Title: A Common Measurement System for K-12 STEM Education: adopting an educational evaluation methodology that elevates theoretical foundations and systems thinking.

Keywords: common measurement system; STEM education; K-12 education; formative evaluation; college and career readiness

Abstract: Science, Technology, Engineering, and Mathematics (STEM) education is important on a national, regional, local, and individual level. However, there are many diverse problems facing STEM education in the US, one of the most critical is the limitation of current measurement tools and evaluation methodologies. The development of a common measurement system is an important step in addressing these problems. This paper describes the conceptualization stage of the development of a common measurement system. The resulting STEM Common Measurement System includes constructs that span from student learning to teacher practice to professional development to school-level variables. The authors detail the constructs and measurement tools associated with each construct. The interconnections within the STEM Common Measurement System are also discussed.
A Common Measurement System for K-12 STEM Education: adopting an educational evaluation methodology that elevates theoretical foundations and systems thinking.

Authors: Saxton, Emily a*, Burns, Robin b, Holveck, Susan c, Kelley, Sybil d, Prince, Daniel e, Rigelman, Nicole d, & Skinner, Ellen. A. f

a Center for Science Education, Portland State University, 1136 SW Montgomery Street, Portland, Oregon 97201
b Intel Corporation
c Beaverton School District, 16550 SW Merlo Road, Beaverton, Oregon 97006
d Graduate School of Education, Portland State University, 1136 SW Montgomery Street, Portland, Oregon 97201
e Outdoor School Program, Multnomah Education Service District
f Department of Psychology, Portland State University, 1136 SW Montgomery Street, Portland, Oregon 97201

*Corresponding Author.

Email addresses: esaxton@pdx.edu (E. Saxton), robin.burns@intel.com (R. Burns), Susan_Holveck@beaverton.k12.or.us (S. Holveck), sybilkel@pdx.edu (S. Kelley), dprince@mesd.k12.or.us (D. Prince), rigelman@pdx.edu (N. Rigelman), skinnere@pdx.edu (E.A. Skinner)

Vitae

Emily Saxton, MST, is the Portland Metro STEM Partnership's Director of Research and Assessment and a research associate with the Center for Science Education at Portland State University. Her Masters degree in Science Teaching is from Portland State University and her B.S. degree in Environmental Science is from the University of Florida. Her research interests include assessment development and validation, formative assessment, and evaluation of professional development. Ms. Saxton served as the committee chair of the Portland Metro STEM Partnership's common measures committee that is described in this paper.

Robin Burns, B.S. is a Senior Technical Program Manager at Intel Corporation. She has worked in industry for over 20 years in roles ranging from Manufacturing Engineering to Systems Integration / Engineering, along with various project and program management roles in the high tech sector. She uses a systems approach to problem solving in her day to day job. Her B.S. degree is in Industrial Engineering (Manufacturing Option) from Oregon State University. Her interests in STEM education have led to participation on the Portland Metro STEM common measures committee and many volunteer activities supporting STEM through Intel Involved programs.

Susan Holveck, D.Ed, is the Beaverton School District’s Science Specialist for Secondary Schools. After working in a lab, she turned to teaching and has recently earned a D.Ed in Educational Management, Policy, and Leadership from the University of Oregon. Susan was honored to receive the 2007 OSTA Middle School Science Teacher of the Year Award and the Lower Columbia River Estuary Partnership Steward of the Year in 2009. She also has a MS in Genetics and Cellular Biology from Washington State
University and a BS in Genetics from Ohio State University. Her research interests include scientific inquiry and conceptual understanding.

**Sybil S. Kelley**, PhD, is Assistant Professor of Science Education and Sustainable Systems at Portland State University in the Leadership for Sustainability Education program. In addition, she teaches the Elementary Science Methods courses in the Graduate Teacher Education Program. Dr. Kelley has spent over a decade working with K-12 students conducting authentic projects in classroom and community-based settings. Prior to her work in education, Dr. Kelley worked as an environmental scientist and aquatic toxicologist.

**Daniel Prince**, MS, is the Coordinator of Outdoor School Programs at Multnomah Education Service District in Portland, Oregon. His Masters degree is in Education Administration is from Portland State University, and his B.A. in Philosophy is from Whitman College. His twenty-five years in experiential, environmental education has generated a strong interest in understanding how non-formal, co-curricular learning experiences support student learning in STEM and other disciplines.

**Nicole Rigelman**, EdD, is an Associate Professor of Mathematics Education in the Graduate School of Education at Portland State University. Her doctorate in Educational Leadership with a focus on Mathematics Curriculum and Instruction is from Portland State University as is her Masters in Education and her BS in Integrated Science. Her research interests lie in mathematics teacher preparation and professional development, influence of curriculum materials on student and teacher learning, and standards implementation and assessment – specifically ways that standards and assessments can positively influence classroom instruction. Ms. Rigelman was a member of the Portland Metro STEM Partnership's common measures committee.

**Ellen Skinner, Ph.D.** is a Professor in the Psychology Department at Portland State University. She received her Ph.D. in Human Development from Pennsylvania State University, and held tenured positions at the Max Planck Institute for Human Development and Education in Berlin and University of Rochester before coming to PSU. As part of Psychology’s concentration in Developmental Science and Education, her research focuses on the development of children’s academic identity and motivational resilience in school. As part of the Common Measures team, she contributed ideas about how to assess students’ perspectives on their engagement, identity, and supportive teacher relationships in STEM.
Highlights:

- The conceptualization stage of a K-12 STEM Common Measurement System was completed.
- The STEM Common Measurement System includes construct definitions for nine diverse variables.
- The STEM Common Measurement System spans K-12 school-level to student learning variables.
- Eight existing measurement tools were selected and an additional five tools will be developed.
- When fully developed, the STEM Common Measurement System will contain 13 measurement tools.
A Common Measurement System for K-12 STEM Education: adopting an educational evaluation methodology that elevates theoretical foundations and systems thinking.

**Introduction**

**Significance of STEM Education.**

Science, Technology, Engineering, and Mathematics (STEM) education is important on a national, regional, local, and individual level for several reasons. First, multiple stakeholders including government and businesses acknowledge that investments in STEM education are key to the United States’ global economic competitiveness (Department of Commerce, 2012; National Academy of Sciences, 2010). To keep that competitive edge, a greater number of students pursuing STEM careers are needed to ensure a STEM capable work force (Carter, 2007). Second, increasing the quality of STEM education is viewed as an important avenue towards creating an informed citizenry that will benefit policy decisions at the national, regional, and local levels (National Research Council (NRC), 2011). Understanding scientific discoveries and methods is relevant for citizenship in a democracy that is dependent on scientific products and technologies (Bencze & Di Giuseppe, 2006). Third, the economic benefits of STEM education to individual students in the form of higher income and greater job security have also been clearly established (Carnevale, Strohl, & Melton, 2011; Zaback, Carlson, & Crellin, 2012). For example, in their analysis of the ten college majors associated with highest median income eight out of the ten majors were engineering majors with the remaining two majors also in STEM fields (Carnevale et al., 2011).

Given the above-described significance, a new emphasis on STEM education as using a coherent, integrated, and multi-disciplinary approach has gained prominence. This new approach to STEM education reflects the ways that STEM concepts and higher-order thinking skills are actually applied in the real world by scientists, engineers, and other professionals in order to recognize, evaluate, and solve complex problems and discover and advance new knowledge (Lewis, 2006; Newmann, Marks, & Gamoran, 1996). Stakeholders from diverse backgrounds recognize the value of high quality STEM education for delivering not only STEM content knowledge, but also the 21st century skills that all students will need to engage as successful members in our workforce and society. Despite wide-spread recognition of the importance of exemplary STEM education, the United States (US) educational system has found it difficult to deliver on this promise, therefore, change is needed to an education system that is fraught with many problems.

**Evidence of the Shortfalls of the Current US STEM Education System.**

The importance of and increased emphasis on STEM education is not a trend unique to the US (Chubb & Kusa, 2012), but for the purposes of this paper the US Education System is critically examined as a case and as the context in which this study took place. There are many problems facing STEM education in the US. In considering the most critical problems facing STEM Education, the diversity of these problems is striking. These diverse problems span from student underperformance in STEM disciplines to shortfalls in teacher practice to limitations of current measurement systems.
Student underperformance in STEM. Both national and international assessments of student mathematics and science performance reveal shortfalls in STEM education. Only 39% of 4th grade students and 34% of 8th grade students score as proficient or advanced in mathematics on the National Assessment of Educational Progress (NAEP) (National Center for Educational Statistics, 2012a). In science, only 31% of 8th grade students score as proficient or advanced on the NAEP (National Center for Educational Statistics, 2012b). The Trends in International Mathematics and Science Study (TIMSS) provides further evidence of poor student performance in STEM. These findings are especially relevant when the percentage of students meeting TIMSS advanced benchmarks are considered, given the close association between the high level of content and skills described in the advanced benchmarks and college readiness indicators (Conley, 2007a; 2007b; National Center for Educational Statistics, 2013). Nationally, only 13% of 4th grade students and 7% of 8th grade students score at advanced benchmarks on the mathematics version of TIMSS (National Center for Educational Statistics, 2013). Similarly, only 15% of 4th grade students and 10% of 8th grade students score at advanced science benchmarks (National Center for Educational Statistics, 2013). Taken together, recent NAEP and TIMSS scores provide clear evidence that US students’ are underperforming in STEM.

Shortfalls in teacher practice. There is an increasing awareness that current K-12 teaching practices tend to isolate STEM disciplines, emphasize rote memorization of STEM content, and neglect higher-order thinking skills (Niemi, Baker, & Sylvester, 2007; NRC, 2001). The purpose of this article is not to review the current findings on teaching practices in STEM; however, a few examples from the literature are provided to illustrate the need for improvement of teacher practice in STEM. In their study of elementary mathematics instruction, Rowan, Harrison, and Hayes (2004) note the repetition of content across grade levels, emphasis on breadth of curriculum coverage over depth of coverage, and inconsistency of both “content coverage and teaching practice among teachers within the same school, even when these teachers work at the same grade level” (p.121). Findings from research on science education convey similar trends. Gallagher (2000) notes a predominance of teaching for lower-order skills and a failure to target deep conceptual understanding. Typical science instructional practice at the elementary level is depicted as “…teacher centered, textbook driven, and geared toward achieving lower level knowledge and comprehension objectives… [with] at best, few student-centered, inquiry-based activities” (Schwartz, Abd-El-Khalick, & Lederman, 2000, p.191).

