Menu of Measures: High-Quality Measures of Middle Years Math Student Outcomes

Julie Bruch, Jasmine Forde, Ijun Lai, and Veronica Sotelo Muñoz

Measuring the effects of math instruction and interventions is one way of learning how to improve students' math performance. Identifying and administering high-quality measures will help the field understand how to improve students' engagement with and understanding of math. This menu identifies a set of student-level measures and approaches to understand the extent that math instruction and interventions change what students know and how they feel about math, as well as how they use math outside of the classroom. Since these outcomes are direct consequences of how students are taught and supported in classrooms, these measures should be paired with measures of classroom context and opportunity structures.

Expert panel members

- **Steven Culpepper**, Associate Professor, Department of Statistics, University of Illinois at Urbana-Champaign
- **Russell Gersten**, Director of Instructional Research Group; Professor Emeritus, College of Education, University of Oregon
- **Chris Hulleman**, Associate Professor, Frank Batton School of Leadership and Public Policy and Department of Psychology, University of Virginia; Founder and Director, Motivate Lab
- **Jamaal Sharif Matthews**, Associate Professor, Educational Studies and the Combined Program in Education and Psychology, School of Education, University of Michigan
- **Ming-Te Wang**, Professor, School of Education and Department of Psychology, University of Pittsburgh

The authors originally developed this menu of measures, in consultation with a panel of experts, as a guide for the Bill & Melinda Gates Foundation's Middle Years Math grantees. The goal of the Middle Years Math grant portfolio is to develop, refine, and scale evidence-based solutions (programs, products, or practices) that demonstrate success in improving math outcomes for students in grades 3 through 9 who are Black, Latino, and/or experiencing poverty (the communities in focus for the grants). Grantees, in collaboration with their evaluation technical assistance providers, used the guide to select high-quality, common measures for their measurement and evaluation work. This menu of measures has been adapted for broader use.

In particular, this menu of measures is designed to be a resource to help educators, districts, researchers, funders, and organizations implementing programs learn about how to measure students' middle years math skills and knowledge, math perceptions and beliefs, and math use beyond the classroom in a valid, reliable, and accessible way. When used at multiple points over time, these measures can provide information about how a particular approach to teaching or supporting learning of math affects students. The menu is not a comprehensive list of every math measure available in the field. It is a curated list of measures developed in partnership with a panel of researchers in the field of math instruction and measurement. The menu includes measures that meet some or all of a set of preferred criteria developed by the panel (see the appendix at the end of the menu for full criteria).

One goal of math instruction and interventions is to have a longer-term impact on students beyond passing grades or scores on standardized tests. Figure 1 shows a conceptual framework of how classroom instruction, engagement, and supportive relationships could impact different aspects of math skills and knowledge, math perceptions and beliefs, and ultimately math use beyond the classroom. Each of these three key student outcome areas is defined below.

- **Math skills and knowledge.** This area encompasses the knowledge of how to flexibly and accurately solve mathematical problems. We focus on two domains of math skills and knowledge (shown in **Figure 1**):
  - **Math ideas and competencies.** Math ideas are overarching categories for describing problem types or structures. When approaching a problem, students need to first identify and understand the overarching idea. Students then apply one or more math competencies to solve the problem. Together, ideas and competencies contribute to students' ability to successfully solve problems (Organisation for Economic Co-operation and Development [OECD], 2019). See **Box 1** for examples of frameworks of math ideas and competencies.

Box 1. Math ideas and competencies

There are multiple frameworks of overarching ideas and competencies. Below are two examples.

Overarching ideas from the Organisation for Economic Co-operation and Development's Programme for International Student Assessment:

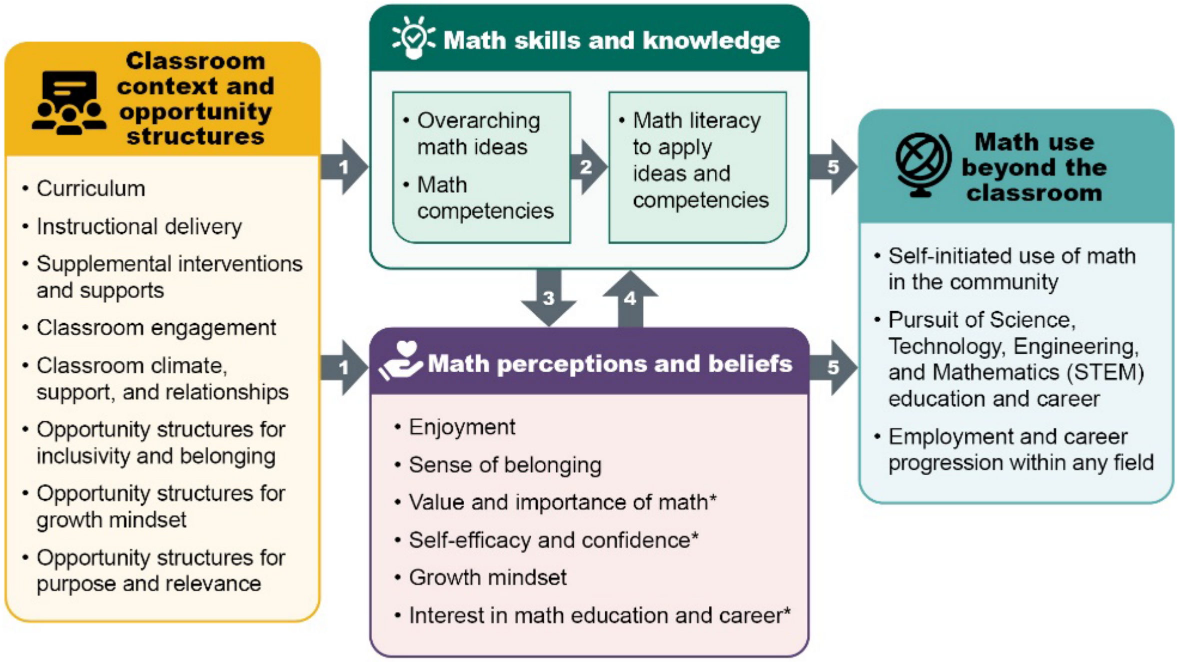
- Quantity
- Space and shape
- Change and relationships
- Uncertainty and data

Competencies from Common Core Standards for Mathematical Practice:

- Make sense of problems and persevere in solving them
- Reason abstractly and quantitatively
- Construct viable arguments and critique of the reasoning of others
- Model with mathematics
- Use appropriate tools strategically
- Attend to precision
- Look for and make use of structure
- Look for and express regularity in repeated reasoning
- Communication
- Use symbolic, formal, and technical language and operations
- Modeling
- Problem posing and solving
- Representation
- Use aids and tools

– **Math literacy.** Math literacy is the capacity to apply math ideas and competencies in a variety of real-world situations relevant to the needs of the student's life, often without explicit math terms (Ojose, 2011; OECD, 2019). As shown in Figure 1, students need to understand math ideas and develop competencies before being able to apply them elsewhere.

Figure 1. Conceptual framework for improving math outcomes and use



* Indicates constructs that might be especially important for supporting math use. When it is not possible to directly measure math use beyond the classroom, these could be useful constructs to measure in the short term.

- **Math perceptions and beliefs.** This area captures a range of student perceptions and beliefs related to learning and applying math (see Table 1 for definitions of specific constructs of math perceptions and beliefs covered in this menu). The panel believes these are malleable constructs that are shaped by the support and opportunities offered to students in school. This menu uses "perceptions and beliefs" rather than "attitudes" to convey the link between these constructs and the context in which students learn and live.

Table 1. Math perceptions and beliefs constructs

Construct	Definition
Sense of belonging	A student's feeling of being connected to the math learning content and activities and being an accepted, respected, and valued math learner.
Self-efficacy and confidence	A student's perception of their ability to successfully complete a specific math task, which can vary by task. Math confidence is a student's global assessment of their math ability.
Enjoyment	A student's interest, liking, or enjoyment of mathematics as well as affective reactions to mathematics.
Growth mindset	A student's belief that math abilities can be developed and improved through hard work, practice, good strategies, and input from others (as opposed to the belief that math abilities are fixed, stable, and unable to change).
Value and importance of math	A student's perception about the importance of doing well on mathematics and usefulness of mathematics for many aspects of daily life or fulfilling future goals.
Interest in math education and career	A student's interest in using and pursuing math beyond the classroom in their future education and career.

- **Math use beyond the classroom.** This area captures the ways that students continue using math outside of the classroom and in future education and career. This includes using math in the short term outside of classroom activities and assignments (for example, what a student does after school or on the weekend in the current school year). It also includes the longer-term use of math in future education and career (for example, in the subsequent school year or many years in the future). In both the short term and longer term, math use beyond the classroom is about the choices that students make and the behaviors they engage in to apply their skills, knowledge, perceptions, and beliefs in other settings.

While *math skills and knowledge* indicate students' capacity to use math beyond the classroom, and *math perceptions and beliefs* signal interest and motivation to use math beyond the classroom, the *math use beyond the classroom* area captures the actual behaviors of students.

Figure 1 shows our hypothesized pathways between math interventions and to the longer-term use of math. The numbered arrows in the figure correspond to the numbered steps in the pathway in the list below. We hypothesize that:

1. What happens in the classroom—including math instructional approaches and interpersonal supports—impact what students learn about math and how they feel about math. Box 2 provides suggestions for measuring classroom context and opportunity structures.
2. Students who understand math ideas and can solve problems can translate these competencies into math literacy to approach real-world problems that might not have an explicit math structure.
3. The capacity to solve problems with and without explicit math structure contributes to positive perceptions and beliefs about math.
4. There is a feedback loop in which positive perceptions and beliefs also motivate students to further build skills and knowledge.
5. A combination of math literacy and positive perceptions and beliefs about math lead to continued use of math beyond the classroom.

