning styles, attitudes toward science education, and views about the nature of science.	survey HIGH POST	science	internal consistency test-retest alternate form
VNOS	Views of Nature of Science Questionnaire	Empirical NOS: Science is based, at least partially, on observations of the natural world; Tentative NOS: Scientific knowledge is subject to change and never absolute or certain; Inferential NOS: The crucial distinction between scientific claims (e.g., inferences) and evidence on which such claims are based (e.g., observations); Creative NOS: The generation of scientific knowledge involves human imagination and creativity; Theory-laden NOS: Scientific knowledge and investigation are influenced by scientists' theoretical and disciplinary commitments, beliefs, prior knowledge, training, experiences, and expectations; Social and cultural NOS: Science as a human enterprise is practiced within, affects, and is affected by, a lager social and cultural milieu; Myth of the "Scientific Method": The lack of a universal step-wise method that guarantees the generation of valid knowledge; and Nature of, and distinction between scientific theories and laws (e.g., lack of a hierarchical relationship between theories and laws).	survey K-12	science	CONTENT CONVERGENT CONSTRUCT
VOSI	Views of Science Inquiry	Elicits details of learners' ideas of what scientists do in the production of valid scientific knowledge. Responses on the VOSI provide descriptors and examples, as opposed to scores (numbers) or dichotomous categories (naïve/informed).	survey K-12	science	CONTENT

Appendix E: Instruments to Assess Students' Emotional Attributes (alphabetical by acronym)

Acronym	Name	Emotional attribute variables measured/ scales	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
AGQ	Achievement Goals Questionnaire	Measures achievement goal orientation in the form of four factors: mastery approach, mastery avoidance, performance approach, and performance avoidance	survey POST	goals	internal (accept) CONVERGENT DISCRIMINANT CRITERION
EATQ-R	Early Adolescent Temperament Questionnaire - Revised	Temperament Scales: Activation Control (The capacity to perform an action when there is a strong tendency to avoid it), Affiliation (The desire for warmth and closeness with others, independent of shyness or extraversion), Attention (The capacity to focus attention as well as to shift attention when desired), Fear (Unpleasant affect related to anticipation of distress), Frustration (Negative affect related to interruption of ongoing tasks or goal blocking), High Intensity Pleasure (The pleasure derived from activities involving high intensity or novelty), Inhibitory Control (The capacity to plan, and to suppress inappropriate responses), Perceptual Sensitivity (Detection or perceptual awareness of slight, low-intensity stimulation in the environment), Pleasure Sensitivity (Pleasure related to activities or stimuli involving low intensity, rate, complexity, novelty, and incongruity), Shyness (Behavioral inhibition to novelty and challenge, especially social) Behavioral Scales: Aggression (Hostile and aggressive actions, including person- and object-directed physical violence, direct and indirect verbal aggression, and hostile reactivity), Depressive Mood (Unpleasant affect and lowered mood, loss of enjoyment and interest in activities)	survey AGES 9-15	temperament	internal consistency (accept) CONVERGENT DISCRIMINANT
ERC	Emotion Regulation Checklist	Assessing lability/negativity (lack of flexibility, lability, and disregulated affect) and emotion regulation (situationally appropriate emotional displays, empathy, and self-awareness)	survey PK-5	emotion regulation	internal consistency (high) CONVERGENT DISCRIMINANT CONSTRUCT

Acronym	Name	Emotional attribute variables measured/ scales	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
GOALS-S	Goal Orientation and Learning Strategies Survey	Measure students' motivational goal orientations and their cognitive and metacognitive strategies. Three academic goals, four social goals, three cognitive strategies, and three metacognitive strategies were defined and operationalized in the developed GOALS-S instrument.	survey HIGH	goals	 CONSTRUCT
PALS	Patterns of Adaptive Learning Scales	Assesses: personal achievement goal orientations; perceptions of teacher's goals; perceptions of the goal structures in the classroom; achievement-related beliefs, attitudes, and strategies; and perceptions of parents and home life.	survey K-12	goals	internal consistency (accept) CONVERGENT
SDQII	Self-Description Questionnaire II	Measures the following areas (child and early adolescent versions exclude some of these factors): Math, Physical Appearance, General Esteem, Honesty/Trustworthiness, Physical Abilities, Verbal, Emotional Stability, Parent Relationships, Academic (General), Same-Sex Relationships, and Opposite-Sex Relationships.	survey 7-12	self-concept	test-retest (accept) internal (high) CONVERGENT DISCRIMINANT CONSTRUCT
SSIS-RS	Social Skills Improvement System-Rating Scales	Social Skills: Communication, Cooperation, Assertion, Responsibility, Empathy, Engagement, Self-Control Competing Problem Behaviors: Externalizing, Bullying, Hyperactivity/Inattention, Internalizing, Autism Spectrum Academic Competence: Reading Achievement, Math Achievement, Motivation to Learn	norm PK-12	social skills	test-retest (high) internal consistency (high) CONVERGENT
WKCT	Wallach-Kogan Creativity Test	Five measures of creativity, each of which yields a score for originality and fluency: instances test-- generate instances for four categories (e.g., things that are round); alternate uses test-- generate uses for eight common objects (e.g., newspaper, knife, cork); similarities test-- generate ways in which pairs of objects are similar for 10 object pairs (e.g., milk and meat, a train and a tractor); pattern meanings test-- view 8 abstract visual designs and suggested interpretations of the patterns; line meanings test-- view 9 pictures of continuous lines and suggested interpretations of the lines.	norm ELEM	creativity	internal (low) DISCRIMINANT CONSTRUCT

Appendix F: Instruments to Assess Students' Motivational Attributes (alphabetical by acronym)

Acronym	Name	Motivational attribute variables measured/ scales	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
	Children's self- and task perceptions during elementary school	Beliefs about academic subjects, instrumental music, and sports, as well as other constructs. In the academic area, children were asked about mathematics and reading. The subjective task value scale for each domain except music includes the items tapping perceived usefulness, perceived importance, intrinsic interest, and liking. For the music domain, the scale includes only the importance and interest items. The math subscale was used for the DR-K12 project.	survey ELEM	math literacy sports music motivation	internal (accept) DISCRIMINANT CRITERION CONSTRUCT
	Children's Self-Efficacy Scale	Assesses beliefs in one's capabilities to do or achieve tasks and skills in all arenas of children's lives. Subscales: self-efficacy in enlisting social resources; self-efficacy for academic achievement; self-efficacy for self-regulated learning; self-efficacy for leisure time skills and extracurricular activities; self-regulatory efficiency; self-efficacy to meet others' expectations; social self-efficacy; self-assertive efficacy; self-efficacy for enlisting parental and community support.	survey K-POST	self-efficacy	internal (accept) CONTENT CONVERGENT CONSTRUCT CRITERION
FFT	Framework for Teaching Evaluation Instrument	Four domains of teaching responsibility: Planning and Preparation, Classroom Environment, Instruction, Professional Responsibilities	observation scoring rubric (T) K-12	Student engagement	CONVERGENT
IMI	Intrinsic Motivation Inventory	Assesses six subscales: participants' interest/enjoyment, perceived competence, effort, value/usefulness, felt pressure and tension, and perceived choice while performing a given activity. The interest/enjoyment subscale is considered the self-report measure of intrinsic motivation.	survey K-ADULT	motivation	internal (high) CONSTRUCT
IMMS	Instructional Material Motivational Survey	Attention: perceptual arousal, inquiry arousal and variability; Relevance: goal orientation and motive matching; Confidence: learning requirements, success opportunities and personal responsibility; and Satisfaction: intrinsic reinforcement, extrinsic rewards and equity.	survey POST	motivation	