Moreover, national accountability policies further compound the shortfalls in STEM teaching practices. In a review of research literature on test-based accountability policies in science education, Anderson (2011) finds a conflict between research grounded, reform-based efforts and accountability policies based on standardized testing. In this research review, the conclusion from teacher and administrator perspectives is clear: “accountability limits time and effort spent on science, drives the remaining science instruction toward memorization of facts, and constrains student learning” (Anderson, 2011 p. 121). Given that mathematics is also a subject associated with high stakes testing, it is reasonable to infer that the conflict between reform efforts and accountability policies is relevant to K-12 mathematics reform efforts as well.

Limitations of measurement systems. Current measurement systems in STEM education are associated with multiple limitations that hinder efforts to improve outcomes for students and teachers and thus are a contributing factor to the persistence of the problems described above.
Generating and sharing information about how to improve STEM teaching and learning depends on the quality of measurement systems, therefore, addressing these limitations is key to transforming STEM education. Measurement of key constructs in STEM education are difficult or problematic at nearly every level of the education system.

The current capacity for the measurement of key student outcomes for college and career readiness in STEM education is associated with significant limitations. Many scholars advocate for changes to the format, goals, and frequency of assessment of student learning (Anderson, 2011; Black & William, 1998; 2009; Gallagher, 2000; Yeh, 2006). Similarly, Ball and Hill (2009) characterize the measurement of teacher practice, in general, as under-developed. Further, the measurement of teacher practice in STEM education specifically is an area that is relatively new, and as a result, associated with few measurement tools that work across the STEM disciplines. Pianta and Hamre (2009) emphasize the need for better teacher practice measurement instruments, “…we need more evidence on why and how classrooms, and teachers, matter; the need for evidence is not trivial…” (p.110). The need for teacher practice measurement instruments that are effective, feasible, valid, and allow for the investigation of links between specific teaching strategies and student learning outcomes is also a need that is common across STEM disciplines (Ball & Hill, 2009; Hiebert & Grouws, 2007; Lewis, 2005).

Moving beyond assessment of individual students and teachers, measurement strategies for programs and systems are also lacking and underdeveloped. There is a clear need for appropriate measurement of teacher professional development programming. Crow (2011) stressed “using [teacher practice] data for professional learning at the individual level has not yet been systematic or widespread” (p.26). Finally, the limitations of measurement instruments extend into organizational variables that are of relevance to the quality of STEM education. Matsumura, Garnier, Pascal, and Valdes (2002) demonstrate the extension of measurement needs to the school level “New indicators that help schools, districts, and states monitor and support efforts to improve the quality of instruction are clearly needed. These indicators are important for providing feedback to schools and districts about their interim progress towards reform goals” (p. 208). Taken together, these sources demonstrate the limitations of current measurement tools across many of the constructs that are most relevant to STEM education.

Shulman (2009) asserts that “…assessment is a powerful tool for raising the quality of teaching and learning. It should be used diagnostically and interactively, not as a form of autopsy” (p. 237). These limitations in both measurement tools and the methodologies that currently dominate educational evaluation and research are contributors to the shortfalls in STEM education because without meaningful, iterative assessments, teachers, administrators, schools, and districts are lacking critical information and direction for improvement. If applied differently, however, in a common and formative manner, measurement tools could be a large contributor to innovative changes in STEM education.

Changes Needed in Evaluation, Research, and Measurement Priorities.

Given the importance of STEM education to the future of national economies, communities, and individual’s lives and given the significant challenges that face K-12 STEM education, new approaches are needed in evaluation, research, and measurement. Scholars both within the
The adoption of common measurement tools is representative of a systems approach to addressing problems in education. The value of common measurement systems is related to 1) the ability of the systems approach to empirically investigate questions about interconnections among important variables of complex school systems, and 2) a shift in evaluation and research studies towards an emphasis on formative data to promote ongoing change and learning among collaborators.

First, a common measurement system should be designed to enable the investigation of questions about the interconnections among important variables of complex school systems. The variables that indirectly and directly impact student learning in STEM education are multi-faceted (Leithwood, 1992; Reynolds & Walberg, 1992; Tschannen-Moran & Barr, 2004). While diverse studies that isolate specific variables that impact student learning have been conducted and shed light on the importance of and interactions between some variables, many questions regarding the connections in educational system remain unanswered. These questions span the full spectrum of variables that contribute to student learning from school-level to teacher practice to student learning variables. For example, the NRC (2011) articulates the difficulty of identifying successful schools because of a lack of common outcomes to judge success. Many questions regarding the connections between teacher professional development and student learning remain unanswered (Thoonen et al., 2011b). Finally, scholars note the need for definitive answers about specific effects of teacher practice on student learning (Ball & Hill, 2009; Gittomer, 2009). The development of a common measurement system for STEM education with a focus on high leverage, relevant, and interconnected variables enables these questions to be empirically investigated (Sun & Leithwood, 2012).

Second, common measurement systems allow for a shift in priorities for evaluation and research from a focus on individual program impact and summative data to an emphasis on formative data designed for optimizing learning environments, promotion of change, and learning among collaborators. For evaluators, common measurement systems can create evaluation plans that contribute to positive change by formatively assessing project goals and providing timely feedback rather than just documenting change (Kramer, Parkhurst, & Vaidynathan, 2009; Matsumura et al., 2002). For researchers, common measures can empower researchers to shift focus towards optimizing learning environments by abandoning objective observer roles and controlled experiments for interventionist strategies and theory-based design with multiple iterations to optimize the design of learning environments (Lamberg & Middleton, 2009; Sandoval & Bell, 2004). It stands to reason that these shifts in evaluation and research priorities,
supported through common measurement systems, could significantly contribute to the transformation of STEM education.

Importance of theory development. Sound measurement begins with construct conceptualizations based on strong theoretical foundations (Wilson, 2005; Wolfe & Smith, 2007). Organizations addressing complex social problems can extend this best practice of sound measurement to the development of common measurement systems. Thoonen et al. (2011b) assert, “To describe and understand the nature of organizational changes, scholarship requires both dynamic theories of organizational processes and sociocultural interactions and dynamic methods that can model changes in an organization’s capacities and growth” (p. 522). The requirements emphasize the need for theory development and a measurement system that enables evaluators, researchers, and practitioners to measure the influence of changes made in one part of the complex education system to other related areas of the same system. Following this line of reasoning, the first stage in the development process for common measurement systems should be a conceptualization stage that involves an iterative cycle of theory development and construct definition. This paper describes an example of a common measurement system development project by describing both the process and the results of the conceptualization stage.

In conclusion, given the clear importance of high-quality STEM education and the limitations of the current US education system to reach the goals of high-quality STEM education, the educational research and evaluation community must contribute a viable solution. What is needed is a clear understanding of what variables are the most high leverage variables to create change, a sound understanding of the relationships between those key variables to explicate how interventions in one part of the system impact outcomes in the other parts, and mechanisms to provide key stakeholders in STEM education with timely and high-quality data to drive decision making. Arguably, educational researchers and evaluators are poised to contribute sound theory and measurement which could meet these needs and serve as a key part of improving STEM education. Therefore, the development of a common measurement system for STEM education that is based in strong theoretical foundations is needed to begin address the above outlined limitations of STEM education in a coherent way.

Methods

Project Context.

The Portland Metro STEM Partnership is a Collective Impact Partnership formed to improve STEM education. The partnership is aligned around five key conditions of collective success, as defined by Kania and Kramer (2011): (1) shared agenda, (2) common measurement, (3) mutually reinforcing activities, (4) continuous communication, and (5) backbone organization. The partnership is composed of diverse organizations that represent the multiple stakeholders in STEM education including K-12 school districts, higher education universities, informal and community education organizations, and businesses. The partners have agreed to coordinate their programs and resources and adopt common measures to achieve a collective impact on STEM education in the Portland metropolitan region. This paper describes the conceptualization stage of the development of the partnership’s common measurement system.
Formation of the Common Measures Committee. Complex social problems require solutions that are collaborative and leverage cross-organization resources. Vaidyanathan (2012) contends “Shared Measurement Systems offer an alternative approach to evaluating complex social problems. By defining measures in a common manner across organizations working on [a common] issue … and then providing the associated data collection tools, data capture and reporting mechanisms, organizations in the field can have more meaningful discussions about the outcomes they are achieving and sharing lessons. Shared measurement provides a mid-point between the extreme evaluation approaches of experimental design and anecdotal case studies.” (p. 56).

The Portland Metro STEM partnership began the development of their common measurement system with the formation of a committee to guide the conceptualization stage of the process. The Common Measures Committee was composed of representatives from the multiple stakeholders in the partnership and a committee leader, the partnership’s appointed Director of Research and Assessment. There are several reasons why the inclusion of multiple stakeholders was important in the conceptualization process. First, the intentional collaboration of stakeholders in education is portrayed as a precursor to successful transformation of education and a key characteristic to achieving “collective impact” (Kania & Kramer, 2011). Second, the involvement of multiple stakeholders allows for the consideration of diverse perspectives in the conceptualization process, which lends credibility to the resulting measurement system (Vaidyanathan, 2012). Finally, Wallace (2008) argues that project stakeholders are the best suited to make sound decisions about evaluation design

“Decisions must be made about what outcomes to measure, how to measure these outcomes, when to measure the outcomes, and who is the appropriate comparison group? Who possesses the most intimate knowledge able to inform the answers to these critical questions? This author and practicing evaluator argues it is the stakeholders” (p.201).