Box 2. Measuring classroom context and opportunity structures

While not covered in depth in this menu, measuring what happens in the classroom is important to understand the structures that enable or constrain students. Opportunity structures are aspects of the school environment – interpersonal, instructional, and institutional factors (Gray et al. 2018). These ways of teaching and interacting in the classroom impact students' agency, motivation, and learning. The student outcomes covered in this menu have the potential to change over time *if teachers provide inclusive and supportive opportunities for students.* In collaboration with the panelists, we have curated a list of reliable classroom observation instruments and surveys that could be used to measure classroom context and opportunity structures. These are suggestions and may not capture the full set of contextual variables in Figure 1. For links to the developers' websites and publications, see the **References** section.

Box 2 (continued)

Observation instruments

- **Belonging-Centered Instruction Observational Protocol (Matthews et al., 2021)**
- **Reformed Teaching Observation Protocol (RTOP) (Piburn & Sawada, n.d.)**
- **Mathematics Scan (M-Scan) (Berry et al., 2012)**
- **Mathematics Classroom Observation Protocol for Practices (MCOP2) (Gleason, 2015)**
- **UTeach Observation Protocol (UTOP) for Math and Science (UTeach Institute, n.d.)**
- **Teaching for Robust Understanding (TRU) Framework (Schoenfeld & Teaching for Robust Understanding Project, 2016)**
- **Mathematical Quality of Instruction (MQI) (Center for Education Policy Research, Harvard University, n.d.)**

Survey instrument

Copilot-Elevate (PERTS, n.d.)

This menu recommends instruments or approaches for measuring each of the three key areas: **math skills and knowledge**, **math perceptions and beliefs**, and **math use beyond the classroom**. For each instrument or approach, we provide a summary, references to key publications, information about reliability, relevant outcome constructs, and other considerations for use.

The menu includes the following measures:

Measure	Constructs	Type of measure	Relevant grades
Math skills and knowledge			
K1. ClearSight	• Overarching math ideas • Math competencies	Customizable math assessment	K-12
K2. Performance Series	• Overarching math ideas • Math competencies	Adaptive math assessment	K-12
K3. PSAT 8/9	• Overarching math ideas • Math competencies	Standardized math assessment	8, 9
K4. Trends in International Mathematics and Science Study (TIMSS)	• Overarching math ideas • Math competencies	Customizable math assessment	4, 8, 12
K5. Mathematics Assessment Collaborative/ Mathematics Assessment Resource Service (MAC/MARS)	• Overarching math ideas • Math competencies	Standardized math assessment	3-10
K6. Balanced Assessment in Mathematics (BAM)	• Overarching math ideas • Math competencies • Math literacy to apply ideas and competencies	Customizable math assessment	7, 8
K7. Star Math	• Overarching math ideas • Math competencies	Adaptive math assessment	1-12
K8. Iowa Algebra Aptitude Test	• Overarching math ideas • Math competencies	Standardized math assessment	7, 8
K9. Program for International Student Assessment (PISA)	• Overarching math ideas • Math competencies • Math literacy to apply ideas and competencies	Standardized math assessment	10, 11*
K10. Tests for Understanding Fractions	• Overarching math ideas • Math competencies	Standardized math assessment	4, 5
Math perceptions and beliefs			
P1. Attainment Value of Mathematics	• Value and importance of math	Written survey	5-9
P2. Beliefs, Engagement, and Attitude of Math Motivation Scale	• Value and importance of math • Self-efficacy and confidence • Enjoyment	Written survey	2, 3
P3. Expectancy-Cost-Value Scale	• Value and importance of math • Self-efficacy and confidence	Written survey	1-PS
P4. Indiana Mathematics Belief Scale	• Self-efficacy and confidence • Growth mindset	Written survey	PS
P5. Math and Me Survey	• Value and importance of math • Self-efficacy and confidence • Enjoyment	Written survey	3-6

Measure	Constructs	Type of measure	Relevant grades
P6. Math and Science Engagement Scales	• Sense of belonging • Enjoyment	Written survey	5-12
P7. Math Identity Measures	• Self-efficacy and confidence • Enjoyment	Written survey	PS
P8. Mathematical Mindset	• Value and importance of math • Self-efficacy and confidence • Enjoyment and Growth mindset	Written survey	6-8
P9. Mathematics Attitude Inventory	• Value and importance of math • Self-efficacy and confidence • Enjoyment	Written survey	7-12
P10. Mathematics Self-Efficacy Scale	• Self-efficacy and confidence	Written survey	9, 10
P11. Patterns of Adaptive Learning Scales (PALS)	• Self-efficacy and confidence	Written survey	5-9
P12. Sense of Connectedness to My Mathematics Classroom	• Sense of belonging	Written survey	5-9
P13. Perceived Utility Value	• Value and importance of math	Written survey	PS
P14. Investigative Scale of Mapping Vocational Challenges	• Interest in math education and career	Written survey	10-12
P15. Interest in a Future STEM Career (IFSTEMC)	• Interest in math education and career	Written survey	9-12
P16. Career Interest Questionnaire	• Interest in math education and career	Written survey	6-8
Math use beyond the classroom			
U1. Self-initiated use of math in the community	• Self-initiated use of math in the community	Custom survey	n.a.
U2. Enrollment and completion of math and STEM coursework	• Pursuit of STEM education and career	Administrative K-12 data or custom survey	6-12
U3. Involvement in math and STEM internships, career technical education (CTE), and other occupational training	• Pursuit of STEM education and career	Administrative K-12 data or custom survey	9-12
U4. Involvement in math and STEM extracurricular activities	• Pursuit of STEM education and career	Administrative K-12 data or custom survey	K-12
U5. Enrollment in math and STEM postsecondary majors	• Pursuit of STEM education and career	Administrative PS data or custom survey	PS
U6. Employment in math or STEM occupations	• Pursuit of STEM education and career	Custom survey	n.a.
U7. Employment and career progression within any field	• Employment and career progression within any field	Administrative labor market data or custom survey	n.a.

Note: Measures identified for math use beyond the classroom are recommended approaches for measurement, either using available administrative data sources or with a custom survey. We did not identify off-the-shelf instruments available in this area. This area needs further measurement development to facilitate reliable, consistent measurement across settings and studies.

* Program for International Student Assessment (PISA) is administered to students who are age 15. Grades 10 and 11 reflect the typical grades a 15-year-old student is enrolled in school in the United States.

n.a. = not applicable; PS = postsecondary.

Assessment of measures

This section describes the criteria used to assess the measures in the menu of math skills and knowledge as well as math perceptions and beliefs. This menu of measures highlights measures that meet all or some of the criteria, and it documents which criteria each measure meets and does not meet. Measures for math use beyond the classroom were not assessed because we did not identify existing instruments in this area. Instead, the menu includes suggestions for how to use administrative data or custom surveys to measure constructs in this area.

Math knowledge and skills and math perceptions and beliefs measures are assessed based on:

- **Reliability.** What is known about whether scores are consistent across individuals, context, time, items, or raters?
- **Validity.** Do the scores represent what they intend to measure?
- **Cultural responsiveness.** Is the measure reliable and valid for use in classrooms that include students who are Black, Latino, and/or experiencing poverty? For instruments that students complete, are the items relevant for students' sociocultural values and experiences?
- **Linguistic accessibility.** For instruments that students complete, are there appropriate adaptations, translations, and resources available for students who need assistance with written or oral English language?
- **Feasibility to administer.** Are instruments feasible to administer and score, and are there few barriers in terms of cost and licensing? Can the instruments be administered virtually, as there might be an intermittent need during the COVID-19 pandemic?

All measures in the menu are reliable, and most are specific to math. Some math perceptions and beliefs measures (identified in Table 2) are not specific to math but cover a gap in the area and could be adapted to be specific to math. Reliability must be greater than a Cronbach's alpha value of 0.70 for standardized tests that have multiple items related to the same construct; surveys with self-reported measures of perceptions, beliefs, and behaviors related to the same construct; and measures that have more than one rater assigning ratings or scores.

For some math skills and knowledge assessments, developers assert that assessments are reliable, but sufficient evidence is not always available to substantiate the authors' claims. For some survey measures, the overall instrument has shown to be reliable, but some individual subscales might not. These situations are indicated in the measure description, and users should not administer the subscales in isolation.

When selecting measures, users can consider the extent to which measures meet the preferred criteria. Preferred criteria (detailed in Steps 2 and 3 of the appendix) include reliability and validity with students who are Black, Latino, and/or experiencing poverty; availability of the instruments themselves and guidance to consistently interpret results; the linguistic accessibility of the measures; and the extent to which the measures are feasible to use and appropriate to use in a virtual setting. There are also content-specific criteria for mathematics knowledge (the extent to which the instrument measures deep knowledge, specific math proficiency strands, and cognitively complex concepts) and perceptions and beliefs (alignment with key constructs and the extent to which they measure a temporary state of thinking rather than a stable trait).

Tables 2 and 3 show which of the preferred criteria each measure meets. Each column in the table lists an assessed measure, and each row describes a relevant criterion that was examined; the intersecting cell presents the panel's assessment of whether each measure meets the criterion of interest. The information used to assess measures in this menu was documented in 2020 and spring of 2021, and updated information might have been published after this time period.

When using this resource, readers should also consider a set of context-specific questions to determine whether the recommended measures are a good fit for the local context (see Box 3).

Box 3. Context-specific questions

Consider context-specific questions to determine fit of measures

Questions to assess validity:

- Is the measure designed to capture an outcome in the program's theory of change?
- Is the measure aligned with an outcome expected to change?
- Is there evidence that the measure is predictive of expected longer-term outcomes?
- Is the measure designed to be a formative or summative assessment?
- Does the measure capture growth or proficiency?

Questions to assess context-specific linguistic accessibility and cultural relevance:

- Is the wording and content relevant and appropriate for the students' level of literacy and cognition? If not, can the measure be adapted to be developmentally appropriate?
- Are there languages other than English that the participating students speak? If so, is the measure linguistically accessible in all relevant languages and dialects?
- Is the interpretation of measure items relevant to students' sociocultural values and experiences and does not presume White middle-class values?

Consider context-specific questions to determine usability of measures

Is the measure feasible to administer given…

- Its training requirements?
- Its scoring requirements?
- The cost?
- The time required to administer?