Acronym	Name	Motivational attribute variables measured/ scales	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) Student behavior: Asks others to explain their thinking (rigorous thinking); Explains thinking (rigorous thinking); Uses evidence & asks others to use evidence to support claims (evidence); Links contributions to other contributions (linking) Classroom culture and climate: Widespread participation in teacher-facilitated discussion (equity)	observation scoring rubric ELEM	literacy math student engagement	internal consistency (Not Reported) interrater % agree (low) DISCRIMINANT
Jr. MAI	Junior Metacognitive Awareness Inventory	Measures elementary school students' general metacognitive awareness in the form of two factors: knowledge of cognition and regulation of cognition. Version A consists of 12 items for grades 3 through 5. Version B includes an additional 6 items for grades 6 through 8.	survey 3-8	Meta-cognition	internal (accept) CONVERGENT CONSTRUCT
MSLQ	Motivated Strategies for Learning Questionnaire	Assesses motivation and use of learning strategies: value (intrinsic and extrinsic goal orientation, task value); expectancy (control beliefs about learning, self-efficacy); and affect (test anxiety). The learning strategies section measures cognitive, metacognitive, and resource management strategies. The cognitive strategies scales include (a) rehearsal, (b) elaboration, (c) organization, and (d) critical thinking. Metacognitive strategies are assessed by one large scale that includes planning, monitoring, and regulating strategies. Resource management strategies include (a) managing time and study environment; (b) effort management, (c) peer learning, and (d) help-seeking.	survey POST	Meta-cognition	CRITERION

Acronym	Name	Motivational attribute variables measured/ scales	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
NASA-TLX	NASA Task Load Index	These subscales include Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort and Frustration. Can be used in various human-machine environments such as aircraft cockpits, command, control, and communication (C3) workstations; supervisory and process control environments; simulations and laboratory tests.	student work ADULT	task demands	test retest (high)
SCE	Student Collective Engagement	Assesses five dependent measures: teachers' autonomy support; teachers' provision of structure; teachers' provision of involvement; and two measures of students' engagement--(1) students' active task involvement during instruction (attention, effort, verbal participation, persistence, and positive emotion) and (2) students' voice and initiative in trying to take personal responsibility for their learning (students' active attempts to influence the flow of classroom events).	observation protocol (student & teacher components) K-12	student engagement	internal consistency (high) interrater (acceptable)
SETS	Self-Efficacy in Technology and Science	Self-concept (self-efficacy) subscales: science inquiry, video gaming, computer gaming, general computer use, problem-solving computer use, synchronous chat use	survey MIDDLE	technology self-efficacy	internal consistency (high) CONTENT CONSTRUCT
SSSE	Sources of Science Self-Efficacy Scale	Mastery experience, vicarious experience, verbal persuasions, and emotional arousal	survey MIDDLE	science self-efficacy	test-retest (high) internal (accept)

Appendix G: Instruments to Assess Students' Career Identity (alphabetical by acronym)

Acronym	Name	Career identity variables measured/ scales	Type of Tool GRADE LEVEL	Subject Domain	Reliability type/ level VALIDITY EVIDENCE
EIDS	Engineering Identity Development Scale	1) academic identity (self-beliefs or self-images in who children think they are as students) 2) school identity (children's affiliation or attachment to their school) 3) occupational identity (children's self-understandings of an occupation) 4) engineering aspirations (children's self-goals, aims, or objectives of becoming an engineer).	survey K-5	engineering identity	test-retest (low) internal consistency (accept) CONTENT DISCRIMINANT
KCS	Kuder Career Search	Reports directly on similarity with groups in six well-known career clusters: outdoor/Mechanical, Science/Technical, Arts/Communication, Social/Personal Services, Sales/Management, and Business Operations.	survey MIDDLE-ADULT	career interests	test-retest (high) internal consistency (mod) CONTENT CONVERGENT
Strong	Strong Interest Inventory	Measures career and leisure interests: general occupational themes, basic interest scales, personal style scales, and occupational scales	survey UNSURE	career interests	test retest (high) internal consistency (high) CONVERGENT DISCRIMINANT CRITERION

Appendix H: Listing of Student Instruments and Access Details (alphabetical by full name)

Instrument Name	Information and Resources	Access Details
4-H Robotics Concept Test	Barker, B. & Ansorge, J. (2007). Robotics as means to increase achievement scores in an informal learning environment. Journal of Research on Technology in Education, 39(3), 229-243. www.csavvy.org/wp-content/uploads/2012/07/Robotics-As-Means-To-Increase-Achievement-Score.pdf	Post-test items reported in Table 3 of Barker & Ansorge (2007) paper.
AAAS Item Bank	Project Overview: http://assessment.aaas.org/ Item analysis form used in expert review for content validity: http://www.project2061.org/research/assessment/assessment_form.htm	Free to educators. For questions about other uses of the resources, such as requests to include those resources in published materials, please contact Barbara Goldstein at bgoldste@aaas.org.
Achievement Goals Questionnaire (AGQ)	One version of the instrument: http://www.uiowa.edu/~c07p075a/class1/Classhandouts/achgoalsquestionnaire.htm Elliot, A., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. Journal of Personality and Social Psychology, 80, 501-519. http://www.selfdeterminationtheory.org/SDT/documents/2001_ElliotMcGregor.pdf	Full instrument published online and in several papers. *Note*: Updated version available: http://www.learnlab.org/research/wiki/images/3/37/Achievement_Goal_Questionnaire.pdf
Acuity Algebra Assessment	Instrument Overview: http://www.acuityforschool.com/details/subject_areas/algebra.shtml Product page: http://www.ctb.com/ctb.com/control/ctbProductViewAction?productId=709&p=products "The Research-Based Acuity In Formative Assessment Solution" http://www.ctb-e.com/emails/pdfs/brAcuityResearchCTB.pdf	Must order from institution.

Instrument Name	Information and Resources	Access Details
American Chemical Society Division of Chemical Education Examinations Institute (ACS Exams)	http://chemexams.chem.iastate.edu/materials/exams.cfm	Must order from institution.
American College Testing of Science Reasoning (ACT Science)	Instrument Overview: http://www.actstudent.org/testprep/descriptions/scidescript.html ACT Science College Readiness Standards: http://www.act.org/standard/planact/science/index.html ACT Technical Manual: http://www.act.org/aap/pdf/ACT_Technical_Manual.pdf	
Attitudes Toward Mathematics Inventory (ATMI)	Tapia, Martha and George E. Marsh II (2004). "An instrument to measure mathematics attitudes." Academic Exchange Quarterly, 8, 2534-2541. http://www.rapidintellect.com/AEQweb/cho25344l.htm	The complete inventory is available from the first author upon request (Martha Tapia, mtapia@berry.edu).
Attitude Toward Science in School Assessment (ATSSA)	ATSSA summary page at PEAR Assessment Tools in Informal Science (ATIS): http://www.pearweb.org/atis/tools/33	Instrument text downloadable from PEAR after submitting name and email. Paper on development requires subscription or purchase.
AWE Pre-College Recruiting Surveys	AWE Surveys homepage: http://www.engr.psu.edu/awe/ Direct link to instrument page (requires registration): https://www.engr.psu.edu/awe/secured/director/precollege/pre_college.aspx	Available at no charge after registering on the website. All AWE surveys are formatted online using SurveyMonkey and are available for transfer to other SurveyMonkey online accounts. AWE will also host your surveys, under specific conditions.