The authors of this paper include the committee leader and six additional committee members that together represent the partnership’s K-12, university, community education, and business partners.

Description of the Process Used in the Conceptualization Stage.

The Portland Metro STEM Partnership’s (PMSP) theory of change guided the work of the Common Measures Committee (Figure 1). Given that the shared goal of the partnership is creating effective student learning environments in STEM, the committee began the conceptualization process with a focus on student outcomes and then moved backwards through the partnership’s theory of change. The Common Measures Committee developed a three-step process for completing the conceptualization of the STEM Common Measurement System: 1) prioritization and definition of constructs, 2) evaluation of currently available measurement tools, and 3) determination regarding measurement tool selection. First, the committee accessed research literature and the collective expertise of the partnership to prioritize and define constructs for each area of the theory of change. Second, the committee collaboratively brainstormed criteria for evaluating available measurement tools, ranked the criteria in importance according to the needs of the partnership, and then used the criteria to guide the evaluation of potential measurement instruments. The highest ranked evaluation criteria and the rationale for those criteria are provided in Table 1. Third, based on the evaluation of existing measurement instruments the committee made a decision regarding instrument selection. Two
types of decisions were made during this third step of the process 1) selection of an existing instrument or 2) determination that no existing instruments met adequate criteria which led to the decision to develop a new instrument. This three-step process was repeated for each construct and each area of the theory of change. The next section in this paper describes and discusses the results of the conceptualization process.

Results of the Conceptualization Process: STEM Common Measurement System Description

Effective student learning environments in STEM.

Rationale. Few of the assessments in K-12 education that are commonly used across schools and districts in any given state directly measure college and career readiness. In fact, Katsinas and Bush (2006) contend that there is no relationship between the contents measured in standardized, state-mandated tests and the essential elements of college or career readiness. Consequently, if STEM educators are to be empowered and informed by assessment data as they strive to ensure that all students are ready for college and careers in STEM, a different set of outcomes and corresponding measures need to be prioritized. Three outcome areas were deemed to be important markers of students’ progress toward college and career readiness, and therefore, most relevant in describing and monitoring effective student learning environments in STEM; these included outcomes in the conceptual, cognitive skill, and affective dimensions of student learning. The prioritization of these outcomes over other possible outcomes was based on the college readiness literature (Conley, 2007a, 2007b) and the collective expertise of the partnership. The partnership identified application of conceptual knowledge, higher-order cognitive skills, and academic identity and motivational resilience as the constructs for effective learning environments in STEM.

Application of STEM conceptual knowledge.

In the STEM Common Measurement System, the construct definition for application of STEM conceptual knowledge is drawn from Lingard, Mills and Hayes (2006): “students' understanding of and thinking about ideas, theories and perspectives considered critical or essential within an academic or professional discipline or in STEM interdisciplinary fields recognized in authoritative scholarship....References to isolated factual claims, definitions, or algorithms are not indicators of significant disciplinary content unless the task requires students to apply powerful disciplinary ideas which organize and interpret information” (p. 94) [emphasis added].

In their analysis of middle school science curricula, Stern and Ahlgren (2002) found that the materials studied largely did not provide assessments that test application of concepts; rather typical items require only knowledge recall. These authors’ findings regarding the shortfalls in terms of expectations and goals for application of conceptual knowledge in science education also holds true for the broader field of STEM education. The purposeful move towards application of conceptual knowledge, rather than rote memorization of content is important for college and career readiness in STEM. Silva (2009) contends, “An emphasis on what students can do with knowledge, rather than what units of knowledge they have, is the essence of 21st century skills.” (p.630). The focus on deep understanding and application of conceptual knowledge is key to student success in STEM because it more accurately reflects the way concepts are applied in the real world by scientists, engineers, and other STEM professionals.
This outcome stands in stark contrast to rote memorization of isolated facts, definitions, or formulas because the application of conceptual knowledge outcome is hypothesized to result in deeper and longer lasting understanding of STEM content and the ability to transfer understanding to new contexts.

Measurement of application of conceptual knowledge. Feltovich, Spiro and Coulson (1993) argue for closer ties between desired outcomes, assessment, and instruction and note that although this seems an obvious requirement for the achievement of student application of content knowledge, it is rarely implemented in practice. Stern and Ahlgren (2002) provide tangible examples and detail 10 categories of tasks that measure application of science knowledge, for example, one task category is:

“Consider the appropriateness of a representation for an idea or compare a representation with the real thing. For example, relevant to the idea that plants make their own food: Some people say that plants are like food factories. Do you agree? Why or why not?” (p. 900).

This task category framework is aligned to the application of content knowledge construct and will guide the development of application of conceptual knowledge of key STEM concepts. Key STEM concepts for each grade level will be determined by consulting the core concepts in the new national standards in mathematics and science (National Governors Association. 2010; Achieve, 2013).

In addition to an assessment task framework, there is a need to balance the tensions between item format and feasibility of use in classrooms and informal education programs. Lee, Liu, and Linn (2011) report that multiple choice items sometimes fail to differentiate between students, but constructed response items provided more useful and detailed information about student ability levels. Constructed response items are, however, more time intensive to analyze and as a result more difficult to use effectively as formative assessments. The STEM Common Measurement System will use a combination of carefully designed multiple choice and constructed response items that will be developed to balance the limitations of both item types with the feasibility of their use in classrooms.

Higher-order cognitive skills.

The construct definition for higher-order cognitive skills was adapted by committee members from a definition provided by Wood, Darling-Hammond, Neill, and Roschewski (2007): students’ higher order thinking or cognitive skills refer to their abilities to:

1. Problem solve: (a) identify, frame, and solve complex problems, (b) apply knowledge and skills to novel problems and/or situations, and (c) assess the reasonableness of a process and/or solution;
   Develop an argument based on evidence: (a) find, evaluate, analyze, and synthesize information and (b) evaluate complex ideas or chains of reasoning;
   Communicate ideas, solutions, arguments, or conclusions in oral and/or written form; and
   Utilize metacognitive skills: (a) reflect on one’s own thinking and reasoning and (b) choose and strategically use tools (technological and otherwise).

The emphasis on higher-order cognitive skills as educational outcomes is not by any means new (Silva, 2009); in fact they are prevalent in older standards. There is ample recent evidence, however, that suggests the US educational system continues to fail to reach these goals. In her
review of research focused on elementary and middle school students’ scientific thinking skills, Zimmerman (2007) notes a trend of educators underestimating children’s reasoning skills, which results in fewer opportunities for students to practice those higher-order cognitive skills. Zohar and Nemet (2002) similarly note that an emphasis on outdated hierarchical theories of learning result in a trend of instruction that fails to address high-order learning goals because perceived “prerequisite skills” are given priority with the resulting implication being that “students often never get to...engage in higher-order skills” (p. 36). These higher-order cognitive skills are believed to be important for student success in college, employment in STEM careers, and participation as informed citizens because they represent the use of cognitive skills to recognize, evaluate, and solve complex problems, discover and advance new knowledge, and create solutions to complex real world problems (Lewis, 2006; Silva, 2009). The further emphasis on these skills in the new national standards’ through clear definitions of mathematical and scientific practices (National Governors Association. 2010; Achieve, 2013), gives additional evidence of their importance to college readiness.

Measurement of higher-order cognitive skills. In K-12 STEM education, cognitive skill constructs are often measured with science inquiry, engineering design, and mathematical problem solving performance-based work samples. Performance-based assessments represent a type of assessment instrument that show potential for the measurement of complex constructs such as higher-order cognitive skills (Ennis, 1993; Jackson, Draugalis, Slack, & Zachry, 2002; Johnson, Penny, & Gordon, 2008; Lombask & Baron, 2003; Ruiz-Primo & Shavelson, 1996). Student generated responses are essential as evidence of thought processes and reasoning. Performance-based assessments, therefore, are well suited for the assessment of cognitive skills because students are required to independently generate the evidence of these cognitive skills. In addition, performance assessments have been demonstrated to have positive impacts on instruction and student learning (Goldschmidt, Martinex, Niemi, & Baker, 2007; Lane, 2004).

The exact design and implementation of performance-based work sample tasks, however, are extremely relevant to the validity of these assessment instruments. Shavelson, Baxter, and Gao (1993) report that rater and task all pose threats to reliability and that measurement method (oral presentation versus written work sample) poses threats to convergent validity. In the STEM Common Measurement System, the higher-order cognitive skill assessment instruments will be performance-based written work samples with consistent, common tasks developed by the partnership across each grade level and each STEM discipline. The partnership will also invest in strategies shown to increase the quality of performance-based assessment systems through rater training, calibration discussion, and teacher professional development focused on assessment practices (Niemi et al., 2007).

Finally, the partnership’s definition of STEM higher-order cognitive skills is complimentary to the definitions of mathematical and scientific practices in the new national standards, which are widely considered college readiness standards (National Governors Association, 2010; Achieve, 2013). Therefore, the emphasis on these higher-order cognitive skills in the STEM Common Measurement system will help bring classroom assessments in closer alignment and congruence with the new standardized assessments that will accompany the national standards.

Academic identity and motivational resilience.
To be prepared to succeed in STEM college majors and careers, students must develop an academic identity and the motivational resources that allow them to work hard and persist despite obstacles and setbacks. From the large literatures on academic motivation and self-perceptions (Wigfield, Eccles, Schiefele, Roese & Davis-Kean, 2006), the common measure committee selected four markers of Academic Identity, which are defined broadly as comprising students’ deeply held views of themselves and their potential to enjoy and succeed in STEM classes and careers. The four components of academic identity that will be measured in the STEM Common Measurement System are: (1) a sense of belonging or relatedness, which refers to students’ feelings about whether “people like them” are welcome and would be accepted in the study and professions of STEM (Deci & Ryan, 2000; Furrer & Skinner, 2003; Ryan and Deci, 2000; Skinner & Wellborn, 1997); (2) perceived competence or self-efficacy, which describes students’ beliefs about whether they have the ability to succeed in STEM classes and fields (Deci & Ryan, 2000; Ryan and Deci, 2000; Skinner, Wellborn, & Connell, 1990); (3) autonomy or ownership, which refers to whether students are personally committed to the work in STEM classes and careers (Deci & Ryan, 2000; Ryan and Deci, 2000); and (4) purpose, which relates to whether students are convinced that classwork and professional work in STEM is meaningful, important, and worthwhile (Damon, Menon, Bronk, 2003; Ford & Smith, 2007). Together, these four facets of academic identity have been shown to be strong predictors of students’ motivation, engagement, learning, and success in school (Wigfield et al., 2006).