Table 2. Math skills and knowledge

Criterion	K1. ClearSight	K2. Performance Series	K3. PSAT 8/9	K4. Trends in International Mathematics and Science	K5. Mathematics Assessment Collaborative/ Mathematics Assessment Resource Service MAC/MARS	K6. Balanced Assessment in Mathematics (BAM)	K7. Star Math	K8. Iowa Algebra Aptitude Test	K9. Program for International Student Assessment (PISA)	K10. Tests for Understanding Fractions
Grades tested in prior research[a]	K-12	K-12	8, 9	4, 8, 12	3-10	7, 8	1-12	7, 8	10, 11‡	4, 5
Item availability	I	I	F	S	F	F	F, I†	F	S	F
Demonstrated reliability and validity with communities in focus	No	No	No	Yes	No	No	Yes	No	No	No
Measure is available for use with-out restrictions on access	No	No	No	Yes	No	No	No	No	No	Yes
Guidance available to help users consistently interpret results	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Measure is linguistically accessible	No	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes
Measure can be implemented in virtual setting	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes
Assesses deep knowledge of math	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Assesses one of the math proficiency strands[b]	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Assesses cognitively complex concepts	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: No = The panelists concluded evidence was not sufficient to determine whether the measure meets the criterion. This could mean either that the studies reviewed did not report the necessary evidence OR that the instrument is not available for assessment by the panel. Yes = The panelists concluded evidence was sufficient to determine whether the measure meets the criterion.

Item availability indicates the ability of programs to use individual items or subscales from a larger assessment. I = items; programs may be able to select individual items that are most applicable. S = subscale; programs may be able to use scores from individual subscales that are most applicable; F = full measure; programs must use a comprehensive score and cannot customize the assessment.

† Star Math is marked as both whole test and items because there are two versions of the assessment. One version is narrower and does not allow for customization, while the custom version allows programs to customize the tests.

‡ PISA is an international standardized test that students take at age 15. Grades are estimated to reflect the typical grades a 15-year-old student is enrolled in school in the United States.

[a] Prior research refers to the key publication reviewed by panelists to assess each measure. The grade range indicates the grades included in each study. For standardized tests, we list grades with available tests.

[b] Math proficiency strands include procedural fluency (ability to carry out procedures flexibly, accurately, efficiently, and appropriately), conceptual understanding (fundamental understanding of mathematical ideas and students' ability to integrate new ideas into their understanding by connecting with previous knowledge), strategic competence (ability to formulate, represent, and solve mathematical problems), and adaptive reasoning (capacity for logical thought, reflection, explanation, and justification).

Table 3. Math perceptions and beliefs

Criterion	P1. Attainment Value of Mathematics	P2. Beliefs, Engagement, and Attitude of Math Motivation Scale	P3. Expectancy-Cost-Value Scale	P4. Indiana Mathematics Belief Scale	P5. Math and Me Survey	P6. Math and Science Engagement Scales	P7. Math Identity Measures	P8. Mathematical Mindset	P9. Mathematics Attitude Inventory	P10. Mathematics Self-Efficacy Scale	P11. Patterns of Adaptive Learning Scales (PALS)	P12. Sense of Connectedness to My Mathematics Classroom	P13. Perceived Utility Value*	P14. Investigative Scale of Mapping Vocational Challenges	P15. Interest in a Future STEM Career (IFSTEMC)	P16. Career Interest Questionnaire
Grades tested in prior research[a]	5-9	2, 3	1-PS	PS	3-6	5-12	PS	6-8	7-12	9, 10	5-9	5-9	PS	10-12	9-12	6-8
Demonstrated reliability and validity with community in focus	Yes	No	No	No	No	Yes	No	No	No	No	No	Yes	No	No	No	No
Measure is available for use without restrictions on access	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Guidance available to help users consistently interpret results	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes	No	Yes	No	No
Measure is linguistically accessible	No	No	No	No	No	Yes	No	No	No	No	No	Yes	No	No	No	No
Measure can be implemented in virtual setting	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Assesses at least one key math perceptions and beliefs construct[b]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Assesses state-based instead of trait-based characteristic[c]	No	Yes	Yes †	Yes †	Yes †	Yes	No	No	Yes †	Yes†	Yes †	Yes	Yes †	Yes	Yes †	Yes

Note: No = The panelists concluded evidence was not sufficient to determine whether the measure meets the criterion. This could mean either that the studies reviewed did not report the necessary evidence OR that the instrument is not available for assessment by the panel. Yes = The panelists concluded evidence was sufficient to determine whether the measure meets the criterion.

* Measure not specific to math but covers a gap in the area and could be adapted to be specific to math.

[a] Prior research refers to the key publication reviewed by panelists to assess each measure. The grade range indicates the grades included in each study.

[b] The panel identified a set of key math perceptions and beliefs constructs that are distinct from one another and can be measured. The list of constructs is in Table 1.

[c] State-based characteristic refers to a temporary way of thinking, feeling, behaving, and so on, while a trait-based characteristic is considered to be a stable and enduring characteristic. There is emerging research on the extent that different perceptions and beliefs are state- and trait-based; measures that are mostly state-based are coded with Yes, potentially state-based are coded with Yes†, and mostly trait-based are coded with No.

PS = postsecondary.

Description of measures

This section describes each measure included in the menu, including additional considerations for implementation and interpretation. The information in this section can help users select a measure from among several options they are considering. It can also inform plans for implementing and interpreting specific measures.

Measures of math skills and knowledge

Being able to solve math problems requires an understanding of overarching math ideas as well as more specific competencies. Measures of math skills and knowledge do not typically distinguish between ideas and competencies—students use both when successfully solving problems on assessments. The measures included in this section were selected because they assess the extent to which students have deep math knowledge instead of rote memorization of facts; develop skills in a math proficiency strand such as procedural fluency and adaptive reasoning; and understand cognitively complex concepts (see the appendix for a detailed description of math proficiency strands).

Although there are many ways to measure math skills and knowledge (e.g., performance assessments, open responses, student portfolios), this menu focuses on standardized assessments that can be used in consistent ways across different contexts and that are publicly available or easily accessible. The panel assumes that standardized tests have face validity and are reliable. Technical manuals can be requested from developers if users need additional information. For links to the developers' websites and full citations for all publications listed in this section, see the **References** section.

K1. **ClearSight (formerly known as AIRways)** is a web-based is a web-based math assessment with questions aligned to grade-level state standards and is customizable for grades kindergarten through high school. Each benchmark assessment has 6 to 24 items, though no information is available about the amount of time allotted for each assessment.

- Key publication: No publicly available evidence; documentation may be available through direct request to the measure developer. Information about the assessment is available from the developer, Voyager Sopris Learning.
- Considerations for reliability: The developer asserts that measures are reliable and valid. Although this test has been used with students in the communities in focus, reliability is not reported for racial or ethnic subgroups. Additional testing is needed with the communities in focus.
- Relationship to math skills and knowledge: Items appear to assess deep math knowledge, math proficiency strands, and cognitively complex concepts. Specific examples and verification of content can be requested from the developer.
- Other considerations for use: Brochure states that the developer can create assessments for students from kindergarten through high school, and as many interim assessments as the consumer would like. The assessment is only available through purchase.

K2. The **Performance Series** is a web-based computer-adaptive diagnostic assessment to predict student scores on state tests and can track student growth over time. The developer, Scantron, offers both online and paper/hard-copy assessments. Item bank includes questions for students in kindergarten through high school. The average time to administer the test is approximately 20 minutes.

- Key publication: No publicly available evidence; documentation may be available through direct request to the measure developer. Information about the assessment is available from the developer, Scantron.
- Considerations for reliability: Reported reliability from the developer is 0.91. Although this test has been used with students in the communities in focus, reliability is not reported for specific racial or ethnic subgroups. Additional testing is needed with the communities in focus.
- Relationship to math skills and knowledge: The public does not have access to math item banks. Test developer staff can work with grantees to determine which types of items to use. Items appear to assess deep math knowledge, math proficiency strands, and cognitively complex concepts.
- Other considerations for use: Participating schools need to administer assessment and share data, or grantee needs to purchase for use ($1,200 per school per year). The mathematics assessments portion is available in both English and Spanish. The Spanish version is developed through a transadaption process, ensuring that test items have the same meaning, text complexity, and other measurement characteristics as the English version (Scantron, 2014).

K3. The **PSAT 8/9** is a standardized math test with an emphasis on problem solving, modeling, using tools strategically, and using algebraic structure. Most questions are multiple choice, but some require students to write in an answer instead of selecting an answer. This is a 60-minute test with 38 questions and tasks.

- Key publication: No publicly available evidence; documentation may be available through direct request to the measure developer. Information about the assessment is available from the developer, College Board.
- Considerations for reliability: The developer asserts that measures are reliable and valid. Although this test has been used with students in the communities in focus, no public information is available about reliability and validity for these communities. Additional testing is needed with the communities in focus.
- Relationship to math skills and knowledge: This test intentionally has a wide range of mathematical tasks, which are heavily based on the National Council of Teachers of Mathematics standards. Items appear to assess deep math knowledge, math proficiency strands, and cognitively complex concepts.
- Other considerations for use: The cost is $17 per student.

K4. The **Trends in International Mathematics and Science Study (TIMSS)** is an international test of mathematics and science achievement for students in grades 4, 8, and 12. These items are all available for use via download; grantees can select which items and number of items they would like to use.

- Key publication: Reliability and validity tests are analyzed after every test release, so there is more than one key publication about this test. For example, Mullis et al. (2016) evaluate the reliability and validity of the 2015 TIMSS.
- Considerations for reliability: No public information is available about reliability and validity for subgroups of students defined by race, ethnicity, or family income within each country, but the TIMSS has been used in more than 60 nations and in ethnically diverse school districts such as Chicago, Jersey City, and Dade County.
- Relationship to math skills and knowledge: This measure is an assessment of mathematical reasoning and problem solving. Items appear to assess deep math knowledge, math proficiency strands, and cognitively complex concepts.
- Other considerations for use: After each round of testing, TIMSS releases half of the questions for free public use, which can be found on the National Center for Education Statistics' website and are available for download. Because only half of the questions are released, the set of publicly available items may not be comparable across years. Each item has been tested for reliability and validity.

