

Future-Focused Measures for High School Math A Measures Guide



May 2026

Acknowledgements

Kate Place, Menbere Shiferaw, Riley Stone, Grady Deacon, Elizabeth Fei, Kara Conroy, and Mayra Sandoval developed this guide. An external panel of experts—Kate Miller (Concord Consortium), Thema Monroe-White (George Mason University), Joseph Polman (University of Colorado, Boulder), and Helen Zhang (Boston College)—reviewed the content and provided feedback. Representatives from CourseKata, Data Science 4 Everyone, Skew The Script, ThinkData Ed also provided suggestions and feedback on the measures in this guide. Virginia Knechtel conducted a quality assurance review. Sheena Flowers provided design and production support, and Donovan Griffin and Effie Metropolous provided editorial support. This publication was prepared for the Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Gates Foundation.

Future-Focused Measures for High School Math: A Measures Guide



Across the United States, students, families, educators, districts, and states are recognizing the need for more relevant, data-infused high school math courses that prepare students to achieve their postsecondary aspirations, succeed in their careers, and lead informed, productive lives. In response, curriculum developers and professional learning providers are advancing efforts to modernize high school math by strengthening core courses (Algebra I, Geometry, and Algebra II) and expanding advanced offerings (such as data science and statistics). By designing updated instructional materials and aligned professional learning supports, these providers aim to enhance teacher practice and increase student motivation, engagement, and persistence (MEP). When implemented effectively, these efforts have the potential to improve students' learning and better prepare them for success in postsecondary education and their careers.

How to use this guide

This guide presents a curated set of student, teacher, and implementation measures that will (1) provide curriculum developers with data they need to inform curriculum design and improvement and (2) help researchers identify measures to use in studies on data-infused curricula and course offerings. Although it is not exhaustive, this guide provides options of measures that curriculum developers and researchers could use during **implementation pilots** or **efficacy studies**.

Road map to the guide

This guide has the following components:

- Measuring student outcomes
- Measuring teacher outcomes
- Measuring implementation
- Appendix A: Measure criteria

The guide has separate sections on three kinds of measures: **student measures** (measures of student learning and achievement and MEP), **teacher measures**, and **implementation measures**. We prioritized student and teacher measures that have demonstrated reliability and validity, have been tested with high school students or teachers in the United States, are publicly available, and are feasible to administer in school contexts. We have also included some new measures that have not yet demonstrated reliability and validity because they focus on emerging areas (such as data science and data literacy) that do not currently have established measures. Throughout the guide, we highlight gaps in available measures, including measures that still need to be piloted or do not yet exist, and we have included a list of these measure gaps in Appendix C.

Why did we prioritize validated measures?

Validated measures are essential tools for evaluating and improving curricula. This guide prioritizes validated instruments because they:

Ensure accuracy

They reliably assess the concepts they are designed to measure, such as student engagement, learning outcomes, or self-efficacy.

Support consistency

They produce stable results across different contexts, populations, and time points, making findings more trustworthy.

Enable comparability

Results are comparable across studies, programs, or institutions, facilitating broader learning and benchmarking.

Reduce measurement error

They minimize bias and ambiguity, helping researchers draw clearer conclusions and make informed decisions.

Strengthen evidence

Rigorous evaluations, funding proposals, and peer-reviewed publications often require validated measures.




Measuring student outcomes: Student learning and achievement

This section provides guidance on measuring students' learning and academic achievement, which includes their mastery of key concepts and skills and their ability to apply what they learn in meaningful real-world contexts. A central goal of modern math is to provide students with more engaging, relevant, and rigorous math offerings. These offerings emphasize applied, data-infused, and technology-integrated learning, which can help students develop key competencies in data literacy, civic reasoning and discourse, and artificial intelligence (AI) literacy (Exhibit 1). These competencies are increasingly important to prepare students for the demands of postsecondary education, the modern workforce, and participation in civic life. Understanding the depth of students' learning in modern courses is essential for assessing the efficacy of courses and curricula, identifying opportunities for improvement, and building a body of evidence about what works in modern math education.

In Exhibit 1, we define key constructs of student learning outcomes and how they relate to each other.

Exhibit 1. Conceptual framework for advancing student competencies through modern high school math

If learners build understanding of...



Math:
Provides a foundational language and tools (e.g., algebra, calculus, logic) that other fields can use to model data and analyze problems.


Data science:
Integrates statistics, computing, and domain knowledge to organize, analyze, and interpret data with the goals of extracting insights and building predictive models to help solve problems and make decisions.