Instrument Name	Information and Resources	Access Details
Building Blocks Assessment (BBA)	Overview: https://www.mheonline.com/program/view/4/4/2626/0076BB2012 Clements, Douglas H. and Julie Sarama (2007). "Effects of a Preschool Mathematics Curriculum: Summative Research on the Building Blocks Project." Journal for Research in Mathematics Education, 38(2), 136-163. http://gse.buffalo.edu/fas/clements/files/Clements_BB_JRME.pdf (all details on constructs, reliability, and validity found here, pg. 147) Building Blocks project website: http://gse.buffalo.edu/org/buildingblocks/index_2.htm	2007 article references "The Building Blocks Assessment of Early Mathematics, PreK-K (Sarama & Clements, in press)," but a search for that title yielded no results Building Blocks curriculum materials are available for purchase from McGraw Hill/SRA; teacher materials for pre-K include an "Assessment Guide" but no details about its contents are provided on the website
California Achievement Test (CAT)	*Note*: No longer in use from 2008: http://star.cde.ca.gov/star2011/aboutSTAR_programbg.aspx An older version (CAT/5) is available online for purchase by homeschoolers, but it appears not to be used anymore in any other context.	
Chemistry Concept Inventory (ChCI)	http://www.foundationcoalition.org/home/keycomponents/concept/chemistry.html Krause, S., Birk, J., Bauer, R., Jenkins, B., & Pavelich, M. (Oct. 2004). Development, testing, and application of a chemistry concept inventory. Paper presented at 34th ASEE/IEEE Frontiers in Education Conference. Savannah, GA. Paper #0-7803-7961-6/03	Unable to access online. Contact principal investigators.

Instrument Name	Information and Resources	Access Details
Children's self- and task perceptions during elementary school	Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self and task perceptions during elementary school. Child Development, 64, 830-847. http://www.jstor.org/stable/1131221 Electronic adaptations: Beal, C. R., Qu, L., & Lee, H. (2006). Classifying learner engagement through integration of multiple data sources. Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06). Menlo Park: AAAI Press. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.95.9897 Beal, C. R., Cohen, P. R., & Adams, N. (2010). Reading proficiency and mathematics problem solving by high school English Language Learners. Urban Education, 45, 58-74. http://animalwatch.arizona.edu/sites/default/files/BealAdamsCohen2010.pdf	See Constructs for full text of math subscale. Items listed in Eccles et al (1993), pg 834. http://www.jstor.org/stable/pdfplus/1131221.pdf
Children's Self-Efficacy Scale	http://www.ravansanji.ir/files/ravansanji-ir/21655425BanduraGuide2006.pdf Bandura, Albert. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.). Self-efficacy beliefs of adolescents, 5, pp. 307-337. Greenwich, CT: Information Age Publishing.	Full text available in linked document, page 326-327.
Computer Science Attitudes and Beliefs Assessment (CSABA)	Heersink, D. & Moskal, B. (2010) Measuring high school students' attitudes toward computing. Proceedings of the 41st ACM technical symposium on Computer science education, 446-250. doi: 10.1145/1734263.1734413 http://dl.acm.org/citation.cfm?id=1734413 Hoegh, A., & Moskal, B. M. (2009). Examining science and engineering students' attitudes toward computer science. Paper presented at the 39th ASEE/IEEE Frontiers in Education Conference, San Antonio, TX. http://www.fieconference.org/fie2009/papers/1035.pdf	Constructs and survey items available in Figure 1 of Hoegh and Moskal (2009).

Instrument Name	Information and Resources	Access Details
Core-Plus Mathematics Project assessments (CPMP)	http://www.wmich.edu/cpmp/index.html Coxford, A. F., & Hirsch, C. R. (1996). A common core of math for all. Educational Leadership, 53 (8), 22-25. Schoen, H. L., Ziebarth, S. W., Hirsch, C. R., & BrckaLorenz, A. (2010). A five-year study of the first edition of the Core-Plus Mathematics curriculum. Charlotte, NC: Information Age Publishing, Inc.	For further information about the Core-Plus Mathematics program and implementation support, contact cpmp@wmich.edu or 1-866-407-2767. For review copies of Course 1, 2, 3, and 4 materials, contact your McGraw-Hill representative.
Cornell Scientific Inquiry Series	Instrument homepage: http://ei.cornell.edu/index.html Trautmann, N.M., Carlsen, W.S., Krasny, M.E., & Cunningham, C.M. (2000). Integrated inquiry. The Science Teacher, 67(6), 52-55.	Available for purchase: http://www.nsta.org/store/search.aspx?action=quicksearch&text=Cornell%20Scientific%20Inquiry%20Series
Draw an Engineer Test (DAET)	Knight, Meredith and Christine Cunningham (2004). "Draw an Engineer Test (DAET): Development of a Tool to Investigate Students' Ideas about Engineers and Engineering." In Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition. http://www.mos.org/eie/pdf/research/DAET_ASEE_2004.pdf Weber, Nicole et al (2011). "The Development of a Systematic Coding System for Elementary Students' Drawings of Engineers." Journal of Pre-College Engineering Education Research, 1:1, 49-62. http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1030&context=jpeer Dyehouse et al. (2011). "Draw-an-Engineer (DAET) Coding System with Interview Triangulation." Proceedings of the Research in Engineering Education Symposium. http://web.ics.purdue.edu/~jstrobel/documents/strobel_dyehouse_rees2011_submission.pdf	Instrument questions available in development paper (Knight & Cunningham 2004). DAET coding system developed and validated in Weber et al. (2011) and Dyehouse et al. (2011).
Draw-A-Scientist Test (DAST)	Chambers, David Wade (1983). "Stereotypic Images of the Scientist: The Draw-A-Scientist Test." Science Education, 67:2, 255-265. doi:10.1002/sce.3730670213 http://onlinelibrary.wiley.com/doi/10.1002/sce.3730670213/pdf	

Instrument Name	Information and Resources	Access Details
Early Adolescent Temperament Questionnaire (EATQ-R)	Overview: http://www.bowdoin.edu/~sputnam/rothbart-temperament-questionnaires/instrument-descriptions/early-adolescent-temperament.html Ellis, Lesa K. and Mary K. Rothbart (2001). "Revision of the Early Adolescent Temperament Questionnaire." Poster presented at the 2001 Biennial Meeting of the Society for Research in Child Development, Minneapolis, Minnesota. http://www.bowdoin.edu/~sputnam/rothbart-temperament-questionnaires/pdf/lesa-ellis-srcd-poster-reprint.pdf	To request one or more of the instruments: http://www.bowdoin.edu/~sputnam/rothbart-temperament-questionnaires/request-forms/ For questions regarding the EATQ, contact Lesa Ellis at lellis@westminstercollege.edu
Early Algebra Student Assessment (EASA)	Brief project summary: http://www.terc.edu/work/954.html *Note*: Developed by Susan Jo Russell, Deborah Schifter, and Virginia Bastable at TERC, Inc. for another NSF-funded project (ESI-0550176, Foundations of Algebra in the Elementary and Middle Grades: Supporting Students to Make, Represent, and Justify General Claims about Operations).	Not currently available, though project summary indicates Foundations of Algebra "sourcebook" may be published in the future.
Early Expository Comprehension Assessment (EECA)	Hall, Kendra M. et al (2005). "The Development of the Early Expository Comprehension Assessment (EECA): A Look at Reliability." Communication Disorders Quarterly, 26:4, 195-206. doi: 10.1177/15257401050260040201 http://cdq.sagepub.com/content/26/4/195.abstract	Kendra M. Hall-Kenyon kendra_hall@byu.edu
Early Mathematics Classroom Observation (EMCO)	Klein, A., Starkey, P., Clements, D. H., Sarama, J., & Iyer, R. (2008). Effects of a pre-kindergarten mathematics intervention: A randomized experiment. Journal of Research on Educational Effectiveness, 1, 155-178.	Unable to obtain instrument.
EcoMUVE assessments	http://ecomuve.gse.harvard.edu/index.html	Currently in year 3 of development. Demo videos and materials are being finalized.
Emotion Regulation Checklist (ERC)	Summary profile: http://www.excellenceforchildandyouth.ca/support-tools/measure-profile?id=507	Dante Chicchetti Institute of Child Development University of Minnesota cicchett@umn.edu
Engineering Design Process Knowledge	Instrument: http://goo.gl/XmMzW Bailey, Reid and Zsuzsanna Szabo (2006). "Assessing Engineering Design Process Knowledge." International Journal of Engineering Education, 22:3, 508-518. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.157.9034	Available in Microsoft Word format at link.