K5. The **Mathematics Assessment Collaborative/ Mathematics Assessment Resource Service (MAC/ MARS)** is a 40-minute standardized test that includes a set of detailed math performance tasks. Tests are available for use in grades 3 to 10. A new test for each grade is created each year.

- Key publication: No publicly available evidence; documentation may be available through direct request to the measure developer. Information about the assessment is available from the developer, MARS.
- Considerations for reliability: The developer asserts that measures are reliable and valid. No public information is available about reliability and validity for students in the communities in focus. Additional testing is needed with students in the communities in focus.
- Relationship to math skills and knowledge: This test assesses a wide range of mathematical tasks. Items appear to assess deep math knowledge, math proficiency strands, and cognitively complex concepts.
- Other considerations for use: This assessment costs about $2 per student test, plus local scoring costs including scoring training (one or more sessions) and scoring time (10 to 20 student papers per scorer hour). For outside scoring and reporting services, costs are about $10 per student test. Annual membership fees are tiered based on the number of students served: $1,260 for entities serving fewer than 250 students to $31,500 for entities serving more than 100,000 students.

K6. The **Balanced Assessment in Mathematics (BAM)** is a collection of K-12 assessment tasks developed by the Harvard Graduate School of Education in collaboration with teams from MARS. The Middle School collection contains 24 assessment tasks that are appropriate for use with middle years students in grades 6 to 9. A scoring system was also designed that evaluates four dimensions of mathematical thinking on a 4-point scale.

- Key publication: Stancavage et al. (2009).
- Considerations for reliability: Stancavage et al. (2009) reported reliability for samples of 7th-grade and 8th-grade students. Reliability is not reported for racial or ethnic subgroups. Additional testing is needed with the communities in focus.
- Relationship to math skills and knowledge: Items appear to assess deep math knowledge, math literacy, math proficiency strands, and cognitively complex concepts. In addition, the items assess math literacy.
- Other considerations for use: Tasks can be viewed and purchased at the Balanced Assessment (2015) website.

K7. **Star Math** is a web-based adaptive multiple-choice assessment for use in grades 1 to 12. No public information is available about the average number of items; the average time to administer is 24 minutes.

- Key publication: White paper on Star Math (Renaissance Learning, Inc., 2020). Information about the assessment is available from the developer, Renaissance.
- Considerations for reliability: No public information is available about reliability and validity for students in the communities in focus. Additional testing is needed with students in the communities in focus.
- Relationship to math skills and knowledge: Star Math has a large math item bank. Items appear to assess deep math knowledge and cognitively complex concepts. Not enough information is available to determine if the measure assesses math proficiency strands.
- Other considerations for use: Cost is estimated at $4.95 per student for the 2020–2021 year. However, there is also a one-time fee of $1,599 per school for first-time users and an annual $750 per-school fee for hosting Renaissance software. Star Math is also available in Spanish.

K8. The **Iowa Algebra Aptitude Test** is a 50-minute exam for students in grades 7 and 8 and assesses students' readiness for Algebra I. It is aligned with the National Council of Teachers of Mathematics standards and is split into four parts: pre-algebraic number skills and concepts, interpreting mathematical information, representing relationships, and using symbols.

- Key publication: No publicly available evidence; documentation may be available through direct request to the measure developer. Information about the assessment is available from the developer, Riverside Insights.
- Considerations for reliability: The developer asserts that measures are reliable and valid. No public information is available about reliability and validity for students in the communities in focus. Additional testing is needed with students in the communities in focus.
- Relationship to math skills and knowledge: Items appear to assess deep math knowledge, math proficiency strands, and cognitively complex concepts.
- Other considerations for use: Cost is estimated to be about $5 per student, but it is not clear if other fees are associated with use. Scoring and reporting services are available for an

additional fee. Additionally, this assessment is designed for use in 7th and 8th grades, so it is not clear how useful it is for other grades.

K9. The **Program for International Student Assessment (PISA)** is an internationally standardized instrument developed by the Organisation for Economic Corporation and Development (OECD). The instrument measures reading, math, and science literacy in 15-year-old students. Math literacy is assessed using open-ended responses, closed responses, and multiple-choice responses. Items assessing math literacy require students to apply math ideas and competencies to real-world situations. The instrument is composed of seven 30-minute timed sections, with the entire instrument taking 210 minutes.

- Key publication: No publicly available evidence; documentation may be available through direct request to the measure developer. Information about the assessment is available from the developer (OECD, 1997).
- Considerations for reliability: The developer asserts that measures are reliable and valid. No public information is available about reliability and validity for students in the communities in focus. Additional testing is needed with students in the communities in focus.
- Relationship to math skills and knowledge: Items appear to assess deep math knowledge, math proficiency strands, and cognitively complex concepts. In addition, the items assess math literacy.
- Other considerations for use: The 2015 and 2018 PISAs were available in both computer- and paper-based formats and in over 90 languages. More information about the instrument is available from the developer (OECD, 2003).

K10. There are two different **Tests for Understanding Fractions,** one for 4th grade and one for 5th grade. The 4th-grade test includes 26 multiple-choice questions and uses items from the National Assessment of Educational Progress (NAEP), Illustrative Mathematics, and a fraction item bank. The 5th-grade test includes 18 multiple-choice questions and draws from items used in the NAEP and the Partnership for Assessment of Readiness for College and Careers.

- Key publication: Jayanthi et al. (2017).
- Considerations for reliability: Jayanthi et al. (2017) reported reliability for samples of 4th-grade and 5th-grade students. The sample was composed of about 30 percent Black students, 15 percent Latino students, 9 percent students with disabilities, and 11 percent English language learners. Reliability is not reported for racial or ethnic subgroups. Additional testing is needed with the communities in focus.
- Relationship to math skills and knowledge: Items assess deep math knowledge, math proficiency strands, and cognitively complex concepts.
- Other considerations for use: The instrument is available for free but requires permission from the authors. A Spanish version of the test is not currently available, but the panel believes that a translation would be easy to create. By request, the developer, Instructional Research Group, can draw from a large normative sample to help with interpretation.

Measures of math perceptions and beliefs

Measures in this section assess students' perceptions and beliefs about math. Table 4 provides information about the specific constructs included in each measure. Although some measures have individual items related to a specific construct, the panel strongly advises against pulling individual items from scales or subscales. It is unclear how individual items perform in isolation. If users extract individual items or subscales from instruments or combine them with their own items, they should psychometrically assess the combined instrument. For most of these measures, the items can be found in the key publication. For full citations for all publications listed in this section, see the **References** section.

Table 4. Math perceptions and beliefs constructs and structure

Criterion	P1. Attainment Value of Mathematics	P2. Beliefs, Engagement, and Attitude of Math Motivation Scale	P3. Expectancy-Cost-Value Scale	P4. Indiana Mathematics Belief Scale	P5. Math and Me Survey	P6. Math and Science Engagement Scales	P7. Math Identity Measures	P8. Mathematical Mindset	P9. Mathematics Attitude Inventory	P10. Mathematics Self-Efficacy Scale	P11. Patterns of Adaptive Learning Scales (PALS)*	P12. Sense of Connectedness to My Mathematics Classroom	P13. Perceived Utility Value	P14. Investigative Scale of Mapping Vocational Challenges	P15. Interest in a Future STEM Career (IFSTEMC)	P16. Career Interest Questionnaire
Sense of belonging						◐						●				
Self-efficacy/confidence		○	◐	◐	◐		○	○	◐	●	◐					
Enjoyment		○			◐	●	○	○	◐							
Growth mindset				◐				○								
Value and importance of math	●	○	◐		◐			○	◐				●			
Interest in math education and career														●	●	●

● = full measure needs to be used to assess construct.

◐ = the instrument includes a subscale related to the construct.

○ = there are individual items related to the construct.

P1. The **Attainment Value of Mathematics** is an 8-item questionnaire on motivation in mathematics and positive math identity that uses a 6-point Likert scale. The measure has two subscales: importance of mathematics as self-defining and generalized value of mathematics. The measure is designed for grades 5 to 9. No information is available on time to administer.

- Key publication: Matthews (2018).
- Considerations for reliability: The instrument meets the required reliability threshold overall. All subscales meet the required reliability threshold based on prior research. Reliability was established with a sample in which 56 percent of students were Black, 27 percent were Latino, and 85 percent were eligible for free or reduced-price lunch. Reliability is not reported for racial or ethnic subgroups. Additional testing is needed with the communities in focus.
- Relationship to math perceptions and beliefs: The full measure assesses the value and importance of math.
- Other considerations for use: This instrument is available for free and could be administered virtually. It measures self-perceptions based on abilities that can change and grow over time. The scale was developed as part of an ongoing longitudinal study on motivation in mathematics.

P2. The **Beliefs, Engagement, and Attitude of Math Motivation Scale** is a 10-item survey that measures beliefs, engagement, and attitudes toward math on a dichotomous (0 or 1) scale. This assessment is designed for grades 2 and 3. Time to administer is three minutes.

- Key publication: Orosco (2016).
- Considerations for reliability: The instrument meets the required reliability threshold overall. Item-level reliability also meets the required reliability threshold based on prior research. Reliability was established with a sample in which more than 50 percent of students were eligible for free or reduced-price lunch and 30 percent were Latino. Reliability is not reported for racial or ethnic subgroups. Additional testing is needed with the communities in focus. Prior research found the instrument to be psychometrically valid at the elementary level (grades 2 and 3). No evidence is available at the secondary level.
- Relationship to math perceptions and beliefs: Individual items assess self-efficacy/confidence, enjoyment, and the value and importance of math.
- Other considerations for use: Measure is available without restrictions. Orosco (2016) included a sample of 2nd- and 3rd-grade students, so the instruments might need to be adapted for middle grades. The measure could be administered virtually.