Statistics:
Focuses on collecting, summarizing, and analyzing data to understand patterns, quantify uncertainty, make inferences, and draw conclusions about a larger population.

And have the opportunity to make connections to...

- their lived experiences
- current events
- other subjects (e.g., history, social studies, civics)

Then they will develop these core competencies...




Data literacy:
The ability to understand, use, and critically interpret data; to tell clear, responsible stories with data; and to critically evaluate data-based claims from others.

Civic reasoning and discourse:
The ability to understand public issues through rigorous inquiry skills and methods, including examination of available evidence. Civic discourse concerns how to communicate with one another around the challenges of public issues in order to enhance both individual and group understanding.

AI literacy:
The ability to understand and apply the concepts behind AI systems, including mathematical functions, probabilistic reasoning, and non-deterministic outputs; to identify and critically assess AI-generated information; and to engage with AI responsibly.

And ultimately become...



A learner who can ask meaningful questions about the world, use mathematical tools and data to explore questions of interest, critically evaluate evidence and AI-generated solutions, and make informed decisions that minimize negative impacts and promote balanced, positive outcomes for themselves, their community, and their future career.



Exhibit 2 presents currently available measures that are related to the conceptual and applied student learning outcomes in Exhibit 1. For each measure, the exhibit includes the construct, availability, strengths, and considerations for use. The measures of student learning and achievement are grouped into the following three categories:

- 1. Standardized assessments of content knowledge.** Tests that are independently developed, administered with consistent scoring and broad participation, and useful for benchmarking across schools or districts.
- 2. Data on course assessments, completion, and grades.** Data generated as part of normal course operations that provide insight into student enrollment and performance within a course, such as formative checks and unit assessments, end-of-course exams, course grades, and course completion rates.
- 3. Measures of the transfer or application of learning.** Tools that evaluate students' ability to apply learning to new contexts, such as civic reasoning.

When selecting a measure of student learning and achievement, curriculum developers and researchers should weigh the rigor of the measure, the feasibility of using it, and how it aligns with the modern math curriculum's theory of change and a study's research questions. Standardized student achievement measures that can be administered consistently across multiple districts and settings support benchmarking, cross-site comparison, and comparison group studies. However, in some cases, course-specific measures (such as unit assessments) may be useful to understand student learning patterns or changes to the curriculum or professional learning—even if they are not well suited for cross-site comparisons or comparison group designs.

To assess the transfer or application of learning to real-world contexts, we have included measures that assess constructs related to data literacy, civic reasoning, and AI literacy. The student learning and achievement measures were selected for the advanced math context (for data science and statistics, for example), but many of the measures are applicable across all high school modern math courses.



Gaps in measures

Currently, there is no national standardized assessment of data science. However, a coalition of national organizations has developed and endorsed a set of learning progressions for data science, found at <https://datasciencelearning.org/>. In addition, ThinkData Ed, in partnership with two multidisciplinary advisory panels, is developing a standardized assessment of data science.

Many national and state standardized assessments (such as the SAT, ACT, Smarter Balanced Assessment Consortium [SBAC], and state-developed assessments) include elements of statistics, data analysis, and problem solving but do not currently capture modern math competencies. As states modernize their math offerings, standardized assessments may also be updated over time.