Instrument Name	Information and Resources	Access Details
Engineering Identity Development Scale (EIDS)	Capobianco, Brenda M. et al (2009). "Generating measures of engineering identity development among young learners." Proceedings of the 39th ASEE/IEEE Frontiers in Education Conference. http://fie-conference.org/fie2009/papers/1049.pdf	Contact Information: Brenda M. Capobianco bcapo@purdue.edu
Engineering is Elementary Research Instruments (EiE)	Instrument: http://www.mos.org/eie/unitpdfs/researchInstruments.php EiE Educator Resources: https://www.mos.org/eie/unitpdfs/login.php Engineering is Elementary Research and Assessment: http://www.mos.org/eie/research_assessment.php	Available at no charge after registering on the website.
Entering Geometry Test (EG)	Overview: http://ucsmp.uchicago.edu/resources/van-hiele/ Usiskin, Zalman (1982). "Van Hiele Levels and Achievement in Secondary School Geometry." http://ucsmp.uchicago.edu/resources/van_hiele_levels.pdf http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED220288&ERICExtSearch_SearchType_0=no&accno=ED220288	No fee for the use of these tests, although permission is needed to duplicate them. Contact Zalman Usiskin (z-usiskin@uchicago.edu) with: (a) a description of your study in which the instruments are to be used; (b) an approximate number of copies to be made of the instruments used in your study; (c) assurance that you will write on each copy of the test: "Copyright ©1980 by the University of Chicago. Reprinted with permission of the University of Chicago"; (d) assurance that you will send us copies of any written reports involving results using these instruments.
Expressive Vocabulary Test-2 (EVT-2)	http://www.pearsonassessments.com/evttwo.aspx Williams, K.T. (2007). Expressive Vocabulary Test-Second Edition, Manual. AGS Publishing, Pearson Assessment. Technical Specifications: http://psychcorp.pearsonassessments.com/hai/images/Products/EVT-II/evt2.pdf	

Instrument Name	Information and Resources	Access Details
Force Concept Inventory (FCI)	http://modeling.asu.edu/r&e/fci.pdf	Free. Can be downloaded. School information needed to receive the password.
Framework for Teaching Evaluation Instrument	www.teachscape.com/binaries/content/assets/teachscape-marketing-website/products/classroom-walkthrough/danielson_newframeworkforteachingevaluationinstrument.pdf For more information: http://www.danielsongroup.org/Default.aspx	Available via link. Instrument highlights: http://www.bcsberlin.k12.nj.us/225810325214955677/lib/225810325214955677/Why_FFT.pdf
Full Option Science System assessments (FOSS)	http://fossweb.schoolspecialty.com/foss-modules Foss: Full Option Science System. (2003). Regents of the University of California, Lawrence Hall of Science. New Hampshire: Delta Education.	Available for purchase: http://www.delta-education.com/science/foss3/fossall.shtml
Gates-McGinitie Reading Test (GMRT)	http://www.riversidepublishing.com/products/gmrt/index.html	
Goal Orientation and Learning Strategies Survey (GOALS-S)	Dowson, Martin and Dennis M. McInerney (2004). "The Development and Validation of the Goal Orientation and Learning Strategies Survey (Goals-S)." Educational and Psychological Measurement, 64:2, 290-310. http://epm.sagepub.com/content/64/2/290 http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED408315	Unavailable online.
Inquiring into Science Instruction Observation Protocol (ISIOP)	http://isiop.edc.org/	To access videos for training, you need to request an account.
Instructional Material Motivational Survey (IMMS)	Keller, J.M. (1983). "Motivational design of instruction. In C.M. Reigeluth (Ed.). Instructional design theories and models: An overview of their current status." Hillsdale, NJ: Erlbaum. http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8535.2005.00582.x/full	Instrument available (as Word document): http://goo.gl/vwkqh
Instructional Quality Assessment (IQA)	Junker, Brian et al (2006). "Overview of the Instructional Quality Assessment." CSE Technical Report 671. http://www.cse.ucla.edu/products/reports/r671.pdf	The IQA rubrics are free to the public and can be accessed through the Center for Standards, Evaluation, & Student Testing website.

Instrument Name	Information and Resources	Access Details
Interactive Mathematics Program assessments (IMP)	http://www.mathimp.org/ Interactive Mathematics Program, IMP Evaluation Update, No. 2, Fall 1996, Emeryville, CA, 1996. Schoen, H. L. (1993). Report to the National Science Foundation on the impact of The Interactive Mathematics Project. Madison, WI: Wisconsin Center for Education Research.	http://www.keycurriculum.com/products/interactive-mathematics-program For more information about the program and its implementation, call 1-888-MATH-IMP (1-888-628-4467).
Intrinsic Motivation Inventory (IMI)	http://selfdeterminationtheory.org/questionnaires/10-questionnaires/50 Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. Journal of Personality and Social Psychology, 43, 450-461. http://selfdeterminationtheory.org/SDT/documents/1982_Ryan_ControlandInfo_JPSP.pdf	Many versions available. Scales available for academic research after registering and logging into the website: http://selfdeterminationtheory.org/questionnaires
Junior Metacognitive Awareness Inventory (Jr. MAI)	Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. Contemporary Educational Psychology, 27, 51-79. http://www.sciencedirect.com/science/article/pii/S0361476X01910914	Full text in appendix to Sperling et al (2002), available by subscription or individual purchase.
KeyMath-3 Diagnostic Assessment (KeyMath-3 DA)	http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Mode=summary&Pid=PAaKeymath3 *Note:* Available for purchase from the publisher in kits of various configurations. KeyMath–3 DA Publication Summary Form (contains reliability and validity details): http://www.pearsonassessments.com/hai/images/pa/products/keymath3_da/km3-da-pub-summary.pdf	
Kuder Career Search (KCS)	http://www.kuder.com/ Technical manual: http://www.visions-unltd.com/PublicWeb/kcs_manual.aspx	Available for purchase.