P3. The **Expectancy-Cost-Value Scale** is a 10-item student survey instrument based on the expectancy-value theory of motivation and uses a 6-point Likert scale. The survey is designed to measure motivation in math and science and can be used with middle school students in grades 6 to 9. The instrument has three subscales: expectancy, cost, and value. The value scale includes utility value, attainment value, and interest/enjoyment value subscales and components. Time to administer is less than five minutes.

- Key publications: Kosovich et al. (2014); Lauermann et al. (2017); Simpkins et al. (2006), Wang (2012).
- Considerations for reliability: The instrument meets the required reliability threshold overall. All subscales meet the required reliability threshold based on prior research. This scale has been widely used and validated with many socioeconomically and racially diverse samples ranging from 3rd to 12th grade since the 1990s.
- Relationship to math perceptions and beliefs: The instrument includes subscales related to self-efficacy/ confidence and the value and importance of math.
- Other considerations for use: The instrument is available for use for free. It is a survey that could be administered virtually. It assesses the utility value, attainment value, interest/enjoyment value, and cost value for students (decision process weighing school work against other activities they could participate in, assessment of effort required for school work, and emotional cost).

P4. The **Indiana Mathematics Belief Scale** is a 35-item scale that students complete to measure their beliefs on constructs using a 5-point Likert scale. The measure assesses self-efficacy, identity, growth mindset, and related beliefs for secondary school and college levels. The instrument includes five subscales: effort, usefulness, difficult problems, understanding, and steps. The scale takes about 15 minutes to administer.

- Key publications: Ayebo & Mrutu (2019); Kloosterman & Stage (1992).
- Considerations for reliability: The instrument meets the required reliability threshold overall. Three of the five subscales meet the required reliability threshold. Two subscales (steps and difficult problems) do not meet the required threshold. Reliability was established with a sample of undergraduate college students, and no demographic information was presented. Reliability is not reported for racial or ethnic subgroups. Additional testing is needed with the communities in focus and targeted age groups.
- Relationship to math perceptions and beliefs: The instrument includes subscales related to self-efficacy/confidence and growth mindset.
- Other considerations for use: The instrument is available for free and could be administered virtually.

P5. The **Math and Me Survey** is a 27-item scale students in grades 3 to 6 complete to measure their beliefs on three constructs using a 5-point Likert scale. Subscales include mathematical self-perceptions, enjoyment of mathematics, and perceived usefulness of mathematics. No information is available on time to administer.

- Key publication: Adelson & McCoach (2011).
- Considerations for reliability: The instrument meets the required reliability threshold overall. All subscales meet the required reliability threshold based on prior research. Reliability was established with a sample of students in grades 3 to 6 of whom 34 percent were non-White. Reliability is not reported for racial or ethnic subgroups. Additional testing is needed with the communities in focus.

- Relationship to math perceptions and beliefs: The instrument includes subscales related to self-efficacy/ confidence, enjoyment, and the value and importance of math.
- Other considerations for use: The instrument is available for free and could be administered virtually.

P6. The **Math and Science Engagement Scales** comprise of 33 items (17 positively worded and 16 negatively worded) that are on a 5-point Likert scale. Subscales include behavioral engagement, emotional engagement, cognitive engagement, and social engagement. The measure was tested on students in grades 5 to 12. The full scale takes about 7 to 10 minutes to administer.

- Key publications: Fredricks et al. (2016); Wang et al. (2016).
- Considerations for reliability: The instrument meets the required reliability threshold overall. All subscales meet the required reliability threshold based on prior research. Reliability was established with a sample of students in grades 6 to 12. This measure has been validated with racially/ethnically and socioeconomically diverse student samples through qualitative methods (focus groups and cognitive interviewing) and quantitative methods (measurement invariance by race/ethnicity, socioeconomic status, and grade level) to support cultural responsiveness.
- Relationship to math perceptions and beliefs: The instrument includes a subscale related to sense of belonging. Overall, the instrument also assesses enjoyment.
- Other considerations for use: The instrument is free and could be administered virtually. A shorter version is also available upon request. The survey has been used with English learners and translated into more than 10 languages. An audio version of questions is also available for student use. A teacher engagement scale is also available, which has 20 items on a 5-point Likert scale. Observational tools are also available.

P7. The **Math Identity Measures** is a 10-item subscale that measures students' math identity and attitudes. Subscales include interest, recognition, and competence/performance. This measure was originally included as a part of the Factors Influencing College Success in Mathematics Survey. The measure was tested on students enrolled in college-level calculus classes across the United States. No information is available on time to administer.

- Key publication: Cribbs et al. (2015).
- Considerations for reliability: The instrument meets the required reliability threshold overall. All subscales meet the required reliability threshold based on prior research, except for the recognition subscales. The scale was used with communities in focus, but reliability is not reported for racial or ethnic subgroups. Reliability was established in an undergraduate postsecondary sample in which 7 percent of students were Latino and 5 percent were Black. Additional testing is needed with targeted age groups and communities in focus.
- Relationship to math perceptions and beliefs: The instrument includes items related to self-efficacy/confidence and enjoyment.
- Other considerations for use: The instrument is free, but permission for use is required. It could be administered virtually.

P8. The **Mathematical Mindset** is a 29-item survey students complete to measure their beliefs and attitudes on math mindset, math as creative and connected, and fear of math on a 6-point Likert scale. The measure was designed for grades 6 to 8. The full scale takes about 7 to 10 minutes to administer.

- Key publication: Boaler et al. (2018).
- Considerations for reliability: Reliability information for the instrument overall is not available. One subscale (fear of math) meets the required reliability threshold, while the mindset and math as creative and connected subscales did not meet the required reliability threshold based on prior research. Analysis was conducted on a subset of nine items. Reliability was established with a sample of students in grades 6 to 8 in which 20 percent were Latino, 2 percent were Black, and 21 percent were eligible for free or reduced-price lunch. Reliability is not reported for racial or ethnic subgroups. Additional testing is needed with the communities in focus using the full survey items.
- Relationship to math perceptions and beliefs: The instrument includes items related to self-efficacy/confidence, enjoyment, growth mindset, and the value and importance of math.
- Other considerations for use: Contact the authors for information about the cost and availability of the instrument. This instrument has been used in a virtual setting.

P9. The **Mathematics Attitude Inventory** is a 48-item survey students complete to measure their beliefs and attitudes on six constructs: perception of mathematics teacher, anxiety toward mathematics, value of mathematics in society, self-concept in mathematics, enjoyment of mathematics, and motivation in mathematics. The measure was designed for students in grades 7 to 12. The measure takes about 20 minutes to administer.

- Key publication: Sandman (1980).
- Considerations for reliability: The instrument meets the required reliability threshold overall. All subscales meet the required reliability threshold based on prior research. Reliability was established with a sample in grades 7 to 12; no racial demographic information was provided. Reliability is not reported for racial or ethnic subgroups. Additional testing is needed with targeted grades and communities in focus.
- Relationship to math perceptions and beliefs: The instrument includes subscales related to self-efficacy/ confidence, enjoyment, and the value and importance of math.
- Other considerations for use: This instrument is available for purchase, and it is unclear whether the publisher would allow use in a virtual setting.

P10. The **Mathematics Self-Efficacy Scale** is a 9-item survey students complete to evaluate their self-efficacy in the classroom and a test setting using a 5-point Likert scale. The study examined students in grades 9 and 10. Administration should take about 10 to 15 minutes.

- Key publication: Nielsen & Moore (2003).
- Considerations for reliability: This instrument meets the required reliability threshold based on prior research. Reliability was established based on 300 high school students in Australia, with no information about sample demographics. Reliability is not reported for

racial or ethnic subgroups. Additional testing is needed with targeted grades and communities in focus.

- Relationship to math perceptions and beliefs: The instrument assesses self-efficacy/confidence.
- Other considerations for use: The measure is free and publicly available.

P11. The **Patterns of Adaptive Learning Scales (PALS)** survey is a 94-item survey developed by a group of researchers to assess the relationship between students' learning environment and student motivation, affect, and behavior. The subscales were developed to be used all together or individually. To measure math perceptions and beliefs, this menu recommends using an adapted 4-item version of the academic self-efficacy subscale that was used in Simzar et al. (2016). All survey questions are measured on a 5-point Likert scale.

- Key publications: Midgely et al. (2000); Simzar et al (2016).
- Considerations for reliability: The internal consistency of each subscale was assessed in the original study and meets the required reliability threshold. Some subscales were assessed with different grades ranging from grades 5 to 9. The adapted academic self-efficacy subscale also meets the required reliability threshold based on research from Simzar et al. (2016). In this study, reliability was established with a sample of students in grade 8 in which 67 percent were Latino and 75 percent were eligible for free or reduced-price lunch.
- Relationship to math perceptions and beliefs: The 4-item subscale used in Simzar et al. (2016) assesses self-efficacy/confidence.
- Other considerations for use: The subscale is available for free and could be administered virtually. According to Midgley et al. (2000), the PALS subscales have been used across various contexts: from elementary through high school, schools serving families with low and middle incomes, and schools with a student population that is majority Black.

P12. The **Sense of Connectedness to My Mathematics Classroom** is a 9-item survey using a 6-point Likert scale. It includes three subscales: emotional support from math teacher, my math class is like a family, and my contributions are valued. The study examined students in grades 5 to 9. No information is available on time to administer.

- Key publication: Maloney & Matthews (2020).
- Considerations for reliability: Reliability information for the instrument overall is not available. Two subscales (emotional support from math teacher and my math class is like a family) meet the required reliability threshold based on prior research. The subscale my contributions are valued did not meet the required threshold. Reliability was established in a sample of students in which 56 percent were Black and 30 percent were Latino. Reliability is not reported for racial or ethnic subgroups. Additional testing is needed with targeted grades and communities in focus.
- Relationship to math perceptions and beliefs: The instrument assesses sense of belonging. Survey responses are dependent on students' perceptions of the current classroom environment, rather than math more broadly.

- Other considerations for use: The instrument is available without restrictions on access. The means and standard deviations are provided for the subscales. It can be implemented in a virtual setting.