Exhibit 2. Measures of student learning and achievement

Construct	Potential measures	Source and availability	Pros	Considerations/limitations
Standardized assessments of math knowledge				
Conceptual understanding in statistics	Advanced Placement (AP) Statistics Exam	<input type="checkbox"/> School or district administrative records <input checked="" type="checkbox"/> National or third party <input type="checkbox"/> Curriculum developer Available in AP-offering schools upon request from the College Board (Exam Overview ; College Board Request for Data)	<ul style="list-style-type: none"> Externally validated Supports benchmarking Widely accepted High comparability across courses or curricula 	<ul style="list-style-type: none"> Only for AP students Cost Access to student-level data may be limited
	Levels of Conceptual Understanding in Statistics (LOCUS)	<input type="checkbox"/> School or district administrative records <input checked="" type="checkbox"/> National or third party <input type="checkbox"/> Curriculum developer Items available for administration from LOCUS team	<ul style="list-style-type: none"> Externally validated Supports benchmarking Measures deep understanding of statistics Endorsed by the American Statistical Association and the National Council of Teachers of Mathematics (NCTM) High comparability across courses or curricula Free 	<ul style="list-style-type: none"> Not widely recognized Requires planning and administration Pre-set questions Not designed to assess hands-on data skills that may be integrated into modernized statistics courses
Achievement in subject-aligned standards	District or state end-of-course assessments (e.g., Smarter Balanced Assessment Consortium)	<input checked="" type="checkbox"/> School or district administrative records <input type="checkbox"/> National or third party <input type="checkbox"/> Curriculum developer Limited availability (varies by district)	<ul style="list-style-type: none"> Embedded in accountability systems where offered Widely recognized 	<ul style="list-style-type: none"> Not typically administered in upper high school grades May include items related to statistics, data analysis, and problem solving, but not fully aligned with modern high school math courses, particularly advanced math
Data on course assessments, completion, and grades				
Course performance	Course grades, pass rates, and completion rates	<input checked="" type="checkbox"/> School or district administrative records <input type="checkbox"/> National or third party <input type="checkbox"/> Curriculum developer Widely available from school records	<ul style="list-style-type: none"> Easy to collect Reflects overall performance 	<ul style="list-style-type: none"> Limited comparability across sites because grading policies could vary (not standardized) May be complex to interpret and analyze because of lack of standardization
Mastery of course content	Knowledge checks, unit assessments, summative end-of-course assessments (e.g., final exams, applied capstone project, or performance assessment)	<input type="checkbox"/> School or district administrative records <input type="checkbox"/> National or third party <input checked="" type="checkbox"/> Curriculum developer	<ul style="list-style-type: none"> Aligned with course content Supports instructional and curricular adjustments Capstone projects capture transfer of learning Comparable over time (if consistent) 	<ul style="list-style-type: none"> Not comparable across courses and curricula Cannot be used with comparison group in efficacy study Potentially time-consuming to design and score

Future-Focused Measures for High School Math:




A Measures Guide for Curriculum Developers, Researchers, and Education Leaders



Construct	Potential measures	Source and availability	Pros	Considerations/limitations
Transfer or application of learning				
Civic reasoning	Civic Online Reading Database—Evaluating Data Assessment <i>Performance task</i>	<input type="checkbox"/> School or district administrative records <input checked="" type="checkbox"/> National or third party <input type="checkbox"/> Curriculum developer Available from Evaluating Data Civic Online Reasoning	<ul style="list-style-type: none"> • Strong real-world relevance for the application of statistics to everyday civic engagement • High comparability • Free 	<ul style="list-style-type: none"> • Requires creating an account on the platform and planning for and administering the assessment to students • Limited adoption outside civic literacy contexts
Data literacy	Critical Thinking Assessment Test (CAT) <i>Open-ended, short-answer test</i>	<input type="checkbox"/> School or district administrative records <input checked="" type="checkbox"/> National or third party <input type="checkbox"/> Curriculum developer Available from TN Technological University	<ul style="list-style-type: none"> • Externally validated • Sensitive enough to detect changes over a single course • Automatic scoring in online format 	<ul style="list-style-type: none"> • Not free • Validated with college, not high school (HS) students (although it has been used with HS students) • One test form taken pre/post
AI literacy	AI Literacy Concept Inventory assessment <i>20-item multiple choice</i>	<input type="checkbox"/> School or district administrative records <input checked="" type="checkbox"/> National or third party <input type="checkbox"/> Curriculum developer Available from Zhang, Perry, & Lee, 2024	<ul style="list-style-type: none"> • High internal consistency; validated with Item Response Theory • Measure grounded in research • High comparability • Free 	<ul style="list-style-type: none"> • Designed for middle and early high school grades • Items examining student understanding of specific AI models (e.g., decision trees) may be less relevant or applicable within mathematics-learning contexts.


Measuring student outcomes: Motivation, engagement, and persistence

A central hypothesis of the modern math initiative is that if teachers and students could access more engaging, relevant, and rigorous math offerings, students would be more motivated to learn, engage more in the content, and report increased interest in future STEM courses and careers. MEP measures refer to:

-  1. **Motivation:** the reasons why students engage in learning
-  2. **Engagement:** how much students participate and apply themselves to learning
-  3. **Persistence:** the extent to which students continue to engage in the face of near or long-term challenges




This section provides guidance on measuring specific MEP subconstructs. Specifically, it highlights measures related to student motivation (for example, utility value, expectancy value, and cost), engagement (for example, behavioral engagement, work completion, and attendance), and persistence (for example, interest in future pursuit of advanced math and STEM). It also offers measures related to identity, or whether students see themselves as a math or data person or capable of using and apply math and data across contexts (for example, math identity, data identity, and data agency).