Motivational resilience refers to students’ enthusiastic hard work and persistence in the face of challenging STEM coursework. The STEM Common Measurement System will measure two components of motivational resilience including academic engagement and constructive coping/resilience. Academic engagement is defined as high quality participation in academic work, including effort (hard work, exertion, follow-through) and enthusiasm (interest, curiosity) (Skinner, Kindermann, & Furrer, 2009; Skinner & Wellborn, 1997). The resilience or constructive coping component is defined as adaptive strategizing (including problem-solving and help-seeking) and persistence in the face of academic challenges, obstacles, and setbacks (Skinner & Zimmer-Gembeck, 2007; Skinner & Wellborn, 1997).

Academic Identity as a STEM capable learner is conceptualized as a fundamental transformation that students need to accomplish if they are to be ready for college and careers in STEM. The various components that make up academic identity are considered basic psychological needs and are precursors for the academic engagement and resilience students (or motivational resilience) need to achieve in STEM. Each component of academic identity is conceptualized as a necessary condition, but any condition alone is not a “…sufficient condition for engagement in learning activities” (Skinner, Wellborn, & Connell, 1990). Whole-hearted engagement and tenacity in demanding STEM classwork is, in turn, essential to student learning and achievement (Furrer & Skinner, 2003; Skinner & Wellborn, 1997; Skinner, Wellborn & Connell, 1990). In summary, educational contexts can invigorate engagement by supporting student’s basic needs, which include envisioning themselves as competent to succeed, related (belonging) in school, and autonomous or self-determined learners (Deci & Ryan, 2000; Ryan & Deci, 2000).

Measurement of academic identity and motivational resilience. The partnership will use student self-report and teacher-report surveys to measure academic identity and motivational resilience (Skinner, Kindermann, & Furrer, 2009). Motivational resilience includes a set of items tapping engagement, adapted from Skinner et al. (2009), who found that the two assessment instruments
(student self-report survey and teacher report survey) were moderately correlated; however, students tended to report higher levels of behavioral engagement and teachers tended to report lower levels of emotional engagement. In addition, coping items were adapted from a survey instrument structure that employs survey scales that include “stems,” which describe the stressful academic situation, and “items,” which describe the children’s coping responses (Skinner & Wellborn, 1997; Skinner, Pitzer, & Steele, in press). The combination of these two assessment instruments will create indicators of student’s academic identity and motivational resilience that balance the limitations of either instrument in isolation.

Further, the STEM Common Measurement System will include core items for each component of academic identity and motivational resilience, and also offer a „menu” of additional items for each component that individual programs can use to conduct a deeper investigation of the impacts of specific interventions. For example, if a program is working explicitly on the academic engagement component of motivational resilience they will be able to administer not only the core items survey that already contains items associated with academic engagement, but will also be able to add more items to further measure that component. Skinner, Chi, and LEAG (2012) found that, although academic identity and motivational resilience are multi-dimensional constructs, the “dimensions were sufficiently inter-correlated to allow the creation of aggregated measures with satisfactory internal consistencies” (p. 31) and convergent validity. These authors’ findings indicate that brief measures like the core items proposed for the STEM Common Measurement System can be used as reliable and valid measures of student academic identity and motivational resilience. Further, this particular study was conducted in an informal education setting, which indicates that these measures should work across the diverse contexts of the partnership (Skinner, Chi, & LEAG, 2012).

**Summary.** The three student constructs will guide the design of STEM learning environments and inform educators with assessment data for learning. Yeh (2006) emphasizes the importance of frequent, fast, and common diagnostic assessment data to create “…fundamental changes in the organization and culture of schooling, encouraging dialogue among teachers by providing a common point of discussion, increasing collaboration among teachers to improve instruction and resolve instructional problems, reducing teacher isolation, and supporting both new and experienced teachers in implementing sound teaching practices.” (p. 657).

With this goal in mind, the partnership will seek to design common student assessment tools that can be applied in a formative and summative manner in STEM classrooms and informal education environments. Common measurement of these student outcomes will be particularly important with the advent of the Common Core State Standards in Mathematics and the Next Generation Science Standards, both of which significantly raise expectations for students toward performance levels that relate to college and career readiness.

**Effective practices for STEM learning environments.**

**Rationale.** As briefly outlined above in the introduction of this paper, the reliable, valid, and feasible measurement of important teacher and informal educator practices is a high needs area in STEM education. Current teacher practice measurement instruments either isolate STEM disciplines, lack feasibility for large-scale use, or provide only summative data, rather than
formative data to improve instructional practices (Crow, 2011; Kane, 2012; Matsumura, Garnier, Pascal, & Valdes, 2002; Matsumura, Garnier, Slater, & Boston, 2008). Additionally, there is a mismatch in terms of the rigor of required work and the structure of problems (i.e. inauthentic, structured vs. real-world, ill-structured) between both K-12 and undergraduate educational systems and between the K-16 education system and STEM careers (Hazari, Potvin, Tai, & Almarode, 2010). This lack of continuity across the educational system represents a failure to focus on characteristics of learning environments that will achieve college and career readiness outcomes for students. Teacher and informal educator practices that were determined to be the most important contributors to student STEM college and career readiness were pedagogical content knowledge, instructional practices, and supportive teacher-student relationships.

**STEM pedagogical content knowledge**

The common measures committee’s search of STEM education literature found that pedagogical content knowledge (PCK) studies are most prevalent in technology, mathematics, and science education literature. An analysis of the available literature revealed a lack of a consistent PCK definition even within a single discipline (Park & Oliver, 2008), but also revealed some common trends in definitions, which informed the creation of a STEM PCK construct definition. The STEM Pedagogical Content Knowledge construct definition is composed of three parts:

1. Teachers’ knowledge of student thinking about specific STEM topics including prior knowledge, misconceptions, learning progressions, common difficulties, and developmentally appropriate levels of understanding (Carpenter, Fennema, Peterson, & Carey, 1988; Grouws & Schultz, 1996; Hill, Ball, & Schilling, 2008; Lee, Brown, Luft, and Roehrig, 2007; Magnusson, Krajcik, & Borko, 1999; Manizade & Mason, 2011, Park, Jang, Chen, & Jung, 2011; Schmidt et al., 2009; Schneider & Palsman, 2011; Shulman, 1986, 1987; Tamir, 1988; Tirosh, 2000).

2. Teachers’ understanding and use of the effective strategies for specific STEM topics including strategies to engage students in inquiry, represent STEM phenomena, and guide discourse about the STEM topic (Carpenter et al., 1988; Hill et al., 2008; Lee et al., 2007; Magnusson et al., 1999; Manizade & Mason, 2011, Park & Oliver, 2008; Schmidt et al., 2009; Schneider & Palsman, 2011; Shulman, 1986, 1987; Tamir, 1988).

3. Teachers’ integration of technology to enhance instruction of specific STEM topics in meaningful and appropriate ways to promote key student College and Career Readiness outcomes. (Graham et al., 2009; Niess, 2005; Schmidt et al., 2009).

Conceptions of effective teaching that are limited to instructional practices are incomplete because in addition to these practices a specialized knowledge for teaching, called pedagogical content knowledge (PCK), is also needed (Shulman, 1986). A teacher’s PCK, therefore, is an important outcome in describing the quality of teaching because it impacts every part of a teacher’s professional practice from planning, implementation, assessment, reflection and revising for the future (Park et al., 2011; Shulman, 1987).

**Measurement of STEM pedagogical content knowledge.** In the literature, scholars stress that PCK is challenging to measure (Lee et al., 2007; Loughran, Milroy, Berry, Gunstone, & Mulhall, 2001; Loughran, Mullhall, & Berry, 2004). Loughran et al. (2001) describe the inherent complexity and context dependence of PCK. These authors further categorize the idea that there is a single correct approach to the integration of PCK into teacher practice as “untenable.” With no clear cut “right answer” to PCK, measuring PCK is a parallel measurement challenge to that
of measuring student reasoning about ill-structured problems. In the context of student reasoning, performance-based assessments that use rubrics to score student constructed answers are often touted as suitable measurement tools (Jackson et al., 2002; Lomask & Baron, 2003). Given that the “ability to reason about teaching” (Shulman, 1987) is the target in PCK measurement, performance-based assessments that involve rubrics as their scoring mechanism are expected to be a useful measurement tool.

In addition, rubric-based measurement tools afford a level of versatility in terms of what product they can be applied to. Loughran et al. (2001) note that PCK may not be readily observable at one single point in time such as during an observation or an individual lesson. Given that many scholars emphasize the importance of PCK throughout every portion of a teacher’s professional practice (Park et al., 2011; Shulman, 1987), then a PCK measurement tool would ideally be versatile so that it can be used to measure PCK in planning, implementation, and other stages of a teacher’s practice. While other scholars have attempted to create PCK rubrics, these rubrics are discipline specific (Park et al., 2011; Lee et al., 2007). The partnership, therefore, has determined that a STEM PCK rubric should be developed for the measurement of the STEM PCK outcome. The STEM PCK rubric will be an analytic rubric with at least one rubric category designed to align to the three components of STEM PCK in the construct definition.
**Effective Instructional practices in STEM.**

The construct definition for effective instructional practices in STEM is derived from a literature review of predominately science, mathematics and engineering education research literature. The following five practices were found to be important to student college and career outcomes across the disciplines, therefore, these five practices compose the partnership’s construct definition.