Instrument Name	Information and Resources	Access Details
Language Assessment Scales Links (LAS Links)	Overview: http://www.ctb.com/ctb.com/control/productFamilyViewAction?p=products&productFamilyId=454 CTB/McGraw-Hill (2010). "English Language Development: Progress Monitoring Assessment using LAS Links." http://www.ctb.com/ctb.com/control/researchArticleMainAction?p=ctbResearch&articleId=9864	Must be purchased from publisher: http://www.ctb.com/ctb.com/control/childNodesViewAction?categoryId=106&rootCatFlag=T&m=menu&p=store Pricing depends on: number of EL students; number of teachers; number of schools; number of assessments per year; scoring; number of staff training days; and assessment configuration. Call CTB/McGraw-Hill (800.538.9547) for a price quote.
Mathematical Argument as Joint Activity in the Classroom (MAJAC)	Shechtman, N., Knudsen, J., Michalchik, V., Stevens, H., & Kim, H. (2011). The bridging teacher professional development project: Supporting mathematical argumentation in distressed urban middle school contexts. Manuscript submitted for review.	No information found online. Developed by principal investigators for prior project.
Mathematics Classroom Observation Protocol	http://mathconnect.hs.iastate.edu/documents/Aguirre.pdf Adapted from National Center for Research in Mathematics Education. (1992). Wisconsin Center for Educational Research. Madison, WI: University of Wisconsin-Madison.	*Note*: Scale for Academic Language Support for ELLs developed by E. Rubenstein-Avila (2006), The University of Arizona.
Measures of Academic Progress for Science (MAP for Science)	http://www.nwea.org/products-services/computer-based-adaptive-assessments/map "The Reliability of NWEA Results." http://www.nwea.org/support/article/533	Must order from institution. Information request form: http://www.nwea.org/about-nwea/contact-us
Mechanics Baseline Test (MBT)	Hestenes, D., & Wells, M. (1992). A mechanics baseline test. The Physics Teacher, 30, 159-165. http://modeling.la.asu.edu/R%26E/MechBaseline.pdf	Available to educators: http://modeling.asu.edu/R&E/Research.html Available in multiple languages.

Instrument Name	Information and Resources	Access Details
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.
Modified Attitudes towards Science Inventory (mATSI)	Overview: http://cyfernetsearch.org/sites/default/files/PsychometricsFiles/Science-Interest%20(5th%20Grade)_0.pdf Additional info here: http://www.pearweb.org/atis/tools/7	Scoring guide also available.
Motivated Strategies for Learning Questionnaire (MSLQ)	http://www.indiana.edu/~p540alex/MSLQ.pdf Pintrich, Paul R. (1991). "A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)." Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning. http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED338122&ERICExtSearch_SearchType_0=no&accno=ED338122	Paper on reliability and predictive validity requires subscription or purchase. Technical manual is temporarily unavailable ("If you would like to request a PDF to be returned online, please email ERICRequests@ed.gov with the record number [MSLQ manual is ED338122, ed. note].")
NASA Task Load Index (NASA-TLX)	http://humansystems.arc.nasa.gov/groups/TLX/downloads/TLXScale.pdf Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press. Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 Years Later. Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, 904-908. Santa Monica: HFES. Chapter on the development of the index: http://humansystems.arc.nasa.gov/groups/TLX/downloads/NASA-TLXChapter.pdf	Article on how this tool has been applied since its inception: http://humansystems.arc.nasa.gov/groups/TLX/downloads/HFES_2006_Paper.pdf
National Assessment of Educational Progress (NAEP)	http://nces.ed.gov/nationsreportcard/itmrlsx/default.aspx	

Instrument Name	Information and Resources	Access Details
Oral Proficiency Interview (OPI)	Overview: http://www.actfl.org/i4a/pages/index.cfm?pageid=3348 ACTFL Proficiency Guidelines-Speaking (Revised 1999): http://www.actfl.org/files/public/Guidelines.pdf	For more information on OPI Testing or to schedule a personal Oral Proficiency Interview, contact: LTI E-mail: testing@languagetesting.com http://www.languagetesting.com
Patterns of Adaptive Learning Scales (PALS)	Project website: http://www.umich.edu/~pals/ Midgley, Carol et al (2000). "Manual for the Patterns of Adaptive Learning Scales (PALS)." Ann Arbor, MI: University of Michigan. http://www.umich.edu/~pals/PALS%202000_V13Word97.pdf	Full survey text provided in technical manual.
Peabody Picture Vocabulary Test-Fourth Edition (PPVT-4)	http://psychcorp.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=PAa30700&Mode=summary Dunn LM, Dunn LM. Peabody Picture Vocabulary Test--Fourth Edition. In. Bloomington, MN: Pearson Assessments; 2007.	
Program for International Student Assessment (PISA)	Overview: 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/assessingscientificreadinganddmathematicalliteracy.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	Versions available at links. All accompanying resources including technical assessment information available: http://www.oecd.org/pisa/pisaproducts/#d.en.192289
Renfrew Bus Story, North American Edition (RBS-NA)	http://www.busstory.us/ Glasgow, C., & Cowley, J. (1994). Renfrew Bus Story test - North American Edition. Centreville, DE: Centreville School.	Must be purchased from publisher. http://busstory.us/orderbusstory.html

Instrument Name	Information and Resources	Access Details
Research-Based Early Maths Assessment (REMA)	Clements, Douglas H. et al (2008). ""Development of a Measure of Early Mathematics Achievement Using the Rasch Model: The Research-Based Early Maths Assessment."" Educational Psychology, 28:4, 457-482. DOI:10.1080/01443410701777272 http://www.tandfonline.com/doi/abs/10.1080/01443410701777272	Full text of paper requires subscription or costs $36 to purchase individually; unable to determine if instrument text is included. Contact information: Douglas Clements at clements@buffalo.edu
Scholastic Aptitude Test (SAT)	http://sat.collegeboard.org/home	
Science Attitudes, Skills, & Knowledge Survey (SASKS)	Form 1: http://www.public.asu.edu/~anton1/AssessArticles/Assessments/Science%20Assessments/SASKS1.pdf Form 2: http://www.public.asu.edu/~anton1/AssessArticles/Assessments/Science%20Assessments/SASKS2.pdf Form 3: http://www.public.asu.edu/~anton1/AssessArticles/Assessments/Science%20Assessments/SASKS3.pdf Teacher version (S-TASKS): http://cosmos.bgsu.edu/communities/research_community/MeasurementInst/pdfs/S-TASKS.pdf	
Science Education for Public Understanding Program assessments (SEPUP)	http://sepuplhs.org/assess.html Assessment system development: http://sepuplhs.org/pdfs/SiegelNagleBarterAERA.pdf SEPUP related research: http://sepuplhs.org/pdfs/sepup_bibliography_2011.pdf	Available from publisher: http://lab-aids.com/
Science Learning Assessment (SLA)	Samarapungavan, A., Mantzicopoulos, P., Patrick, H., and French, B. (2009) The development and validation of the Science Learning Assessment (SLA): A measure of kindergarten science learning. Journal of Advanced Academics; 20 (3): 502-35. http://www.eric.ed.gov/PDFS/EJ860959.pdf	Descriptions of items available in linked paper.