Exhibit 3 shows measures for each construct, including the type of measure, example items, and their validity and reliability. Curriculum developers and researchers should use their discretion to select measures that align with their own theory of change and measurement needs.


Gaps in measures

Several of the MEP measures focus on math and may need to be adapted for courses such as data science and statistics and reassessed for reliability and validity.

Exhibit 3. Measures of student motivation, engagement, and persistence

	Construct	Measure	Type of measure	Example items	Validity and reliability (citation link)
 <p>Motivation</p>	Utility value	Math Mind Survey , utility value	Four-item student survey	"I think that learning math is useful for my future goals." "I think math is important to do because it can be useful in real life."	Validation study (internal consistency = 0.90)
	Expectancy value	EVC Scale , expectancy	Three-item student survey	"I believe that I can be successful in my [math or science] class." "I am confident that I can understand the material in my [math or science] class."	Validation study (internal consistency = 0.88)
	Cost	EVC Scale , cost	Four-item student survey	"I'm unable to put in the time needed to do well in my [math or science] class." "I have to give up too much to do well in my [math or science class]."	Validation study (internal consistency = 0.86)
 <p>Engagement</p>	Engagement	Student Engagement in Schools Questionnaire , behavioral effort and persistence ^a	Nine-item student survey	"When I'm in class, I participate in class activities." "I pay attention in class." "When I run into a difficult homework problem, I keep working at it until I think I've solved it."	Validation study (internal consistency = 0.85)
	Engagement	Time in minutes, number of visits, or number of distinct pages visited (per chapter or unit)	Curriculum platform data	On average, how many minutes per chapter are students engaged in the learning materials? On average, how often do students visit pages within a chapter?	n.a.
	Engagement	Class attendance	Administrative data	n.a.	n.a.
 <p>Persistence</p>	Completion	Proportion of completed pages or proportion of submits (per chapter or unit)	Curriculum platform data	The proportion (or percentage) of pages completed within a chapter The proportion (or percentage) of submits or attempts per chapter	n.a.
	Future interest, advanced math	Attitudes toward Mathematics Instrument , Questions 23, 28, 32–34	Five-item survey	"I am confident that I could learn advanced mathematics." "I would like to avoid using mathematics in college." "I am willing to take more than the required amount of mathematics." "I plan to take as much mathematics as I can during my education." "The challenge of math appeals to me."	Validation study (internal consistency = 0.88)

	Construct	Measure	Type of measure	Example items	Validity and reliability (citation link)
Persistence (continued)	Future interest, STEM ^b	Future interest scale	Three-item survey	"I look forward to learning more about math, statistics, or data science." "I want to take more math, statistics, or data science classes in the future." "I would be interested in having a job someday that involves things like math, science, statistics, or data science." <i>[items slightly adapted by curriculum developer CourseKata]</i>	Validation study (internal consistency = 0.86)
		Mathematical mindset or identity	Math Mind Survey , math identity	Four-item student survey	"I see myself as a 'math person.'" "Others see me as good at math."
 Identity	Data identity	Data Identity , data identification and positioning (Slide 14)	Nine-item student survey	"I see myself as a data person." "My teacher sees me as able to use data successfully." "Topics that involve data are exciting to me."	<i>Independent validation study not yet available</i> (internal consistency = 0.95 for full survey)
	Data agency	Data Identity , data agency and relationship (Slide 16)	10-item student survey	"I take initiative in using data to support my learning in school." "I view data as relating to all subjects, not just math and science."	<i>Independent validation study not yet available</i> (internal consistency = 0.95 for full survey)
	Data self-efficacy	Data Identity , confidence and efficacy with data (Slide 15)	Six-item student survey	"I can do well on schoolwork that involves data." "I feel confident in my ability to learn how to use data."	<i>Independent validation study not yet available</i> (internal consistency = 0.95 for full survey)
		Confidence in data science skills relevant to STEM	Four-item student survey	Students rate how confident (1–10) they are in their skills to: "Create a convincing argument using data as evidence" "Assess the reliability of claims made in online sources (e.g., social media, news sites, etc.)" "Interpret data presented in graphs, charts, and tables"	Validation study (internal consistency = 0.90)

^a The Student Engagement in Schools Questionnaire includes items on engagement and persistence.