P13. The **Perceived Utility Value** instrument is a 5-item survey that measures a student's perception of the value and importance of math on a 7-point scale. Although this scale is similar to other utility value scales, it focuses on the utility value of specific techniques within a topic or subject as opposed to generally valuing the topic or subject. Although this measure is not specific to math, it is generally applicable and can be adapted to measure math specifically. The study examined postsecondary students. No information is available on time to administer.

- Key publication: Canning & Harackiewicz (2015).
- Considerations for reliability: This instrument meets the required reliability threshold based on prior research. Reliability was established based on 88 undergraduate postsecondary students enrolled in an introductory psychology course at a university in the United States. Reliability was established in a sample of students in which 5 percent were Latino and 1 percent were Black. Reliability is not reported for racial and ethnic subgroups. Additional testing is needed with targeted grades and communities in focus.
- Relationship to math perceptions and beliefs: The instrument assesses value and importance of math.
- Other considerations for use: The instrument is available without restrictions on access. It can be implemented in a virtual setting.

P14. The **Investigative Scale of Mapping Vocational Challenges** is a 15-item survey using a 3-point scale (1 = low to 3 = high) that measures the degree of interest students have in 15 STEM careers and the actions they took to prepare for that occupation (e.g., taking advanced coursework). The study examined students in grades 10 to 12. No information is available on time to administer.

- Key publication: Turner et al. (2019).
- Considerations for reliability: This instrument meets the required reliability threshold based on prior research. Reliability was established with a sample of students in grades 10 to 12 in which 41 percent of students were Black and 3 percent were Latino. Reliability is not reported for racial and ethnic subgroups. Additional testing is needed with targeted grades and communities in focus.
- Relationship to math perceptions and beliefs: The instrument assesses interest in math education and career.
- Other considerations for use: Contact the authors for information about the cost and availability of the instrument. It can be implemented in a virtual setting.

P15. **Interest in a Future STEM Career (IFSTEMC)** is a 4-item survey using a 6-point Likert scale that measures students' self-reported interest in pursuing a STEM career. The study examined students in grades 9 to 12. No information is available on time to administer.

- Key publication: LaForce et al. (2017).

- Considerations for reliability: This instrument meets the required reliability threshold based on prior research. Reliability was established with a sample of students in grades 9 to 12 in which 28 percent of students were Latino and 10 percent were Black. Reliability is not reported for racial and ethnic subgroups. Additional testing is needed with targeted grades and communities in focus.
- Relationship to math perceptions and beliefs: The instrument assesses interest in math education and career.
- Other considerations for use: The instrument is available without restrictions on access. It can be implemented in a virtual setting.

P16. The **Career Interest Questionnaire** is a 13-item survey using a 5-point Likert scale that measures students' self-reported interest in pursuing STEM-related careers. The survey is divided into three subscales: measuring perception of a supportive environment for pursuing a career in science, interest in pursuing educational opportunities that would lead to a career in science, and perceived importance of a career in science. The study examined students in grades 6 to 8. No information is available on time to administer.

- Key publication: Tyler-Wood et al. (2010).
- Considerations for reliability: This instrument and its subscales meet the required reliability threshold based on prior research. Reliability was established with a sample of students in grades 6 to 8 in which 1 percent of students were Latino and 1 percent were Black. Reliability is not reported for racial and ethnic subgroups. Additional testing is needed with communities in focus.
- Relationship to math perceptions and beliefs: The instrument assesses interest in math education and career.
- Other considerations for use: The instrument is available without restrictions on access (Knezek & Christensen, 2013). It can be implemented in a virtual setting.

Measures of math use beyond the classroom

Math use beyond the classroom refers to the use of math in the short-term outside of classroom activities and assignments as well as in the longer-term in future education and career. Measuring math use beyond the classroom is more complex than the other domains in this menu because, by definition, it occurs outside of the classroom and sometimes far in the future. Observing behaviors outside of school is costly and often not feasible; asking students to self-report on behaviors outside of school requires them to recognize and recall math use activities reliably; and collecting postsecondary education and employment data many years in the future requires student consent and more time and resources than many studies have. In addition, students can only use math if they have the opportunities to do so. Considerations for understanding and measuring opportunities for math use beyond the classroom are in Box 3. For links to specific data sources, see the **References** section.

Box 3. Consider students' opportunities to use math

The extent of math use beyond the classroom depends on students' opportunities for math use in and beyond the classroom. Students in different classrooms, schools, families, and communities may have different opportunities to use math. For example, students may have differences in advanced course options, internship opportunities, extracurricular activities, and their daily life routines, so it is important to confirm that you are measuring the actual use of math and not differences in the opportunities to use math. To learn if you are measuring the actual use of math you could:

- Solicit student and community input on opportunities to use math outside of school that are relevant to the communities in focus.
- Directly measure opportunities to use math, for example by surveying students or documenting course options.
- Select instruments that are applicable across different settings.

Compounding these complexities is the lack of well-developed and tested ways to measure choices and behaviors related to using math beyond the classroom. As a result, this section does not include existing instruments. Instead, it includes recommendations about what kinds of surveys could be developed and what kinds of administrative data could be collected to assess different aspects of short- and longer-term math use. This is a field in need of further measure development to facilitate consistent and reliable measurement across studies.

U1. **Self-initiated use of math in the community** captures the short-term application of math in students' everyday lives. This might be an early indicator that students will continue to use math in education and their career, and might be a desired outcome on its own for interventions.

- Data sources: A survey could be developed to ask people to self-report on their math use in the community.
- Examples of measures: A custom survey could assess the frequency and ways that a person self-initiates using math in the community. For example, it might include a checklist of potential ways that math is used in the community, specific scenarios that could require

math and ask how students approach them, or open-ended responses about the ways that they use and see math in their everyday lives.

- Considerations: Opportunities for using math in the community will vary across settings and should be accounted for (see Box 3).

U2. **Enrollment and completion of math and STEM coursework** is important because it reflects independent interest and motivation to engage and develop math skills, and is an intermediate outcome that facilitates use of math in higher education and career.

- Data sources: School administrative data systems could be used to track enrollment and completion of math and STEM coursework. A survey could also be developed to ask students to self-report on course taking and performance.
- Examples of measures: Enrollment in math and STEM coursework could be measured by calculating the percentage of students who enroll in advanced or elective math or STEM courses in the year(s) following the intervention out of all students enrolled in the grade level. Completion could be measured by calculating the percentage of students who complete with a passing grade out of all students enrolled in the course.
- Considerations: Opportunities for course taking will vary across settings and should be documented and accounted for (see Box 3). Additionally, there might be eligibility criteria for course taking within a school. It will be important to understand this and account for it when determining who to include in the denominator of the enrollment calculation. Another challenge is determining which courses are advanced, elective, and STEM-related. This could be time intensive and difficult to standardize across schools.

U3. **Involvement in math and STEM internships, Career Technical Education (CTE), and other occupational training** is important because it reflects interest to use math skills in the longer term, or even in the pursuit of a career. It can also reinforce learning of knowledge and skills, and it is an intermediate outcome that facilitates use of math in higher education and career.

- Data sources: School administrative data systems could be used to track enrollment in CTE. They might track involvement in internships and occupational training, but this will not be done consistently across schools. A survey could be developed to ask students to self-report on involvement.
- Examples of measures: Involvement in math or STEM CTE or occupational training could be measured by calculating the percentage of students who enroll in math or STEM CTE courses out of all students enrolled in any CTE or any occupational training. Involvement in math or STEM internships could be calculated as a percentage of students who apply to, participate in, or complete internships. This could be a percentage of all students enrolled in the grade level or as a percentage of all students involved in any internship.
- Considerations: Opportunities to participate in CTE or occupational training will vary across settings and should be documented and accounted for (Box 3). Additionally, there might be eligibility criteria or a limit on the number of students who can participate in CTE or occupational training. In addition to a binary measure of involvement, it might be informative to measure frequency and duration of participation. It will be important to understand these

considerations and account for them when determining who should be included in the denominator of the involvement calculation. Another challenge is determining which opportunities are STEM-related. This could be time intensive and difficult to standardized across schools.

U4. **Involvement in math and STEM extracurricular activities** is important because it reflects interest and motivation to use math skills outside of the classroom. It can also reinforce learning of knowledge and skills and is an intermediate outcome that facilitates use of math in higher education and career.

- Data sources: School administrative data systems might track school-affiliated extracurricular activities, but this will not be done consistently across schools. A survey could be developed to ask students to self-report on involvement.
- Examples of measures: Involvement in math and STEM extracurricular activities could be measured by calculating the percentage of students who are involved in math or STEM extracurricular activities. This could be a percentage of all students enrolled in the grade level or as a percentage of all students involved in any extracurricular activity.
- Considerations: Opportunities to participate in math and STEM extracurricular activities will vary across settings and should be documented and accounted for (Box 3). Additionally, there might be eligibility criteria or a limit on the number of students who can participate in a particular math or STEM extracurricular activity. In addition to a binary measure of involvement, it might be informative to measure frequency and duration of participation. It will be important to understand these considerations and account for them when determining who to include in the denominator of the involvement calculation. Another challenge is determining which opportunities are STEM-related. This could be time intensive and difficult to standardized across schools.

U5. **Enrollment in math and STEM postsecondary majors** is important because it is a key step on the pathway to use math skills in employment or other aspects of adult life.

- Data sources: University administrative data systems or the National Student Clearinghouse could be used to track enrollment and completion of credentials in math and STEM postsecondary majors. A survey could also be developed to ask students to self-report on enrollment in postsecondary majors.
- Examples of measures: Enrollment in math and STEM postsecondary majors could be measured by calculating the percentage of students who enroll in math and STEM majors out of all students who enrolled in a postsecondary institution or out of all students who participated in the intervention.
- Considerations: Opportunities for students to enroll in postsecondary majors will vary and should be documented and accounted for (Box 3). Additionally, there may be eligibility criteria for enrolling in a math and STEM postsecondary major. Another consideration is determining which postsecondary majors are STEM-related and which programs to include because they may vary in length and requirements (e.g., a 4-year bachelor’s degree program compared to a 2-year associate degree program). An additional consideration is that some students will choose not to enroll in postsecondary education, so a decision will

need to be made to either include all students or just those who choose to pursue postsecondary education. It will be important to understand these considerations and account for them when determining who to include in the denominator of the enrollment calculation.