Instrument Name	Information and Resources	Access Details
Science notebook as assessment tool rubric	http://www.stanford.edu/dept/SUSE/SEAL/Presentation/Presentation%20PDF/Looking%20into%20Aera%202001%20pdf.pdf Ruiz-Primo, Maria A. et al (2004). "Evaluating students' science notebooks as an assessment tool." International Journal of Science Education, 26(12), 1477-1506. DOI: 10.1080/0950069042000177299 http://www.tandfonline.com/doi/abs/10.1080/0950069042000177299 Ruiz-Primo, Maria Araceli, Min Li, and Richard J. Shavelson (2001). "Looking into students' science notebooks: What do teachers do with them?" UCLA Center for Research on Evaluation, Standards, and Student Testing. Technical report. http://www.cse.ucla.edu/products/reports/TR562.pdf	Most current development paper (Ruiz-Primo et al, 2004) requires subscription or purchase. Ruiz-Primo et al (2001) describes in detail instrument development and coding scheme, but may not be the most current version.
Scientific Attitude Inventory: A revision(SAI II)	Instrument: http://ret.fsu.edu/Research_Tools.htm For additional information: http://www.pearweb.org/atis/tools/12	Available for download.
Scientific Inquiry and Engineering Design Scoring Guides	http://www.ode.state.or.us/search/page/?id=1414 Student work and scoring examples: http://www.ode.state.or.us/search/page/?=519	Supporting documents: http://www.ode.state.or.us/search/page/?id=32
Self-Description Questionnaire II (SDQII)	http://www.self.ox.ac.uk/SDQ2.htm Technical Manual: http://www.self.ox.ac.uk/SDQ2m.htm	Full text available at link provided.
Self-Efficacy in Technology and Science (SETS)	Ketelhut, Diane J. (2011). "Assessing Gaming, Computer, and Scientific Self-Efficacy in a Virtual Environment." In L. Annetta & S. Bronack (Eds.), Serious Educational Game Assessment: Practical Methods and Models for Educational Games, Simulations, and Virtual Worlds (pp 1-18). Rotterdam: Sense Publishing. https://www.sensepublishers.com/files/9789460913297PR.pdf	Unable to find online. Author contact details: Diane Jass Ketelhut (301) 405 3324 djk@umd.edu
Social Skills Improvement System-Rating Scales (SSIS-RS)	https://psychcorp.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=PAa3400&Mode=summary Gresham FM, Elliott SN. (2008). Social Skills Improvement System-Rating Scales. Minneapolis, MN: Pearson Assessments.	

Instrument Name	Information and Resources	Access Details
Sources of Science Self-Efficacy Scale (SSSE)	Britner, S. L. and Pajares, F. (2006), Sources of science self-efficacy beliefs of middle school students. J. Res. Sci. Teach., 43: 485–499. doi: 10.1002/tea.20131 http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ760151 Adapted for middle school from this instrument for undergraduates: Lent, R. W., Lopez F. G., & Bieschke, K. J. (1991). Mathematics self-efficacy: Sources and relation to science-based career choice. Journal of Counseling Psychology, 38, 424-430. http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ438838	Unable to locate online. Technical papers by Britner & Pajares (2006) and Lent et al (1991) available for subscription or purchase; unknown whether full text of instrument is included in either.
Stanford Achievement Test (Stanford 10)	Product overview: http://education.pearsonassessments.com/haiweb/cultures/en-us/productdetail.htm?pid=SAT10C Stanford Achievement Test overview: https://docs.alsde.edu/documents/91/Stanford%20Achievement%20Test.pdf	
Strong Interest Inventory	Overview: https://www.cpp.com/products/strong/index.aspx Donnay, David A. C. et al (2004). "Technical Brief for the Newly Revised Strong Interest Inventory Assessment: Content, Reliability, and Validity." CPP Research Department. https://www.cpp.com/Pdfs/StrongTechnicalBrief.pdf	Available for purchase from publisher (many options): https://www.cpp.com/en/searchresults.aspx?prodName=&author=&code=&prodFam=2
Student Collective Engagement (SCE)	Reeve, Johnmarshall et al (2004). "Enhancing High School Students' Engagement by Increasing Teachers' Autonomy Support." Motivation and Emotion, 28(2), 147-169. http://johnmarshallreeve.org/yahoo_site_admin1/assets/docs/Reeve_Jang_Carrell_Barch_Jeon2004.4731952.pdf	Observer rating sheet provided in Reeve et al (2004), pg. 156; see link.

Instrument Name	Information and Resources	Access Details
Student Perceptions of Classroom Climate (SPOCC)	Yore, L.D., Shymansky, J.A., Henriques, L., Hand, B.M., Dunkhase, J.A., & Lewis, J.O. (1998, January). Students' perceptions of science teaching and attitudes toward science learning and teachers' self-report of using children's ideas, applications of science, and use of print resources as indicators of classroom teaching. Paper presented at the International Conference of the Association for the Education of Teachers in Science, Minneapolis, MN. http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED442653	Full text of development paper (Yore et al, 1998) may contains more details, but is temporarily unavailable. http://www.eric.ed.gov/PDFS/ED442653.pdf
Student Understanding of Science and Scientific Inquiry (SUSSI)	Liang, L. L., Chen, S., Chen, X., Kaya, O. N., Adams, A. D., Macklin, M., & Ebenezer, J. (2006). Student Understanding of Science and Scientific Inquiry (SUSSI): Revision and further validation of an assessment instrument. Paper presented at the annual meeting of the National Association of Research in Science Teaching, San Francisco, CA. http://www.springerlink.com/content/p6065qj00q433842/ http://www.gb.nrao.edu/~sheather/For_Sarah/lit%20on%20nature%20of%20science/SUSSI.pdf	Full text of instrument is in Appendix A of Liang et al (2006), pg 28-29. Also includes scoring guides and illustrative examples of student responses.
Students' Use of Evidence in Written Scientific Explanations	Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. Cognition and Instruction, 23(1), 23-55. http://www.jstor.org/stable/3233896 http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ724914	Example rubrics in development paper, available at no charge after brief registration: http://academia.edu/797319/The_quality_of_students_use_of_evidence_in_written_scientific_explanations
Taking a Good Look at Instructional Technology (TAGLIT)	Overview: http://www.testkids.com/taglit/about.htm Student assessments: http://www.testkids.com/taglit/student.htm	Versions available for teachers, administrators, and students.
Technology Enhanced Learning in Science assessment activities (TELS)	Overview: http://telscenter.org/projects/tels/assessments Liu, O.L., Lee, H.S., Hofstetter, C., & Linn, M.C. (2008). Assessing knowledge integration in science: Construct, measures and evidence. Educational Assessment, 13(1), 33–55. http://www.tandfonline.com/doi/abs/10.1080/10627190801968224	Development paper requires paid access. Assessment items incorporated into curriculum available as WISE projects: http://wise.berkeley.edu/webapp/index.html