^b This exhibit includes measures that capture MEP (motivation, engagement, and persistence) outcomes for high school students while they are still in high school. If planning for multiyear studies in the future, curriculum developers might consider exploring enrollment in a STEM major in college (beyond reporting future interest while in high school) as a measure of persistence.

n.a. = not applicable.

Measuring teacher outcomes

Understanding teachers' readiness to teach modern math course content and curriculum and their instructional practices in the classroom can (1) help curriculum developers and schools and districts improve professional learning and implementation supports and (2) help researchers and the education field understand what is required to successfully implement these courses.

To successfully teach a new or updated modern math course, teachers need to know the curriculum's lesson content and materials and develop Technological Pedagogical Content Knowledge (TPACK), or an understanding of the content, technology, and pedagogy that undergird the course and how these intersect with each other (Koehler & Misra, 2009). The TPACK framework can be used to inform professional learning and assess teachers' readiness to lead a course. Teacher surveys about short-term outcomes such as self-efficacy and their perceptions about teaching math are early indicators of longer-term teacher and student outcomes, such as teacher job satisfaction and student achievement ([Zee & Koomen, 2016](#)). Measuring teachers' perceptions can provide curriculum developers with valuable information that can support continuous improvement of their offerings, such as additional professional learning that teachers may require. The validated survey measures can also be used to assess teacher outcomes in an efficacy study.

In Exhibit 4, we include measures of teacher self-efficacy, mathematical mindsets, and self-reported teaching practice. This guide includes only teacher survey measures because they are often more feasible to implement than other types of measures, but curriculum developers and researchers may want to consider conducting classroom observations if they are interested in measuring pedagogical practice in the classroom.



Gaps in measures

Validated classroom observation measures and assessments of teachers' TPACK for modern statistics, data science, and data literacy do not yet exist. Several TPACK teacher measures of AI literacy have been developed, but they have not been tested or validated in the United States. Most of the teacher measures in this guide are related to mathematics or science and may need to be adapted to apply to statistics, data science, data literacy, or AI.

Exhibit 4. Teacher perception surveys

Construct	Measure	Type of measure	Example items	Validity and reliability (citation link)
Self-efficacy	National Survey of Science and Mathematics Education (NSSME+) Preparedness to teach math scale For the preparedness to teach math scale, see Question 25 in Appendix C, 2018 NSSME+ Mathematics Teacher Questionnaire	3-item teacher survey	"How well prepared do you feel to do each of the following in your mathematics instruction? Develop students' conceptual understanding for the mathematical ideas you teach."	Validation study Q. 23a–d; internal consistency = 0.82 Q. 24a–h; internal consistency = 0.79 Q. 25a–l; internal consistency = 0.84
	Mathematics Practices-Teaching Efficacy and Expectancy Beliefs Instrument (MP-TEBI), Self-Efficacy Scale For self-efficacy, see Questions 24, 28, 5, 22, 14, 31, 30, 10, 21, 1, 23, 15, 29, 19, 13, 26, 33, 12, 3, and 27	33-item teacher survey (with 20 self-efficacy items)	"I know how to prepare students to plan their own approaches to solving problems."	Validation study ; internal consistency = 0.91
Technological Pedagogical Content Knowledge ^a	Technological Pedagogical Content Knowledge Self-Efficacy Scale (TPACK-SeS)	52-item survey (with 6-item subscales associated with TPACK)	"I can use technological tools to determine students' misconceptions about science." "I can use technological tools to assess student learning of science." "I can apply my technological knowledge, content knowledge, and pedagogical knowledge all together to create an effective learning environment." "I can develop quality lesson plans using my technological knowledge, content knowledge, and pedagogical knowledge together."	Validation study ; internal consistency = 0.98
Mathematical mindsets	Mathematics Practices-Teaching Efficacy and Expectancy Beliefs Instrument (MP-TEBI), Outcome expectancy scale For outcome expectancy, see Questions 2, 4, 6, 7, 9, 8, 11, 16, 17, 18, 20, 25, and 32	33-item teacher survey (with 13 outcome expectancy items)	"Increased effort in math teaching produces little change in some students' math achievement."	Validation study ; internal consistency = 0.84
Pedagogical practice for math	Research on Curricular Alignment Partnerships (R-CAP), Ambitious instruction scale	Six-item teacher survey	"In the previous marking period, about how often did you engage in each of the following activities? Work on extended learning activities (e.g., projects or portfolios)."	Validation study ; internal consistency = 0.83

^a This survey was designed for STEM and not specifically for data science or math. Evaluators could adapt the items for a data science context; however, items would lose validity.