U6. **Employment in math or STEM occupations** might be a desired long-term outcome for interventions that want to increase representation of the communities in focus in STEM fields or promote economic self-sufficiency through employment in high-demand or higher-wage occupations.

- Data sources: Administrative data from the National Directory of New Hires or state Unemployment Insurance data systems could be used to collect information about employment and earnings. A survey could be developed to ask people to self-report on occupation.
- Examples of measures: A custom survey could assess employment in a math or STEM field, earnings, job title, years of employment, and job progression.
- Considerations: Access to administrative data might be a barrier (e.g., consent, need for participants' Social Security numbers), and the data may be limited (e.g., may not capture contract work or self-employment). Also, administering a survey to participants years after completion of an intervention will be challenging and may yield low response rates.

U7. **Employment and career progression within any field** might be a desired long-term outcome for interventions that want to promote economic self-sufficiency.

- Data sources: Administrative data from the National Directory of New Hires or state Unemployment Insurance data systems could be used to collect information about employment and earnings. A survey could be developed to ask people to self-report on employment and career progression.
- Examples of measures: A custom survey could assess quarterly employment, earnings, job title, years of employment, and job progression.
- Considerations: Access to administrative data might be a barrier (e.g., consent, need for participants' Social Security numbers), and the data may be limited (e.g., may not capture contract work or self-employment). Additionally, administering a survey to participants years after completion of an intervention will be challenging and may yield low response rates.

References

Gray, D. L., Hope, E. C., & Matthews, J. S. (2018). Black and belonging at school: A case for interpersonal, instructional, and institutional opportunity structures. Educational psychologist, 53(2), 97-113. https://doi.org/10.1080/00461520.2017.1421466.

Ojose, B. (2011). Mathematics literacy: Are we able to put the mathematics we learn into everyday use? *Journal of Mathematics Education*, *4*(1), 89–100.

Organisation for Economic Co-operation and Development. (2019). *PISA 2018 Assessment and Analytical Framework.* https://doi.org/10.1787/b25efab8-en.

Measuring classroom context and opportunity structures (Box 2)

Belonging-Centered Instruction Observational Protocol

Matthews, J. S., Gray, D. L., Lachaud, Q., McElveen, T. L., Chen, X., Victor, T., Okai, E., Boomhower, K., Wu, J., & Cha, E. (2021). *Belonging-Centered Instruction: An approach toward building inclusive mathematics classrooms.* Student Experience Success Network. https://studentexperiencenetwork.org/wp-content/uploads/2021/10/SERN_Research-Snapshot_Matthews_102721.pdf

Reformed Teaching Observation Protocol (RTOP)

Piburn, M., Sawada, D., Falconer, K., Turley, J., Benford, R., & Bloom, I. (2000). *Reformed Teaching Observation Protocol (RTOP): Reference manual*. Arizona Collaborative for Excellence in the Preparation of Teachers. https://www.public.asu.edu/~anton1/AssessArticles/Assessments/Science%20Assessments/RTOP%20Reference%20Manual.pdf.

Mathematics Scan (M-Scan)

Berry, III, R. Q., Rimm-Kaufman, S. E., Ottmar, E. M., Walkowiak, T. A., & Merritt, E. (2012). *The Mathematics Scan (M-Scan): A measure of mathematics instructional quality* [Unpublished measure]. University of Virginia.

Mathematics Classroom Observation Protocol for Practices (MCOP2)

Gleason, J. (2015). *Mathematics Classroom Observation Protocol for Practices (MCOP2)*. https://jgleason.people.ua.edu/mcop2.html.

UTeach Observation Protocol (UTOP) for Math and Science

UTeach Natural Sciences (2014). *The UTeach Observation Protocol (UTOP) for Mathematics and Science Training Guide and User Manual.* University of Texas Austin. https://pd.uteach.utexas.edu/sites/default/files/documents/2020-08/UTOP-Training-Guide-Math-and-Science.pdf.

Teaching for Robust Understanding (TRU) Framework

Schoenfeld, A. H., & Teaching for Robust Understanding Project. (2016). *An Introduction to the Teaching for Robust Understanding (TRU) Framework*. University of California, Berkeley Graduate School of Education. https://truframework.org/wp-content/uploads/2018/03/Introduction-to-TRU-2018-version.pdf.

Mathematical Quality of Instruction (MQI)

Hill, H.C. (2014). *Mathematical Quality of Instruction (MQI): 4-point version*. University of Michigan Learning Mathematics for Teaching Project.

Copilot-Elevate

PERTS. (n.d.). *Copilot-Elevate*. https://www.perts.net/elevate.

Math skills and knowledge measures

K1. ClearSight (formerly known as AIRways)

Voyager Sopris Learning. (n.d.). *ClearSight*. https://www.voyagersopris.com/product/assessment/clearsight/clearsight-overview.

K2. Performance Series

Scantron. (2014). *Scantron's computer adaptive assessment solution selected by the California Department of Education*. https://www.scantron.com/news/scantrons-computer-adaptive-assessment-solution-selected-by-the-california-department-of-education/.

Scantron. (n.d.). *Assessment Solutions*. https://www.scantron.com/assessment-solutions/.

K3. PSAT 8/9

College Board. (n.d.). *PSAT 8/9–Inside the test–Math test*. https://collegereadiness.collegeboard.org/psat-8-9/inside-the-test/math.

K4. Trends in International Mathematics and Science Study (TIMSS)

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). TIMSS 2015 international results in mathematics. TIMSS & PIRLS International Study Center, Boston College.

National Center for Education Statistics. (n.d.). *Assessment questions*. https://nces.ed.gov/timss/educators.asp.

K5. Mathematics Assessment Collaborative/ Mathematics Assessment Resource Service (MAC/ MARS)

Mathematics Assessment Resource Service. (n.d.). *Balanced Assessment in Mathematics: The tests*. http://www.toolkitforchange.org/toolkit/documents/9_108_bamtests.pdf.

K6. Balanced Assessment in Mathematics (BAM)

Balanced Assessment. (2015). Balanced Assessment in Mathematics. https://hgse.balancedassessment.org/.

Stancavage, F., Shepard, L., McLaughlin, D., Holtzman, D., Blankenship, C., & Zhang, Y. (2009). Sensitivity of NAEP to the effects of reform-based teaching and learning in middle school mathematics. American Institutes for Research. https://files.eric.ed.gov/fulltext/ED506807.pdf.

K7. Star Math

Renaissance. (n.d.). Renaissance Star Math. https://www.renaissance.com/products/star-math/.

Renaissance Learning, Inc. (2020). Research foundation for Star Adaptive Assessments [White paper]. http://doc.renlearn.com/KMNet/R001480701GCFBB9.pdf.

K8. Iowa Algebra Aptitude Test

Riverside Insights. (n.d.). Iowa Algebra Aptitude Test. https://www.riversideinsights.com/iowa_algebra_aptitude_test?tab=0.

K9. Program for International Student Assessment (PISA)

Organisation for Economic Co-Operation and Development. (1997). *Measuring student knowledge and skills: A new framework for assessment*. https://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/33693997.pdf.

Organisation for Economic Co-operation and Development. (2019). *PISA 2018 Assessment and Analytical Framework.* https://doi.org/10.1787/b25efab8-en.

K10. Tests for Understanding Fractions

Instructional Research Group. (2015). https://www.inresg.org/.

Jayanthi, M., Gersten, R., Taylor, M. J., Smolkowski, K., & Dimino, J. (2017). *Impact of the Developing Mathematical Ideas professional development program on grade 4 students' and teachers' understanding of fractions* (REL 2017–256). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. https://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/rel_2017256.pdf.

Math perceptions and beliefs measures

P1. Attainment Value of Mathematics

Matthews, J. S. (2018). When am I ever going to use this in the real world? Cognitive flexibility and urban adolescents' negotiation of the value of mathematics. *Journal of Educational Psychology, 110*(5), 726–746. https://doi.org/10.1037/edu0000242.

P2. Beliefs, Engagement, and Attitude of Math Motivation Scale

Orosco, M. J. (2016). Measuring elementary student's mathematics motivation: A validity study. *International Journal of Science and Mathematics Education, 14*(5), 945–958.

P3. Expectancy-Cost-Value Scale

Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2015). A practical measure of student motivation: Establishing validity evidence for the Expectancy-Value-Cost Scale in middle school. *The Journal of Early Adolescence*, *35*(5–6), 790–816. **https://doi.org/10.1177/0272431614556890**.

Lauermann, F., Tsai, Y. -M., & Eccles, J. S. (2017). Math- related career aspirations and choices within Eccles et al.'s expectancy–value theory of achievement-related behaviors. *Developmental Psychology, 53*(8), 1540–1559.

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental Psychology, 42*(1), 70.

Wang, M. T. (2012). Educational and career interests in math: A longitudinal examination of the links between perceived classroom environment, motivational beliefs, and interests. *Developmental Psychology, 48*(6), 1643–1657.

P4. Indiana Mathematics Belief Scale

Ayebo, A., & Mrutu, A. (2019). An exploration of calculus students' beliefs about mathematics. *International Electronic Journal of Mathematics Education, 14*(2), 385–392.

Kloosterman, P., & Stage, F. K. (1992). Measuring beliefs about mathematical problem solving. *School Science and Mathematics, 92*(3), 109–115.

P5. Math and Me Survey

Adelson, J. L., & McCoach, D. B. (2011). Development and psychometric properties of the Math and Me survey: Measuring third through sixth graders' attitudes toward mathematics. *Measurement and Evaluation in Counseling and Development, 44*(4), 225–247.

P6. Math and Science Engagement Scales

Fredricks, J. A., Wang, M. T., Schall, J., Hofkens, T. L., & Parr, A. (2016). Using qualitative methods to develop a survey measure of math and science engagement. Learning and Instruction, 43, 5–15.