1) Teachers facilitate active engagement of students in their learning.
   a. Teachers assume the role of facilitator rather than authority figure (Anderson, 2002; Chapin & O’Conner, 2007; Crawford, Wood, Fowler, & Norrell, 1994; Fortus, Dershimer, Krajcik, Marx, & Mamlok-Naaman, 2004; Hiebert et al., 1997; Marshall, 2009; Mehalik, Doppelt, & Schuun, 2008; Miner, Levy, & Century, 2010; Secada, Foster, & Adajian, 1995)
   b. Students assume the role of active learners making sense of learning activities for themselves (Anderson, Reder, & Simon, 1996; Anderson, 2002; Carpenter, Fennema, Franke, Levi, & Empson, 1999; Fortus et al., 2004; Fosnot & Dolk, 2001; Hiebert et al., 1997; Mehalik et al., 2008)

2) Teachers emphasize deep content knowledge and higher-order cognitive skills by addressing learning goals in both areas. (Apedoe, Reynolds, Ellefson & Schunn, 2008; Ascbacher & Roth, 2002; Gallagher, 2000; Henningsen & Stein, 1997; Lewis, 2006; Miner et al., 2010; Newman & Wehlage, 1993; Puntambekar & Kolodner, 2005; Smith & Stein, 1998; Stein & Lane, 1996; Stephens, McRobbie & Lucas, 1999; Stigler & Hiebert, 2007; Zimmerman, 2007).

3) Teachers create and implement multiple and diverse opportunities for students to develop conceptual knowledge and cognitive skills (Anderson, 2002; Ascbacher & Roth, 2002; Brophy, Klein, Portsmore, & Rogers, 2008; Lesh, Post, & Behr, 1987; Prain & Waldrip, 2006; Puntambekar & Kolodner, 2005; Sadler, Coyle & Schwartz, 2000; Shulman, 1987; Stein, Smith, Henningsen, & Silver, 2009).

4) Teachers use frequent formative assessments (and summative assessments) to facilitate diagnostic teaching and learning. (Bell & Cowie, 2001; Black & Wiliam, 1998; 2009; Miner, Levy, & Century, 2010).
   a. Teachers and students are both stakeholders in the assessment process. The teacher’s role includes setting clear, developmentally-appropriate learning targets or performance criteria (Frederiksen & White, 1997; Webb & Jones, 2009) and selecting or developing formative assessment tasks that align with learning goals (Bell & Cowie, 2001; Marshall, 2009; Ruiz-Primo & Furtak, 2006; Tomanek, Talanquer, and Novodvorsky, 2008). The student’s role includes assuming ownership over their learning (Anderson, 2002; Bell & Cowie, 2001) and engaging in metacognitive activities (Bingham, Holbrook, & Meyers, 2010; Higgins, Harris, & Kuehn, 1994; Fernandes and Fontana, 1996; Frederiksen & White, 1997; Marshall, Horton, & Small, 2008; Stefani, 1994).
   b. Teachers and students contribute to a classroom culture of assessment for learning (i.e. class wide focus on learning, student “honesty about understanding, mistakes, and feedback,” and “an emphasis on dialogue and exploratory talk to support thinking” (Webb & Jones, 2009).

5) Teachers implement learning activities that students find to be relevant, important, worthwhile, and connected to their cultural and personal lives outside of the classroom and encourage students to actively use the connections and real-world examples in their thinking (Newmann et
al., 1996; Newman & Wehlage, 1993; Rivet & Krajcik, 2008; Thoonen, Sleegers, Peetsma, & Oort, 2011a; Zohar & Nemet, 2002).

Any one of these instructional practices in isolation will likely fail to help teachers achieve the outcomes for students that are characteristic of effective learning environments in STEM. The careful combination of these practices, however, is expected to contribute to student college and career readiness in STEM. Additionally, practices such as those listed above can be implemented with various degrees of quality and it is the quality of implementation that has great bearing on the intellectual rigor or cognitive demand expected and maintained within the classroom.

Newmann et al., (1996) clearly articulate that questions regarding quality of implementation are of the utmost importance in reaching the intellectual rigor required for “authentic academic performance.” Therefore, it is hypothesized that through combining the effective instructional practices in STEM and monitoring implementation with measures that provide formative data regarding the quality of the implementation of these practices, teachers may better be able to keep intellectual rigor at the levels required for effective student learning environments.

In looking to the literature, a study of the higher-order skills of argumentation provides an example that garners some initial support for the importance of the careful combination of these practices. Zohar and Nemet (2002) found that instructional strategies such as tying specific content to higher-order skills, requiring students to use metacognitive skills, and posing problems that bear relevance and importance outside the classroom all contributed to higher-quality instruction and increases in student argumentation skills. The instructional strategies outlined in the Zohar and Nemet (2002) study map onto the STEM Common Measurement System’s instructional practices construct specifically linking to three of the practice outlined above (IP#2, 4a2, and 5). Given that argumentation skills require students to apply conceptual knowledge, used higher-order cognitive skills, and engage and persist in the rigorous activity of constructing arguments, this example is particularly supportive of the importance of the careful combination of instructional practices to create learning environments with high intellectual rigor such as those that are required for college and career readiness.

Measurement of effective instructional practices in STEM. Instructional practices for both informal and formal educational settings are frequently acknowledged as being complex constructs (Kane & Staiger, 2012; Martinez, Borko, & Stecher, 2012; Peterson, Wahlquist, & Bone, 2000; Stecher et al., 2006). As a result of the complexity of the construct, many scholars advocate for the use of multiple measures in order to ensure high-quality data can be collected (Berk, 2005; Kane & Staiger, 2012; Follman, 1992, 1995; Martinez et al., 2012; Peterson et al., 2000; Stecher et al., 2006; Wilkerson, Manatt, Rogers, & Maughan, 2000). The use of multiple measures of instructional practice enables evaluators, researchers, and practitioners to draw stronger inferences through data triangulation and a careful balancing of the limitations of different measurement tools (Ing & Webb, 2012; Pianta & Hamre, 2009).

The partnership seeks to achieve this balance through a triad of measurement tools that include a classroom observation protocol, an artifact-based instructional instrument, and a student perception of instructional practices survey. The committee chose the UTEACH observation protocol as the observation protocol for the STEM Common Measurement System for several reasons. First, the UTEACH observation protocol has satisfactory reliability and validity.
evidence (Walkington at al., 2011; Kane & Staiger, 2012). Second, an analysis of the UTEACH manual and rubrics revealed that the instrument was well matched to the majority of the instructional practices except the implementation multiple and diverse opportunities (practice number 3 above). Newman and Wehlage (1993) note that classroom observations will often fail to reveal connection between lessons or multiple instructional days that may combine to culminate into more complex experiences for students. The instructional practice related to multiple and diverse opportunities is, consequently, difficult to measure with classroom observations because it is likely to occur over more than one class period or observation. Other measurement tools will be used to balance the limitations of observation protocols.

The artifact-based instrument will be developed by the partnership through a modification of an existing instrument called the Teacher Instructional Portfolio (Saxton & Rigelman, unpublished). The Teacher Instructional Portfolio is designed to measure full instructional units and creates a rich data set around instructional practices that covers more than one lesson allowing for the exploration of connections across lessons and progression of instruction. In order to scale up the use of this instrument to school-wide transformation efforts, the instrument will be modified to document 3-5 days of practice. While classroom observations and artifact-based instruments measure some similar instructional practices to create data triangulation, they also each measure distinct practices creating breadth of coverage of the construct definition (Martinez et al., 2012).

Finally, the STEM Common Measurement System will use a student perception of instructional practices survey as the third instrument for measuring the instructional practices construct. There are three primary reasons for the selection of student surveys as an appropriate third tool. First, student perception of instructional practices surveys can be used across K-12 grade levels to indicate the frequency with which instructional practices are used (Desimone, Smith, & Frisvold, 2010; Follman, 1992, 1995; Porter, 2002). This type of instrument will provide a different time-scaled data source that focuses more on describing quantity, than quality (Desimone et al., 2010); therefore, this quantitative data source will serve as a complement to the richer, more qualitative data provided by the observation and artifact-based instruments described above (Desimone et al., 2010). Second, student surveys of instructional practices are feasible and inexpensive measure tools. Finally, student surveys will be used to measure instructional practices (IP#1, 4b, and 5) because measurement of these practices requires the consideration of student perceptions that can be most validly measured directly from students. For example, regarding the measurement of the implementation of learning activities of relevance to students personal and cultural lives (IP#5), this practice is most directly measured by attending to student perceptions, rather than relying on teacher or observer inferences regarding relevance. Although existing student perception of instructional practices surveys exist, they were not selected because of either a focus on one STEM discipline (Odom, Stoddard, & LaNasa, 2007), a mismatch with the partnership’s defined construct and/or the cost associated with access and use of the surveys (Crow, 2011; Kane & Staiger, 2012). Therefore, the partnership will develop a student survey to measure instructional practices (IP#1, 4b, and 5).

Table 2 shows how the triad of instruments is projected to cover the partnership’s effective instructional practices in STEM construct.
Supportive teacher-student relationships.
The STEM Common Measurement System’s construct definition for supportive teacher-student relationships is composed of three components:


2) Teachers provide challenging learning activities with high expectations, authentic academic work, and clear feedback (Skinner et al., 1990).

3) Teachers explain the relevance of activities and rules while soliciting input from students and respecting their opinions (Deci & Ryan, 2000; Furrer & Skinner, 2003; Ryan & Deci, 2000; Skinner & Wellborn, 1997; Skinner, Wellborn, & Connell, 1990).