Measuring implementation

Usability, usefulness, and utilization are important indicators of successful implementation. We define these concepts in the following ways:

1. **Usability.** How easy or hard it is to implement or use a particular course offering or product.
2. **Usefulness.** Users' perceptions of the course offering or product's ability to meet their needs.
3. **Utilization.** The rate at which users take up the course offering or product or how much they use it.

These indicators can help curriculum developers understand whether their curricular materials, including professional learning materials, meet the needs of teachers and students and whether people use the materials enough for them to be evaluated and potentially scaled. Assessing these indicators may also provide useful information for understanding the quality of the curricular materials. The indicators could also be used along with the engagement measures introduced in the student outcomes section to assess students' and teachers' engagement with specific materials or products.

The specific methods used to measure usability, usefulness, and utilization will vary according to context, circumstances, and the implementation stage (Manley et al., 2023b). Approaches include survey instruments; qualitative data from observations, interviews, or focus groups; exit tickets; attendance or usage records (for example, metadata); or a combination of these. A single data collection method (for example, a survey or a focus group) may be able to capture multiple constructs (such as usability and usefulness). Exhibit 5 provides examples of the types of measures that organizations could use at the student or teacher levels to evaluate usability, usefulness, and utilization.

Exhibit 5. Examples of usability, usefulness, and utilization

Construct	Methods	Measures
Usability	<p>Conducting observations of how teachers and students explore or implement materials independently (for example, accessing coding platforms or using instructional materials)</p> <p>Gathering user feedback via surveys, interviews, or focus groups on ease of use and fit with current practice</p>	<ul style="list-style-type: none"> • Implementation challenge rate. Proportion of observed challenges or struggles teachers encounter during the implementation of the instructional materials. • Time. Time it takes to complete a discrete task. • Qualitative feedback. Perceived ease of use, barriers, and alignment with current practice.
Usefulness	<p>Gathering user feedback via interviews or focus groups on perceived usefulness and their rationale for answers</p> <p>Administering surveys to assess perceived usefulness or utility value</p>	<ul style="list-style-type: none"> • Qualitative feedback. Reported usefulness and rationale. • Survey items: <ul style="list-style-type: none"> – <i>Students:</i> Utility Value subscale from Math Mind Survey (see Table 1) – <i>Teachers:</i> Focus Grades/Classrooms Math Teacher Survey
Utilization^a	<p>Tracking usage by monitoring platform engagement, attendance, and task completion</p> <p>Administering surveys or gathering logs to determine take-up and completion</p>	<ul style="list-style-type: none"> • Take-up rate. Percentage of users who engage after access. • Usage. Time spent, log-ins, attendance. • Completion rate. Percentage completing full curriculum or core tasks; correlations with overall utilization.

^a Utilization measures may overlap with engagement and completion measures, but they answer different questions. For example, utilization measures may help reveal the extent of students' use of a solution, but engagement and completion measures may help answer whether students engage with the material.

Appendix A: Characteristics of survey measures in the guide

Exhibit A.1. Measures of the characteristics of student learning and achievement

Measure Characteristic	AP Statistics Exam	Levels of Conceptual Understanding in Statistics (LOCUS)	Civic Online Reading Database—Evaluating Data Assessment	Critical Thinking Assessment Test (CAT)	AI Literacy Concept Inventory assessment
Grades tested in prior research	High school (10th, 11th, 12th)	6th–8th, High school	6th–8th, High school	College	Middle school
Demonstrated reliability ^a and validity ^b	✓	✓		✓	✓
Demonstrated reliability and validity with broadly representative student populations ^c	✓	✓		✓	
Measure is available for use without restrictions on access			✓		✓

^a 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.

^b Scores represent what they intend to measure.

^c The measure has demonstrated reliability and validity with a study sample that included students who are Black or African American, Hispanic or Latine, Native American, and/or are in low-income or economically disadvantaged households OR the measure has been qualitatively tested (through focus groups, cognitive interviewing, or talk-aloud protocols) with students from these communities. The measure was modified to reflect findings from qualitative testing, and reliability and validity were reestablished on the modified version.