Instrument Name	Information and Resources	Access Details
TELS Knowledge Integration Scoring Rubric (TELS)	http://telscenter.org/projects/tels/assessments/rubric Background on Knowledge Integration (KI): http://telscenter.org/projects/tels/knowledge-integration	Sample rubric and example responses available at link.
Terra Nova Algebra Assessment	http://www.ctb.com/ctb.com/control/productFamilyViewAction?productFamilyId=449&p=products	
Test of Early Mathematics Ability-Third Edition (TEMA-3)	*Note*: Available for purchase from various sources. Ginsburg, H. P. & Baroody, A.J. (2003). Test of Early Mathematics Ability, 3rd edition. Austin, TX: PRO-ED. http://aei.sagepub.com/content/30/4/57.extract	
Test of Science-Related Attitudes (TOSRA)	Fraser, Barry J. (1981). "TOSRA: Test of Science-Related Attitudes Handbook." Hawthorn, Victoria: The Australian Council for Educational Research Limited. http://www.ecu.edu/ncspacegrant/docs/RESTEPdocs/TOSRA_BJF_paper.pdf	Full text of instrument provided in handbook (Fraser, 1981) or separately as Word document (http://ret.fsu.edu/Files/Tools/TOSRA_2.doc).
Trends in International Mathematics and Science Study (TIMSS)	Overview: http://nces.ed.gov/timss/ Public released item bank: http://nces.ed.gov/timss/educators.asp Williams, T., Ferraro, D., Roey, S., Brenwald, S., Kastberg, D., Jocelyn, L., Smith, C., and Stearns, P. (2009). TIMSS 2007 U.S. Technical Report and User Guide (NCES 2009–012). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.	
Upper-Elementary Mathematics Assessment Modules	Research overview: http://nsf.gov/awardsearch/showAward.do?AwardNumber=0831450 MSPnet profile: http://lmtets.mspnet.org/index.cfm/profile ETS Research Portfolio 2011 (pg 32): www.ets.org/Media/Research/pdf/RDAgenda.pdf	Intended for dissemination but not yet available to the public.
Views About Science Survey (VASS)	Halloun, Ibrahim A. (2001). "Student Views about Science: A Comparative Survey." Monograph published by the Educational Research Center at Lebanese University, Beirut, Lebanon. http://modeling.asu.edu/R&E/IHalloun/VASS-2001Monograph.pdf	Available at no charge.

Instrument Name	Information and Resources	Access Details
Views of Nature of Science Questionnaire (VNOS)	Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learner's conceptions of nature of science. Journal of Research in Science Teaching, 39(6), 497-521. http://www.gb.nrao.edu/~sheather/For_Sarah/lit%20on%20nature%20of%20science/views%20of%20nature%20of%20science%20questionnaire.pdf	VNOS-A: http://ret.fsu.edu/Files/Tools/VNOS(A)[1].pdf VNOS -B: http://www.csss-science.org/downloads/VNOS_B.pdf VNOS-C: http://ret.fsu.edu/Files/Tools/VNOS(C)[1].pdf VNOS-D: http://ret.fsu.edu/Files/Tools/VNOS(D)[1].pdf
Views of Scientific Inquiry (VOSI)	Schwartz, Renee' S. et al (2008). "An Instrument to Assess Views of Scientific Inquiry: The VOSI Questionnaire." Paper presented at the annual meeting of the National Association for Research in Science Teaching, March 30-April 2, 2008. Baltimore, MD http://homepages.wmich.edu/~rschwart/docs/VOSInarst08.pdf	Specific VOSI forms for use in research, evaluation, or instruction can be obtained by contacting the authors: Renee' S. Schwartz r.schwartz@wmich.edu Norman G. Lederman ledermann@iit.edu Judith S. Lederman ledermanj@iit.edu
Wallach-Kogan Creativity Test (WKCT)	Wallach, M. A., & Kogan, N. (1965). Modes of thinking in young children: A study of the creativity-intelligence distinction. New York: Holt, Rinehart, and Winston.	

Appendix I: Alphabetical Listing (by state) of State Academic Achievement Tests with Access Details

State	Assessment Name	Access details and Resources
AL	Alabama Reading and Mathematics Test (ARMT)	http://www.alsde.edu/html/sections/doc_download.asp?section=91&id=8358&sort=10 Item Specifications: http://alex.state.al.us/ccrs/node/76
AK	Alaska Standards Based Assessment (SBA)	Overview: http://www.eed.state.ak.us/tls/assessment/sba.html Practice Test materials: http://www.eed.state.ak.us/tls/assessment/SBA_PracticeTests.html Technical Reports: http://www.eed.state.ak.us/tls/assessment/techreports.html
CA	California English Language Development Test Reading Scale (CELDT)	CELDT Overview: http://www.cde.ca.gov/ta/tg/el/cefceldt.asp Released Questions: http://www.cde.ca.gov/ta/tg/el/documents/celdtrtqs9-2012.pdf CELDT Technical Documentation: http://www.cde.ca.gov/ta/tg/el/techreport.asp Additional Information: http://www.cde.ca.gov/ta/tg/el/resources.asp
CA	California Standards Test (CST)	Overview: http://www.cde.ca.gov/ta/tg/sr/resources.asp Released test questions: http://www.cde.ca.gov/ta/tg/sr/css05rtq.asp and http://starsamplequestions.org/starRTQ/search.jsp Technical reports: http://www.cde.ca.gov/ta/tg/sr/technicalrpts.asp
CO	Colorado Student Assessment Program (CSAP)	*Note*: CSAP has been replaced with Transitional Colorado Assessment Program (TCAP) in 2012 Overview: http://www.cde.state.co.us/assessment/CoAssess.asp Released items: http://www.cde.state.co.us/assessment/CoAssess-Released.asp Additional Resources: http://www.cde.state.co.us/assessment/CoAssess-AdditionalResources.asp
IA	Iowa Test of Basic Skills (ITBS)	http://www.riversidepublishing.com/products/itbs/ Can be ordered from Riverside Publishing
IL	Illinois Standards Achievement Test (ISAT)	Overview: http://www.isbe.state.il.us/assessment/isat.htm Sample questions: http://metacat2.com/iltestlinks.html Technical Manual: http://www.isbe.state.il.us/assessment/isat.htm Interpretive Guide: http://www.isbe.net/assessment/pdfs/ISAT_Interpr_Guide_2011.pdf

IN	Indiana Statewide Testing for Educational Progress (ISTEP)	Overview: http://www.doe.in.gov/achievement/assessment/istep-grades-3-8 Select items released online: http://www.doe.in.gov/achievement/assessment/istep-released-items-and-scoring-notes Guide to Test Interpretation: http://www.doe.in.gov/sites/default/files/assessment/2012-gti.pdf
KY	Kentucky Core Content Test (KCCT)	*Note*: KCCT is no longer in use from 2011-2012. http://www.education.ky.gov/KDE/Administrative+Resources/Testing+and+Reporting+/District+Support/Archives/Kentucky+Core+Content+Test+2001-2011.htm K-PREP Blueprint: http://www.education.ky.gov/NR/rdonlyres/EC1509B4-C7EF-4A51-A79A-46676BE77BC2/0/KPREPBlueprints.pdf
LA	Integrated Louisiana Educational Assessment Program (*i*LEAP)	Overview: http://www.louisianaschools.net/topics/ileap.html Sample items: http://www.louisianaschools.net/topics/sample_test_items_4_8.html Sample practice tests: http://www.louisianaschools.net/topics/practice_test_questions.html Additional resources: http://www.louisianaschools.net/testing/help/
MA	Massachusetts Comprehensive Assessment System (MCAS)	Overview: http://www.doe.mass.edu/mcas/ Test items: http://www.doe.mass.edu/mcas/testitems.html Technical reports: http://www.doe.mass.edu/mcas/tech/
MD	Maryland High School Assessments (HSA)	HSA homepage: http://hsaexam.org/ Sample items: http://mdk12.org/assessments/high_school/index.html Practice Tests: http://hsaexam.org/support/practice.html Test Specifications and Blueprints: http://www.marylandpublicschools.org/MSDE/testing/hsa/ http://marylandpublicschools.org/NR/rdonlyres/AF9068C5-41EC-447A-BA28-95DFB2F05678/9873/2005_TechReport_SECTION_1.pdf 2005 Technical Report: http://marylandpublicschools.org/MSDE/divisions/planningresultstest/2005+HSA+Technical+Report.htm
MD	Maryland School Assessment (MSA)	Overview: http://www.marylandpublicschools.org/MSDE/testing/msa/ Sample items: http://www.mdk12.org/assessments/k_8/index_c.html Technical Report: http://www.marylandpublicschools.org/MSDE/divisions/planningresultstest/2007+MSA+Reading+Technical+Report