Student’s supportive relationships with teachers are critical because they are the basis upon which students construct a positive academic identity and develop motivational resilience (Deci & Ryan, 2000; Skinner & Belmont, 1993; Skinner & Wellborn, 1997). For example, scholars have demonstrated that warm, supportive teacher and student relationships are positively related to students’ beliefs about relatedness, their behavioral and emotional student engagement (Skinner & Belmont, 1993), and students’ use of help seeking or constructive coping strategies (Skinner & Wellborn, 1997). Supportive classroom environments that are associated with authentic learning activities with high expectations can also effectively create a developmentally appropriate “optimal challenge” for students (Deci & Ryan, 2000); the creation of an “optimal challenge” can, in turn, positively impact student perceptions of competence and their ability to persist in the face of challenges (Deci & Ryan, 2000; Skinner & Wellborn, 1997). Finally, teachers can further foster supportive classroom environments by taking the time to explain the importance of activities or learning goals and giving students choice in the learning process (Newman & Wehlage, 1993).

Measurement of supportive teacher-student relationships. The partnership’s construct definition of supportive teacher-student relationships is closely aligned to the Teacher As Social Context Questionnaire’s construct definition, which includes the concepts of involvement (component 1 above), structure (2 above), and autonomy support (3 above) (Skinner & Belmont, 1993; Teacher As Social Context Questionnaire (TASC), 1992). Therefore, a shortened form of the Teacher As Social Context Questionnaire will be used by the partnership to measure supportive teacher-student relationships with one positive and one negatively worded item for each component of the construct. The Teacher As Social Context Questionnaire has previously been shown to have adequate internal consistency within each component as well as discriminability between the components (TASC, 1992), however, the shortened form is expected to function more as a holistic measure of supportive teacher-student relationships because of the reduced number of items. Both student self-report and teacher report surveys of supportive teacher-student relationships will be used by the partnership at the individual student level.

Summary. The three educator practice outcomes represent a comprehensive approach to the measurement of effective practices for STEM learning environments and are designed to address the measurement issues in STEM educator practices that are outlined above in the introduction of this paper. The construct definitions for both instructional practices and pedagogical content knowledge are derived from a literature review of STEM education fields resulting in constructs
that represent areas of commonality and the best practices identified in the four disciplines. This merging of each discipline’s findings represents an important move towards a new interdisciplinary approach to improving K-12 STEM educator practice and a first step in addressing the isolation of STEM disciplines in current K-12 classrooms. The STEM Common Measurement System’s three educator practice outcomes also shift emphasis from retention-oriented content-based instruction to instruction that emphasizes deeper conceptual knowledge and higher-order cognitive skills. Finally, the emphasis on instrument selection that can be used across diverse contexts and in iterative, formative cycles, rather than for solely summative evaluation purposes, will be important in supporting teachers and informal educators in adopting these practices that are key to student college and career readiness outcomes in STEM.

**Effective Teacher Professional Development Experiences in STEM.**

*Rationale.* The goal of this partnership’s professional development is to foster teacher learning in key beliefs, knowledge, and skill areas so that educators can implement effective practices that will create effective STEM learning environments. As a result, the partnership holds its professional development (PD) to a high standard by evaluating PD effectiveness through focusing not only on what is modeled in the PD, but also on what is implemented in the classroom after teachers complete the PD courses. Desimone (2009) contends that specifying a core set of PD outcomes and associated measures will facilitate knowledge gained across the PD field by “…establish[ing] consistency that would contribute to building a knowledge base” (p.186). This contention is complementary to the goal of common measurement systems, which is to facilitate comparative analysis and mutual learning across organizations or programs (Kramer et al., 2009). The partnership defined pedagogical content knowledge, instructional practices, and teacher self-efficacy as the professional development outcomes and measurement priorities.

*STEM pedagogical content knowledge.*

The partnership’s PCK construct and chosen measurement method was already described above, therefore, this section will be constrained to two points of notable relevance to professional development. First, many scholars list a focus on content and student understanding as key traits of effective professional development (Desimone, 2009; Guskey, 2002; Ingvarson, Meiers, & Beavis, 2005). The STEM PCK construct focuses on student thinking about and effective strategies for teaching specific STEM content, therefore, it presents the opportunity to measure these aspects that are considered key to PD effectiveness. Second, in their review of science education research that was focused on PCK, Schneider and Plasman (2011) note a troubling trend of mid- and advanced career teachers either plateauing or declining in their PCK; the authors go on to attribute this as a failure of professional development to serve the needs of these later career teachers. The partnership’s decision to create a STEM PCK rubric will allow for creation of the tool that 1) measures key aspects of effective PD and 2) builds on the concepts of learning progression for teachers and a continuum of PCK across teachers’ careers (Schneider & Plasman, 2011).

*Effective Instructional practices in STEM.*

The construct definition and suite of measurement tools for effective instructional practices in STEM are outlined above. The importance of the instructional practices outcome to professional
development, however, is worth briefly exploring in this section. Kane (2012) asserts the importance of research into effective models for the systematic use of teacher practice data to improve PD design. By using the same data sources to provide a lens on individual teacher practice and on the aggregate effect of PD courses on teachers, the partnership will be able to strive for the “multifocal research lens” advocated for by Borko (2004). This focus on multiple levels of the system that ultimately impacts student outcomes will allow for the partnership to adopt additional effective professional development characteristics. The individual data lens incorporates feedback on and application to practice for the individual teacher, which is considered a trait of effective professional development (Desimone, 2009; Elmore & Burney, 1999; Ingvarson et al., 2005), while the aggregate professional development data lens allows for investigation of the effectiveness of the professional development course overall.

Construct Definition: Teacher Self-Efficacy.

The partnership’s construct definition for teacher self-efficacy is built from the work of Tschannen-Moran & Hoy (2001) and Klassen, Tze, Betts, and Gordon (2011): the belief that teachers hold about their own capabilities to bring about desired outcomes of student engagement, motivation, and learning in STEM.

Teacher self-efficacy was prioritized over other affective outcomes because it is likely the most important affective outcome for teacher professional growth. Thoonen et al., (2011b) characterize teacher self-efficacy as the “most important motivational factor for explaining teacher learning” (p.517). The relationship between teacher self-efficacy and other outcomes of relevance to effective professional development has been demonstrated. For example, teacher self-efficacy was found to have a positive correlation to instructional practices that align fairly well to the partnership’s effective STEM instructional practices (Lakshmanan, Heath, Perlmutter, & Elder, 2011).

Measurement of teacher self-efficacy. According to Klassen et al. (2011) teacher self-efficacy instruments must be designed with high fidelity to self-efficacy theory, which is primarily characterized by two item traits. First, items should be specific in nature providing clear cues to the teaching situation including context, subject, or targeted outcomes (Klassen et al., 2011; Tschannen-Moran & Hoy, 2001). Second, items should focus on measuring teacher’s current capabilities and avoid phrasing that emphasizes past performance (Klassen et al., 2011; Thoonen et al., 2011b).

The Teacher Sense of Self Efficacy (TSES) Survey was selected as the partnership’s measurement tool for the teacher self-efficacy construct for several reasons (Tschannen-Moran & Hoy, 2001). First, the TSES survey has three scales: efficacy for instructional practices, student engagement, and classroom management. These scales were considered in good alignment to both the partnership’s defined teacher self-efficacy construct and complimentary to other areas of the theory of change (i.e. focus on student engagement and instructional strategies that are specific to both higher-order thinking skills and conceptual knowledge). Second, the TSES survey meets the criteria for high fidelity to self-efficacy theory as outlined above with one exception. The current version of the TSES is subject general. In their review of research literature, Klassen et al. (2011) report “only modest empirical support” for a relationship between teacher efficacy and student outcomes. Sixty percent of the studies in their review
focused on teachers’ subject-general efficacy, therefore, the lack of a strong relationship could be due to the subject-general teacher self-efficacy instrumentation. Therefore, the partnership has amended the survey items on the instructional practices and student engagement scales to be subject-specific for the STEM disciplines. The classroom management scale was left subject general because it has been noted that classroom management skills are not specific to the subject being taught (Shulman, 1987).

Summary. Niemi et al. (2007) argue for assessment systems that produce data with validity at multiple levels contending that “coherence in an assessment system between inferences drawn system wide for accountability purposes and those drawn in the classroom…” (p. 197) are essential and as of yet unrealized. Through its common measurement system, this project seeks to meet this need for teacher practice assessments that can be used validly at the individual, programmatic, and system level. In addition, the STEM Common Measurement System will seek to characterize assessment for teaching, not of teaching with a strong emphasis on formative assessment of teacher practice, quick feedback, and data-driven PD.

School-level Support for Implementation of Effective STEM Practices.

Rationale. In the complex US education system, factors that either directly or indirectly impact students’ college and career readiness extend beyond teacher and student learning environments. The NRC’s 2011 report entitled Successful K-12 STEM Education emphasizes the importance of school-level variables. Some scholars also argue that the failure to be mindful of school system complexity and variables outside the classroom is the “…single best explanation for the historical failure of planned change in schools” (Leithwood, Aitken, & Jantzi, 2006, p.58). The partnership identified two school-level outcomes that have been demonstrated to have bearing on multiple levels of the theory of change: collective teacher efficacy and transformational leadership.

Construct definition: collective teacher efficacy.
The construct definition is based on the research literature’s conceptualization of collective teacher efficacy (Goddard, Hoy & Woolfolk Hoy, 2000; Klassen et al., 2011; Tschannen-Moran & Barr, 2004): the beliefs teachers hold about the shared capacity of their school’s staff to positively impact their students’ learning in STEM.

Even at the school-level, taking teachers beliefs into account is important because of the link between affective outcomes and behavioral outcomes (Tschannen-Moran & Barr, 2004). Collective teacher efficacy is a school-level affective variable that is a property of the staff of a school at a particular point in time (Tschannen-Moran & Barr, 2004). Multiple studies, included in a research review of unpublished research projects, have reported a positive relationship between collective teacher efficacy and school-level student achievement (Goddard et al., 2000; Sun & Leithwood, 2012; Tschannen-Moran & Barr, 2004). In addition to the relationship to student achievement, collective teacher efficacy has been shown to strongly predict teacher commitment (Ross & Gray, 2006), and influence the way school staff approach school-level improvement goals (Goddard et al., 2000). Finally, the collective teacher efficacy construct is complementary to the teacher self-efficacy construct, which creates complementary affective
outcomes between the supports and professional development components of the project’s theory of change.