Exhibit A.2. Characteristics of student MEP measures

Measure Characteristic	Math Mind Survey	EVC Scale	Student Engagement in Schools Questionnaire (SESQ)	Attitudes Toward Mathematics	Future Interest Scale	Data Identity
Grades tested in prior research	5, 6, 7, 8	6, 7, 8, 11, postsecondary	7, 8, 9	6, 7, 8, 9, 10, 11, 12	Postsecondary	7 and 8
Full scale or subscale	<ul style="list-style-type: none"> Utility value subscale Math identity subscale 	<ul style="list-style-type: none"> Expectancy subscales Cost subscale 	Behavioral effort and persistence subscale	Self-confidence in mathematics	Full scale	<ul style="list-style-type: none"> Data identification and positioning Confidence and efficacy with data Data agency and relationship
Demonstrated reliability ^a and validity ^b	✓	✓	✓	✓	✓	✓
Demonstrated reliability and validity with broadly representative student populations ^c	✓	✓	✓			
Measure is available for use without restrictions on access	✓	✓	✓	✓		Pending release

^a 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.

^b Scores represent what they intend to measure.

^c The measure has demonstrated reliability and validity with a study sample that included students who are Black/African American, Hispanic/Latine, Native American, and/or are low-income or economically disadvantaged OR it has been qualitatively tested (through focus groups, cognitive interviewing, or talk-aloud protocols) with students from these communities. The measure was modified to reflect findings from qualitative testing, and reliability and validity were reestablished on the modified version.

Exhibit A.3. Characteristics of teacher measures

Measure Characteristic	National Survey of Science and Mathematics Education	Mathematics Practices-Teaching Efficacy and Expectancy Beliefs Instrument	Research on Curricular Alignment Partnerships (R-CAP)
Level tested in prior research (elementary/middle/high school)	Elementary, middle, and high schools	Elementary, middle, and high schools	Elementary, middle, and high schools (though mostly middle schools)
Full scale or subscale	Preparedness to teach math subscale	Self-efficacy subscale Outcome expectancy subscale	Ambitious instruction
Demonstrated reliability ^a and validity ^b	✓	✓	✓
Demonstrated reliability and validity with high school teachers in the United States	✓	✓	✓
Measure is available for use without restrictions on access	✓	✓	✓

^a 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.

^b Scores represent what they are designed to measure.

Future-Focused Measures for High School Math:

A Measures Guide for Curriculum Developers, Researchers, and Education Leaders

Appendix B: How we developed this guide

Advisory panel members

Kate Miller, research associate in data science education, Concord Consortium

Thema Monroe-White, associate professor of AI and innovation policy, George Mason University

Joseph Polman, associate dean for research and professor of learning sciences and human development, University of Colorado, Boulder

Helen Zhang, research associate professor of AI and STEM education, Lynch School of Education and Human Development, Boston College

This measures guide was developed as part of the Gates Foundation’s Modern Advanced Math Curricula for High School Body of Work (BoW). We compiled it using the foundation’s existing repository of MEP and teacher measures developed by Digital Promise, scans of the literature, and recommendations from measures experts and the Gates Foundation’s modern advanced math grantees. We collected feedback from the grantees (Skew The Script, CourseKata, Data Science for Everyone, and ThinkData Ed) on constructs of interest and potential scales. A council of external research experts on data science and statistics education, data literacy, and AI education reviewed the guide and provided feedback on the measures and the focus, usability, and structure of the guide.

We prioritized measures for this guide that aligned with curriculum developers’ measurement needs and would be feasible to implement. For the student outcome measures, we included measures that some grantees already use (for example, Levels of Conceptual Understanding in Statistics, or LOCUS) to ensure immediate relevance and feasibility, along with widely recognized measures (or types of measures) in the education field (such as end-of-course exams). In addition, we conducted a targeted literature scan to identify widely available or validated measures for artificial intelligence (AI) literacy and civic reasoning, which are priority areas for the grantees. For the student motivation, engagement, and persistence measures and the teacher measures, we prioritized survey measures over classroom observation measures because they are more feasible to implement. We included student and teacher survey measures that have demonstrated reliability and validity, have been tested with high school students or teachers in the United States, are publicly available at no cost, and have a small number of items feasible to administer in school contexts. We included some new measures that have not yet demonstrated reliability and validity because they focus on emerging areas that do not yet have established measures.

Appendix C: Gaps in Measures

Exhibit C.1 lists measures that are not currently available for modern math. These include measures that are under development, need to be adapted, or do not yet exist.