Wang, M. T., Fredricks, J. A., Ye, F., Hofkens, T. L., & Schall, J. (2016). The Math and Science Engagement Scale: Scale development, validation, and psychometric properties. Learning and Instruction, 43, 16–26.

P7. Math Identity Measures

Cribbs, J. D., Hazari, Z., Sonnert, G., & Sadler, P. M. (2015). Establishing an explanatory model for mathematics identity. Child Development, 86, 1048–1062.

P8. Mathematical Mindset

Boaler, J., Dieckmann, J. A., Pérez-Núñez, G., Sun, K. L., & Williams, C. (2018, April). Changing students' minds and achievement in mathematics: The impact of a free online student course. Frontiers in Education, 3, 26.

P9. Mathematics Attitude Inventory

Sandman, R. S. (1980). The Mathematics Attitude Inventory: Instrument and user's manual. Journal for Research in Mathematics Education, 11(2), 148–149.

P10. Mathematics Self-Efficacy Scale

Nielsen, I. L., & Moore, K. A. (2003). Psychometric data on the Mathematics Self-Efficacy Scale. Educational and Psychological Measurement, 63(1), 128–138.

P11. Patterns of Adaptive Learning Scales (PALS)

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., & Urdan, T. (2000). Manual for the patterns of adaptive learning scales. University of Michigan, 1-61. **http://websites.umich.edu/~pals/PALS%202000_V12Word97.pdf**.

Simzar, R., Domina, T., & Tran, C. (2016). Eighth-grade algebra course placement and student motivation for mathematics. AERA Open, 2(1), 1–26.

P12. Sense of Connectedness to My Mathematics Classroom

Maloney, T., & Matthews, J.S. (2020). Teachers' critical care and students' feelings of connectedness in the urban mathematics classroom. Journal of Research in Mathematics Education.

P13. Perceived Utility Value

Canning, E. A., & Harackiewicz, J. M. (2015). Teach it, don't preach it: The differential effects of directly-communicated and self-generated utility-value information. *Motivation Science, 1*(1), 47–71. **https://doi.org/10.1037/mot0000015.**

P14. Investigative Scale of Mapping Vocational Challenges

Turner, S. L., Joeng, J. R., Sims, M. D., Dade, S. N., & Reid, M. F. (2019). SES, gender and STEM career interests, goals, and actions: A test of SCCT. *Journal of Career Assessment, 27*(1), 134–150. **https://doi.org/10.1177/1069072717748665.**

P15. Interest in a Future STEM Career (IFSTEMC)

LaForce, M., Noble, E., & Blackwell, C. (2017). Problem-based learning (PBL) and student interest in STEM careers: The roles of motivation and ability beliefs. *Education Sciences, 7*(4), 1–22. **https://doi.org/10.3390/educsci7040092**.

P16. Career Interest Questionnaire

Knezek, G., & Christensen, R. (2013). *Career Interest Questionnaire*. https://iittl.unt.edu/sites/default/files/Instruments/Career%20Questionnaire%20v20.pdf.

Tyler-Wood, T. L., Knezek, G., & Christensen, R. (2010). Instruments for assessing interest in STEM content and careers. *The Journal of Technology and Teacher Education, 18*, 345–368.

Math use beyond the classroom measures

U5. Enrollment in math and STEM postsecondary majors

National Student Clearinghouse. (2021). *StudentTracker for High Schools.* https://www.studentclearinghouse.org/high-schools/studenttracker/.

Appendix: Measure criteria

This set of criteria includes the three steps for assessing math measures. These criteria were used to assess measures in the menu and can be applied to other measures as well. The menu of measures highlights measures that meet all or some of the criteria, and it documents which criteria each measure meets and does not meet. In addition to this set of general criteria, users assessing and selecting measures should consider the context-specific questions listed in the menu. These questions will help assess whether a measure is a good fit in terms of content and feasibility.

The following high-level steps guide decision making when assessing whether a measure should be used for evaluation purposes:

- **Step 1.** Users should aim to use measures that meet all of the Step 1 criteria, which the panel identified as a minimal benchmark for measures to meet.
- **Steps 2 and 3.** Users should prioritize measures that meet some or all of the Step 2 and 3 criteria if possible. In schools that serve English learners, the measure should also meet criteria in Step 2a. Step 2a can be adapted for other languages as well. If no available measures meet criteria in Steps 2, 2a, and 3, then users could adapt measures as recommended in those steps.

Note that “communities in focus” refers to the students whom the Middle Years Math grants seek to serve: students who are Black, Latino, and/or experiencing poverty.

Step 1: Determine if measures meet required criteria

Can be measured reliably across conditions

- For standardized tests that have multiple items related to the same construct: Internal consistency must be >0.70.
- For surveys with self-reported measures of perceptions, beliefs, and behaviors related to the same construct: Internal consistency for subscales must be >0.70. All subscales must meet this minimum level.
- For measures that require subjective judgement to assign ratings or scores: More than one rater should score a subset of assessments, and inter-rater reliability must be >0.70. Even if prior research demonstrates inter-rater reliability, grantees should establish inter-rater reliability with their own raters as well.

Measure is specific to math

Step 2: Determine if measures meet the preferred criteria

Measure has demonstrated reliability and validity with communities in focus

- Measure has demonstrated reliability and validity with students from communities in focus OR has been qualitatively tested (through focus groups, cognitive interviewing, or talk-aloud protocols) with students from communities in focus.
- Measure was modified to reflect findings from qualitative testing, and reliability and validity were reestablished on modified version.

Measure is available for use by programs without restrictions on access

- Programs can access and use the measure without substantial burden.

Guidance is available to help users consistently interpret results

Linguistic accessibility criteria for users serving English learners

Step 2a: Determine if measures meet additional preferred criteria

Measure is linguistically accessible

- Directions and test items are presented in students' native language alongside original English version OR one or more tools is provided to support comprehension, including:
  - Audio versions of questions
  - Dictionaries
  - Pop-up glossaries
  - Audio glossaries
  - Printed glossaries
- Translations and tools include minimal barriers to use (e.g., minimal mouse-clicking skills)
- Has been translated by a culturally competent translator with translation best practices, including:
  - Translator is a native speaker and considers differences in vocabulary, pronunciation, forms of speech, and idiomatic expressions in multiple dialects
  - Translator has math content knowledge (if applicable)
  - Translation performed by independent translators with translation reconciliation to resolve discrepancies
- Test items have been assessed for linguistic accessibility, including:
  - Equivalence of items in different languages assessed with differential item functioning
  - Technical terminology and translation assessed using cognitive interviews with students from communities in focus to ensure that measure elicits intended mental processes
- Measure was modified to reflect findings from qualitative testing, and reliability and validity were reestablished on modified version

Step 3: Determine if measures meet the following preferred content- specific criteria

For math skills and knowledge measures

Measure assesses deep knowledge of math, not rote memorization of math facts

Measure assesses one of the math proficiency strands (National Research Council, 2001)

- Procedural fluency—ability to carry out procedures flexibly, accurately, efficiently, and appropriately
- Conceptual understanding—fundamental understanding of mathematical ideas, and the ability to integrate new ideas into understanding by connecting with previous knowledge
- Strategic competence—ability to formulate, represent, and solve mathematical problems
- Adaptive reasoning—capacity for logical thought, reflection, explanation, and justification
- Productive dispositions regarding big ideas in mathematics

Measure assesses cognitively complex concepts

For math perceptions and beliefs measures

Measure assesses one of the key math perceptions and beliefs constructs identified by the panel

- Key constructs include sense of belonging, self-efficacy/confidence, enjoyment, growth mindset, and value and importance of math.

Measure is meant to assess state-based characteristics instead of trait-based characteristics

- State-based characteristic refers to a temporary way of thinking, feeling, behaving, and so on, while a trait-based characteristic is considered to be a stable and enduring characteristic.

References

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. National Academies Press. https://doi.org/10.17226/9822.

Appendix: Panel of researchers

In Fall 2019, Mathematica formed a panel of researchers in the field of math instruction and measurement to provide input on the development of an initial version of the menu of measures. This initial version of the menu focused on (1) measures of math skills and knowledge and (2) math perceptions and beliefs. In Spring 2021, Mathematica formed a second panel of researchers to advise on the expansion of the menu of measures to include math use and to advise on revisions to conceptual framework.

The goal in selecting the panels was to balance expertise in mathematics content, teacher practice, and psychometrics. We prioritized consultants who identify with the communities served by the Bill & Melinda Gates Foundation's Middle Years Math grants, as well as those who have focused on these communities in their research. Table 1 shows the researchers included in each panel.

Table 1. Panelists who advised on development of menu of measures

Name	Affiliation	Areas of expertise	Panel
Steven Culpepper	University of Illinois, Urbana	Psychometrics, large-scale assessments including the National Assessment of Educational Progress (NAEP) and New York State Regents exams, measures of mathematical knowledge, measure development and administration	Math skills and knowledge; math perceptions and beliefs
Russell Gerstein	College of Education at the University of Oregon; Executive Director of the Instructional Research Group	Measurements of program implementation, math instructional innovations, assessments of math with focus on low-achieving students	Math skills and knowledge; math perceptions and beliefs
Chris Hulleman	University of Virginia	Social and personality psychology, motivation, learning mindsets, relevance of coursework for students, science and math interventions and coursework	Math use
Jamaal Shariff Matthews	University of Michigan	Motivation in mathematics in urban schools, how context and teachers shape students' beliefs about mathematics	Math skills and knowledge; math perceptions and beliefs; math use
Ming-Te Wang	University of Pittsburgh	Achievement motivation and engagement, school-based psychosocial interventions, social and emotional development, STEM learning and career development, identity development, school climate and school discipline, ethnic-racial socialization, racial and gender biases and disparities, risk and resilience	Math skills and knowledge; math perceptions and beliefs; math use

The panels met periodically during the development of the menu of measures via virtual video calls. During the meetings, the panels discussed issues related to the conceptual framework, relevant constructs, common measures used in the field, and criteria for assessing measures. Additional ad hoc communication occurred via email between meetings. Panelists also provided input on drafts of the menu of measures.

This publication is based on research funded by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.