**Measurement of collective teacher efficacy.** In their review of twelve years of collective teacher efficacy research, Klassen et al. (2011) characterize the field of collective teacher efficacy research as not yet reaching “maturity” and the measurement of this construct as an area of underdevelopment. In contrast to other available instruments, the Collective Teacher Belief Scale (CTBS) was characterized as the strongest reported in the literature and was hesitantly praised for a “closer congruence to collective efficacy theory” (Klassen et al., 2011). The CTBS survey contains two subscales: instructional strategies and classroom management/student discipline (Tschannen-Moran & Barr, 2004). This survey was found to have a two-factor structure that matched the survey scales, and high reliability for the whole survey and both sub-scales (.94-.97) (Tschannen-Moran & Barr, 2004). In addition, the CTBS Survey’s sub-scale structure and construct definitions is consistent with the PMSP construct definition and complimentary to the teacher self-efficacy construct. Therefore, the partnership has selected the CTBS Survey and made a similar modification as the one made to the TSES to create a STEM subject-specific instrument for the STEM Common Measurement System.

**Construct definition: transformational leadership**

The project’s construct definition for transformational leadership is drawn directly from Roberts (1985):

“This type of leadership offers a vision of what could be and gives a sense of purpose and meaning to those who would share that vision. It builds commitment, enthusiasm, and excitement. It creates hope in the future and a belief that the world is knowable, understandable, and manageable. The collective action that transforming leadership generates empowers those who participate in the process... In essence, transforming leadership is a leadership that facilitates the redefinition of a people’s mission and vision, a renewal of their commitment, and the restructuring of their systems for goal accomplishment.” (p. 1024).

Transformational school leaders actions are characterized by defining and sustaining a school vision, building structures for teacher collaboration, providing support for individual teacher’s professional growth, inspiring teacher motivation through excitement and high performance expectations, and engaging school communities in STEM education (Leithwood et al., 2006; Roberts, 1985; Sun & Leithwood, 2012; Thoonen et al., 2011b; Wiley, 2001).

School leadership is recognized as an important school-level variable in STEM education (Leithwood et al., 2006; Leithwood, Harris, & Hopkins, 2008; NRC, 2011). While many scholars acknowledge that leaders need to shift between leadership styles given the specific task at hand, these same scholars emphasize that it is transformational leadership that is associated with improvement in other important variables in complex school systems (Bass & Avolio, 1994; Leithwood, 1992; Sun & Leithwood, 2012). For example, Ross and Gray (2006) found transformational leadership to have direct effects on collective teacher efficacy and teacher commitment. In addition, increasingly strong evidence indicates that transformational leadership has indirect, but significant effects on student outcomes (Leithwood, Harris, & Hopkins, 2008; Leithwood & Sleegers, 2006; Sun & Leithwood, 2012). The importance of the transformational leadership outcome is further supported by evidence of the positive impact of transformational
leaders in a variety of organizations that extend beyond the education sector (Bass & Avolio, 1994; Leithwood, 1992).

Measurement of transformational leadership. The STEM Common Measurement System will use a school leadership survey created by school leadership scholars Leithwood et al. (2006). The survey authors iteratively refined the survey over more than a decade of projects and the survey has been used in both elementary and secondary schools (Jantzi & Leithwood, 1996; Leithwood et al., 2001; 2006). The survey has a strong theoretical foundation that is based off a literature review that revealed four characteristics of successful leaders; the survey scales are designed to measure these characteristics as well as key sub-characteristics that are organized under each main characteristic (Leithwood et al., 2008). The survey is also tightly aligned to the PMSP transformational leadership construct with subscales that target each of the five leadership characteristics that are specifically identified in the partnership’s construct.

The partnership will make three modifications to the survey to tailor it to this specific project’s needs. First, the survey will be specific to STEM by using the item stem “STEM leadership in this school…”. Second, this project emphasizes the distribution of STEM leadership in its STEM transformation schools through the appointment of a STEM Teacher on Special Assignment, therefore, for the purposes of the survey STEM leadership will be defined more broadly than the school’s administration. There is ample support for the importance of broadening the definition of leadership beyond school principals, especially when transformational leadership is the desired leadership style (Leithwood et al., 2008; NRC, 2011; Roberts, 1985). Finally, as a result of the broadening of the definition of STEM leadership in this project, the management-focused items on the school leadership survey will be removed from the survey. These items appear to be particularly tailored to administrator activities and are not closely aligned with the partnership’s definition of transformational leadership. The survey sub-scales are all reported to have satisfactory reliability with alphas between 0.75-0.93 (Leithwood et al., 2006). As a result of these reliability indices, it is expected that deleting the management sub-scales will not be problematic; however, the partnership will examine the psychometric properties of the amended survey once applied to the STEM context.

Discussion.

The STEM Common Measurement System described in this paper is the result of a conceptualization process that emphasized a strong theoretical foundation and a systems thinking approach. This prioritization of theory and systems thinking created a cohesive STEM Common Measurement System that is characterized by high levels of interconnections both within areas of the theory of change and across the theory of change (Figure 2). Though the interconnections between the constructs of the STEM Common Measurement System are hypothetical, they are well-grounded in the literature. As was argued in the introduction of this paper, common measurement systems must be designed to enable the investigation of questions about the interconnections among important variables of complex school systems. The completion of the conceptualization process reported in this paper allows for these interconnections to be fully articulated so that they can be empirically investigated in the future.
In examining the interconnections within and between the areas of the theory of change, connections to student achievement were also taken into consideration. Although typical student achievement measurement tools do not align with the rigor that is intentionally sought in the measurement of student outcomes in the STEM Common Measurement system (Katsinas & Bush, 2006), connections to student achievement were included for two reasons. First, district, schools, and teachers are still currently working in an education system that relies on student achievement metrics for accountability policies making these metrics relevant to these educators’ professional practice. Second, a large portion of the education research literature relies on student achievement metrics in examining relationship between variables and school or intervention effectiveness. Until the student achievement tests are adapted to come into better alignment with the new national standards for college and career readiness in mathematics and science, the education community has few other options besides these measures that are used on a broad scale across the K-12 education sector; therefore, connections to student achievement are included in this analysis.

**Interconnections within areas of the theory of change.**

In examining the interconnections within the areas of the theory of change, the two constructs contained in the supports area of the theory of change are related to each other and student achievement in the following ways. Transformational leadership is significantly and positively correlated with collective teacher efficacy (Ross & Gray, 2006) and is indirectly related to school-level student achievement (Leithwood & Sleegers, 2006; Wiley, 2001). Collective teacher efficacy also has a positive relationship with school-level student achievement in math (Goddard et al., 2000; Tschannen-Moran & Barr, 2004).

Regarding professional development and educator practice outcomes, scholars in the respective STEM disciplines have begun to demonstrate the intricate relationship between teacher’s PCK, instructional practices, and teacher self-efficacy beliefs. In their study of secondary mathematics teachers PCK, Baumert et al. (2010) found that teacher PCK had a “…substantial positive effect…on students’ learning gains,” but that the effects of PCK were moderated by instructional practices, which indicates a relationship between PCK and teacher instructional practices. A similar and significant relationship was found in a study of high school biology teachers when teacher PCK and reform-oriented instructional strategies were found to have a significant positive relationship (Park et al., 2011). There is also evidence of the link between teacher beliefs and teacher practices (Tschannen-Moran & Barr, 2004) and it is also reasonable to infer that teacher PCK may impact efficacy beliefs. This body of literature, when taken together, support the hypothesis that a focus on measuring the three professional development outcomes may be key in helping educators create effective learning environments in STEM.

Evidence of the interconnections between the three student outcomes can be drawn from psychological research, studies of specific programs, and longitudinal studies of national data sets. The affective student outcome of motivational resilience has been frequently linked to student achievement (Singh, Granville and Dika, 2002; Hughes, Luo, Kwok, & Loyd, 2008), science and mathematics grades (Skinner, Chi, & LGEAG, 2012; Crombie et al., 2005), general learning (Ryan & Deci, 2000; Wingfield, 2006) and students’ own perceptions of their learning (Skinner, Chi, & LGEAG, 2012) in the literature. The application of conceptual and higher-order cognitive skill student outcomes are less commonly investigated in the literature, however, some
preliminary evidence points towards important interconnections between these outcomes. For example, in the context of studying the development of scientific reasoning, Schauble (1996) concludes that conceptual knowledge and higher-order thinking skills “bootstrap each other so that appropriate knowledge supports the selection of appropriate experimentation strategies, and systemic and valid experimentation strategies support the development of more accurate and complete knowledge” p. 188. Additionally, both the Common Core State Standards in Mathematics and the Next Generational Science standards emphasize the importance and interconnection of student learning in the conceptual and practices domains for college and career readiness. Taken together, this evidence lends strong support for the partnership’s hypothesis that addressing student outcomes in all three domains (affective, conceptual, and practices) will result in STEM college and career readiness for students.

Interconnections between areas of the theory of change.

In the STEM Common Measurement System, there are a number of interconnections between the four main areas of the theory of change (school-level supports, professional development, educator practice, and student learning) that can be hypothesized based on previous findings in the literature (Figure 2). Beginning with the school-level supports area of the theory of change, transformational leadership positively impacts teacher self-efficacy beliefs through leader actions of identifying a clear school vision and goal setting (Ilies, Judge, & Wagner, 2006), holding high academic expectations (Hoy & Woolfolk, 1993), and providing individualized support based on teachers level of experience and unique needs (Walker & Slear, 2006). Transformational leadership was also found to have moderate, indirect effects on teacher instructional practices through leaders impacting their staff’s capacity, motivation, and commitment, and the working conditions in the school (Leithwood, Harris, & Hopkins, 2008). These indirect effects on instructional practices are most likely linked to the transformational leadership actions of providing intellectual stimulation and support, and high performance expectations.