MI	Michigan Educational Assessment Program (MEAP)	Overview: http://www.michigan.gov/mde/0,1607,7-140-22709_31168---,00.html Released items: http://www.michigan.gov/mde/0,4615,7-140-22709_31168-281205--,00.html http://www.michigan.gov/mde/0,4615,7-140-22709_31168-281206--,00.html http://www.michigan.gov/mde/0,4615,7-140-22709_31168-281209--,00.html Technical reports: http://www.michigan.gov/mde/0,1607,7-140-22709_35150-172665--,00.html Curriculum frameworks: http://www.michigan.gov/documents/MichiganCurriculumFramework_8172_7.pdf
MS	Mississippi Curriculum Test (MCT)	Overview: http://www.mde.k12.ms.us/student-assessment/student-assessment-mct2 Practice Tests: https://districtaccess.mde.k12.ms.us/studentassessment/Public%20Access/Forms/AllItems.aspx?RootFolder=%252Fstudentassessment%252FPublic%20Access%252FStatewide_Assessment_Programs%252FPractice_Tests Test blueprint for math: http://www.mde.k12.ms.us/docs/student-assessment/math_test_blueprint.pdf?sfvrsn=2
MS	Mississippi Science Test (MST2)	Overview: http://www.mde.k12.ms.us/student-assessment/student-assessment-mst2 Practice Tests: https://districtaccess.mde.k12.ms.us/studentassessment/Public%20Access/Forms/AllItems.aspx?RootFolder=%252Fstudentassessment%252FPublic%20Access%252FStatewide_Assessment_Programs%252FPractice_Tests
MO	Missouri Assessment Program (MAP)	Overview: http://dese.mo.gov/divimprove/assess/mapa.html 2008 MAP Technical Report: http://dese.mo.gov/divimprove/assess/tech/documents/2008MAPTechnicalReport.pdf Order from institution (DESE at 573-751-3545) or publisher (CTB/McGraw-Hill at 800-544-9868)
NE	Omaha Public Schools Criterion-Referenced Tests (CRT)	*Note*: no longer in use from 2010: http://drs.education.ne.gov/ http://www.ops.org/district/CENTRALOFFICES/GeneralAdministrativeServices/Research/AchievementReports/StatewideAchievementTests/CriterionReferencedTests/tabid/1861/Default.aspx
NJ	Grade Eight Proficiency Assessment (GEPA)	*Note*: GEPA is no longer in use from 2008. Overview: http://www.state.nj.us/education/assessment/ms/gepa/ Sample test and other resources: http://www.state.nj.us/education/assessment/ms/sample/
NJ	New Jersey Assessment of Skills and Knowledge (NJ ASK)	Overview: http://www.state.nj.us/education/assessment/ms/5-8/ Sample tests: http://www.state.nj.us/education/assessment/es/sample/ For additional information: http://www.state.nj.us/education/assessment/
NM	New Mexico Standards Based Assessment (NMSBA)	Overview: http://www.ped.state.nm.us/AssessmentAccountability/AssessmentEvaluation/SBA/index.html Released items: http://www.ped.state.nm.us/AssessmentAccountability/AssessmentEvaluation/dl09/releasedItems/index.html

NY	New York State Regents exam (REGENTS)	Overview: http://www.p12.nysed.gov/assessment/hsgen/ Past Exams: http://www.nysedregents.org/ For more details: http://schools.nyc.gov/daa/test_info/ Technical Reports: http://www.p12.nysed.gov/assessment/reports/
NY	New York State Testing Program (NYSTP)	Overview: http://www.p12.nysed.gov/assessment/ei/eigen.html Past Exams: http://www.nysedregents.org/ For more details: http://schools.nyc.gov/daa/test_info/ Learning Standards for NY: http://www.p12.nysed.gov/ciai/standards.html Technical Reports: http://www.p12.nysed.gov/assessment/reports/
NC	North Carolina End-of-Grade Tests (EOG)	EOG Overview: http://www.ncpublicschools.org/accountability/testing/eog/ Math: http://www.ncpublicschools.org/accountability/testing/eog/math/ Science: http://www.ncpublicschools.org/accountability/testing/eog/science/ Reading: http://www.ncpublicschools.org/accountability/testing/eog/reading/ Test Development process: http://www.ncpublicschools.org/accountability/testing/shared/testdevprocess
OH	Ohio Graduation Tests (OGT)	*Note*: OGTs were eliminated in 2009. OGT homepage: http://www.ode.state.oh.us/GD/Templates/Pages/ODE/ODEPrimary.aspx?page=2&TopicRelationID=216 Released materials: http://education.ohio.gov/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&TopicRelationID=1070&ContentID=7835&Content=128490 Practice Tests: http://education.ohio.gov/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&TopicRelationID=240&ContentID=4760&Content=126216 Blueprints: http://education.ohio.gov/GD/Templates/Pages/ODE/ODEDetail.aspx?page=3&TopicRelationID=216&ContentID=10825&Content=126762

OR	Oregon Assessment of Knowledge and Skills (OAKS)	OAKS overview: http://www.oaks.k12.or.us/ Sample Tests: http://www.ode.state.or.us/search/page/?=1222 Test Specifications and Blueprints: http://www.ode.state.or.us/search/page/?id=496 Online user guides: http://www.ode.state.or.us/search/page/?=391 Technical Report: http://www.ode.state.or.us/search/page/?=1305
PA	Pennsylvania System of School Assessment (PSSA)	Overview: http://www.portal.state.pa.us/portal/server.pt/community/pennsylvania_system_of_school_assessment_(pssa)/8757 Test Samples and Resources: http://www.portal.state.pa.us/portal/server.pt/community/pennsylvania_system_of_school_assessment_(pssa)/8757/resource_materials/507610
TX	Texas Assessment of Knowledge and Skills (TAKS)	Overview: http://www.tea.state.tx.us/student.assessment/taks/ Released items: http://www.tea.state.tx.us/student.assessment/taks/items/ Released tests: http://www.tea.state.tx.us/student.assessment/taks/released-tests/ Test Blueprints: http://www.tea.state.tx.us/student.assessment/taks/blueprints/
UT	Utah Performance Assessment System for Students (U-PASS)	Overview: http://www.schools.utah.gov/assessment/default.aspx Test items: Utah Test Item Pool Service (UTIPS) http://www.schools.utah.gov/assessment/UTIPS.aspx
VA	Virginia Standards of Learning Assessments (SOL)	Practice items: http://www.doe.virginia.gov/testing/sol/practice_items/index.shtml Math Blueprints and other resources: http://www.doe.virginia.gov/testing/sol/standards_docs/mathematics/index.shtml Science blueprints and other resources: http://www.doe.virginia.gov/testing/sol/standards_docs/science/index.shtml
WA	Washington Measurements of Student Progress (MSP)	Overview: http://www.k12.wa.us/assessment/StateTesting/MSP.aspx Practice items: http://www.k12.wa.us/assessment/StateTesting/TestQuestions/Testquestions.aspx 2011 Technical Report: http://www.k12.wa.us/assessment/pubdocs/WCAP2011SpringAdministrationTechnicalReport.pdf