In considering the interconnections between the professional development and teacher practices areas of the theory of change, it is expected that both teacher self-efficacy and PCK will be related to educators’ instructional practices. Teacher self-efficacy beliefs are an important outcome of professional development because of previously established positive relationships with teachers’ instructional practices (Lakshmana et al., 2011; Thoonen et al., 2011a). Additionally, the professional development outcome of PCK is related to the practices teachers implement in their classrooms. Studies have found a relationship between 1) mathematics teachers’ PCK and the cognitive demand of learning activities (Baumert et al., 2010), which links to IP#2 in the STEM Common Measurement System, 2) science teachers’ PCK and their implementation of reform oriented instructional practices (Park et al., 2011), which align to IP#1, 2, & 5, and the intricate role of teacher PCK in attending to student thinking in assessment data (Coffey, Hammer, Levin, & Grant, 2011), which aligns to IP#4.

Interconnections between the educator practices and student learning areas of the theory of change are hypothesized to be diverse and intricate involving multiple links between the educator practices and student learning. Both supportive teacher-student relationships and educator instructional practices have been shown to relate to students’ academic identity and motivational resilience. Furthermore, educator instructional practices have been related to
students’ development of higher-order cognitive skills. Finally, educators’ PCK and instructional practices have been shown to support students’ abilities to apply their conceptual knowledge.

As outlined earlier in the paper, the supportive teacher-student relationships construct of the STEM Common Measurement System is related to students’ academic identity, which in turn impacts student motivational resilience (Deci & Ryan, 2000; Skinner & Belmont, 1993; Skinner & Wellborn, 1997). For example, studies have found that supportive teacher-student relationships were positively linked to student perceptions of autonomy (Skinner, Wellborn, & Connell, 1990; Skinner, Chi, & LGEAG, 2010) and competence (Skinner, Chi, & LGEAG, 2010). Further supporting this interconnection in educational environments for very young students, Hughes et al. (2008) report that the quality of teacher-student relationships in first grade had a positive effect on student effortful engagement (similar to the academic engagement component of motivational resilience) in second grade. Students’ academic identities are also likely to be positively impacted by student-centered instructional practices (IP#1) that allow students to make choices (Skinner & Wellborn, 1997) about learning activities they engage in. For example, allowing students to pose hypotheses in science inquiry or choose their own solution in mathematical problem solving are examples of structured choices that create a student-centered learning environment.

Educator instructional practices have also been related to student learning of higher-order cognitive skills. Instruction that provides students with multiple and ideally diverse opportunities to practice higher-order cognitive skills (IP#3) is considered a requirement for the development of these skills (Aschbacher & Roth, 2002). Across elementary to high school settings, instructional practices that engage students in reflective metacognitive activities and carefully aligned learning goals have been attributed to increased in students’ higher-order cognitive skills (IP#4) (Bingham, Holbrook, & Meyers, 2010; Toth, Suthers, and Lesgold, 2002; Zohar & Nemet, 2002).

Finally, educators’ PCK and instructional practices are related to students’ abilities to apply their conceptual knowledge. Feltovich, Spiro, & Coulson (1993) outline “design principles for effective instruction” for what the authors call “advanced knowledge,” which is closely aligned to the application of conceptual knowledge construct. These principles emphasize the role of an educator’s PCK in anticipating prior beliefs or misconceptions students may hold about specific concepts and the use of strategies to challenge misconceptions. Feltovich et al. (1993) also outline design principles in clear alignment with some of the instructional practices including active engagement of students (IP #1), and the use of multiple representations so students have more than one opportunity to develop a deep understanding of concepts (IP#3). Additional studies further empirically support the interconnections between students’ ability to apply content knowledge and the use of student-centered instructional practices (IP#1) (Lawrenz, Huffman, & Robey, 2003) and multiple representations or opportunities (IP#3) (Prain & Waldrip, 2006). Finally, Rivet & Krajcik (2008) found preliminary evidence of a significant, positive correlation between students’ use of contextualizing features of instruction (which ties closely to IP#5) and their content learning as measured by both pre/post tests and project-based artifacts that were designed in part to measure outcomes complimentary to the application of conceptual knowledge construct.
When combined the description of the interconnections within and among the areas of the theory of change and the visual provided in figure 2, convey that that STEM Common Measurement System is attempting to make sense of, prioritize, and measure a complex network of variables of relevance to increasing the quality of STEM Education and obtain college and career readiness for all students. All interconnections are hypothetical at this point in time, however, the completion of the conceptualization of the STEM Common Measurement System allows for a step towards the next phase of the development process towards empirically testing the hypothesizes articulated above. In doing so, this study is addressing the scarcity of “systematic research” that investigates the connections between variables across complex school organizations as noted by Thoonen et al. (2011b).

**Conclusion.**

With the completion of the conceptualization stage of the STEM Common Measurement System described in this paper, the partnership is entering the next phase of the common measurement development process. This next stage entails instrument development, preliminary data collection, and internal validation investigations. While Vaidyanathan (2012) emphasized that the development and use for common measures systems is time intensive and difficult, she also contends “the benefits far outweigh the costs of developing such systems” (p. 56). The PMSP is organized around the premises of a collective impact partnership including the commitment to using common measures. Despite the relatively recent completion of the conceptualization phase and the vast amount of work still left to be completed in the next stage of the development of the STEM Common Measurement System, the partnership has already gained modest benefits from the common measures in the form of the common constructs, or outcomes, framework which provides clear expectations for programming, evaluation, and research, and allows for more intentional collaboration across partners.
References.


Figure 1 Blk & Wht

- Supports for Implementation of Effective STEM Practices
- Effective Teacher Professional Development Experiences
- Effective Practices for STEM Learning Environments
- Effective Student Learning Environments in STEM
Figure 2: STEM Common Measurement System

- **School-Level Supports**
- **Professional Development**
- **Educator Practices**
- **Student Learning**

**Key Components:**
- **Transformational Leadership**
- **Collective Teacher Efficacy**
- **Teacher Self-Efficacy**
- **Supportive Teacher-Student Relationships**
- **Instructional Practices**
- **Pedagogical Content Knowledge**
- **Academic Identity**
- **Motivational Resilience**
- **Higher-Order Cognitive Skills**
- **Application of Conceptual Knowledge**

**Educator Practices**

**Student Achievement**
Figure 1: PMSP Theory of change. The STEM Common Measurement System is designed and organized around the partnership’s theory of change.

Figure 2: Interconnections among variables in the STEM Common Measurement System. The theoretical model of interconnections between and among school-level support, professional development, educator practice and student learning variables is based on preliminary evidence from the literature. Connections with student achievement are also depicted.
Table 1: Common Measurement Committee’s highest ranked evaluation criteria. The below evaluation criteria are listed in order of importance based on the partnership’s highest ranked priorities and were used while evaluating currently available measurement instruments. A brief rational for each criterion is also provided.

<table>
<thead>
<tr>
<th>Ranked evaluation criteria</th>
<th>Rationale for each criteria’s importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link between defined construct &amp; instrument</td>
<td>In reviewing preexisting instruments, there was a potential pitfall of a lack of congruence between the PMSP prioritized construct and the construct the instrument was designed to measure. Great care was taken in to examine the alignment of PMSP construct to that of existing instruments.</td>
</tr>
<tr>
<td>Validity and reliability evidence</td>
<td>In order to ensure that existing tools would produce high-quality data, evidence of validity and reliability were key in evaluating instruments.</td>
</tr>
<tr>
<td>Potential for data use for formative/reflective purposes</td>
<td>The PMSP’s emphasis on the use of data to drive program improvement and educator decision-making made this criterion a high priority.</td>
</tr>
<tr>
<td>Sustainability of data collection tools</td>
<td>The goal of the committee was that tools in the common measurement system would be able to be used in perpetuity by partnering organizations after the appropriate initial training on use of the tools.</td>
</tr>
<tr>
<td>Relevance of data &amp;/or instrument across diverse contexts: formal and informal education settings</td>
<td>Given that the PMSP involves both formal and informal educational settings and that cross-organization learning is a key benefit to common measurement, the instruments and the data they produce must be applicable and useful in diverse settings.</td>
</tr>
<tr>
<td>Feasibility: cost, data processing and management</td>
<td>In order for instruments to be useful, they must also be feasible for organizations to use, therefore, the cost, data processing and management associated with each instrument were considered.</td>
</tr>
</tbody>
</table>
Table 2: Coverage of selected instructional practices instruments that will measure the STEM Common Measurement System’s Instructional Practice construct.

<table>
<thead>
<tr>
<th>Effective instructional practices in STEM Construct</th>
<th>Student perception of instructional practices survey</th>
<th>Artifact-based instrument</th>
<th>UTEACH Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP#1) Teachers facilitate active engagement of students in their learning.</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>a. Teachers assume the role of facilitator rather than authority figure</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>b. Students assume the role of active learners</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IP#2) Teachers emphasize deep content knowledge and higher-order cognitive skills by addressing learning goals in both areas.</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>IP#3) Teachers create and implement multiple and diverse opportunities for students to develop conceptual knowledge and cognitive skills.</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>IP#4) Teachers use frequent formative assessments (and summative assessments) to facilitate diagnostic teaching and learning.</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>a) Teachers and students are both stakeholders in the assessment process.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Teachers and students contribute to a classroom culture of assessment for learning.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IP#5) Teachers implement learning activities that students find to be relevant, important, worthwhile, and connected to their cultural and personal lives outside of the classroom.</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Acknowledgements.
The authors of this paper would like to gratefully acknowledge the other members of the PMSP Common Measures Committee for the contribution of each of their unique perspectives: Marci Benne, Steve Day, Joe Hansen, Brian Hawkins, Nancy Lapotin, Melissa Potter, Milan Sherman, and Susan Winner. We would also like to thank James Connell for his role as an external advisor to the committee. Finally, we thank Eric Banks for his assistance in designing our graphics so that our ideas could come alive in a visual form of communication, instead of only in words.