Exhibit C.1. Measures that have not yet been piloted or fully developed

Category	Construct	Details
Student outcomes: student learning and achievement	Conceptual understanding in data science	<ul style="list-style-type: none"> No national standardized assessments of data science currently exist. Data science assessment is currently under development by ThinkData Ed, and expected to be available for use beginning in the 2028–29 school year.
Student outcomes: motivation, engagement, and persistence	Motivation Engagement	<ul style="list-style-type: none"> Current measures focus on math or STEM and not specifically on data science, statistics, etc. The measures may need to be adapted and reassessed for reliability and validity.
Teacher outcomes	Teacher perceptions	<ul style="list-style-type: none"> Validated assessments of teachers' TPACK for modern statistics, data science, data literacy, and AI literacy do not yet exist but are currently under development.
	Instructional practice	<ul style="list-style-type: none"> Validated classroom observation measures for modern math courses do not yet exist.

TPACK = Technological Pedagogical Content Knowledge.

References

Student outcomes: Learning and achievement

- College Board. (n.d.). *AP Statistics exam*. <https://apcentral.collegeboard.org/courses/ap-statistics/exam>
- Digital Inquiry Group. (n.d.). *Evaluating data: Civic online reasoning assessment*. <https://cor.inquirygroup.org/curriculum/assessments/evaluating-data/>
- LOCUS. (n.d.). *LOCUS: Levels of conceptual understanding in statistics*. <https://locus.statisticseducation.org/>
- Tennessee Technological University Center for Assessment & Improvement of Learning. (2016). *Critical-Thinking Assessment Test (CAT) Technical Information (Version 8)*. https://www.tntech.edu/cat/pdf/reports/CAT_Technical_Information_V8.pdf
- Zhang, H., Perry, A., & Lee, I. (2025). Developing and validating the Artificial Intelligence Literacy Concept Inventory: An instrument to assess artificial intelligence literacy among middle school students. *International Journal of Artificial Intelligence in Education*, 35, 398–438. <https://doi.org/10.1007/s40593-024-00398-x>

Student outcomes: MEP

- Boaler, J., Conte, K., Cor, K., Dieckmann, J. A., LaMar, T., Ramirez, J., & Selbach-Allen, M. (2025). Studying the opportunities provided by an applied high school mathematics course: Explorations in data science. *Journal of Statistics and Data Science Education*, 33(1), 26–45.
- Hart, S. R., Stewart, K., & Jimerson, S. R. (2011). The Student Engagement in Schools Questionnaire (SESQ) and the Teacher Engagement Report Form-New (TERF-N): Examining the preliminary evidence. *Contemporary School Psychology*, 15, 67–79. <https://doi.org/10.1007/BF03340964>
- Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2014). A practical measure of student motivation: Establishing validity evidence for the Expectancy-Value-Cost Scale in middle school. *Journal of Early Adolescence*. Advance online publication. <https://doi.org/10.1177/027243161455689>
- Lim, S.-Y., & Chapman, E. (2013). Development of a short form of the Attitudes Toward Mathematics Inventory. *Educational Studies in Mathematics*, 82(1), 145–164. <https://eric.ed.gov/?id=EJ998478>
- Polman, J. L., Tran, T. C., & Miller, K. (2025). *Towards inquiry into data identity in interdisciplinary project-based learning* [Conference presentation]. Data Science Education K–12 Research to Practice Conference, San Antonio, TX, United States.
- Sinclair, C. H., Terrio, S. R., & Kim, Y. K. (2025). *The Math Mind Measures: A user guide*. SREC Reports, Louisiana State University. https://repository.lsu.edu/cgi/viewcontent.cgi?article=1015&context=srec_reports
- Tapia, M. (1996, November 6–8). *The attitudes toward mathematics instrument* [Conference paper]. Annual Meeting of the Mid-South Educational Research Association, Tuscaloosa, AL, United States.

Teacher outcomes

- Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 NSSME+*. Horizon Research, Inc.
- Desimone, L. M., Bell, N. S., Lentz, A., Hill, K. L., & Marianno, L. (2025). A holistic examination of how professional learning and curriculum relate to ambitious and culturally relevant instruction and student engagement. *AERA Open*, 11(4). <https://doi.org/10.1177/23328584241310429>
- Horizon Research, Inc. (2018). 2018 NSSME+ mathematics teacher questionnaire: Teacher background and opinions. <https://horizon-research.com/NSSME/wp-content/uploads/2018/12/2018-NSSME-MTQ-Formatted.pdf>
- Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education*, 9(1). <https://citejournal.org/volume-9/issue-1-09/general/what-is-technological-pedagogicalcontent-knowledge>
- Riggs, I. M., Fischman, D. D., Riggs, M. L., Jetter, M. E., & Jesunathadas, J. (2018). Measuring teachers' beliefs in relation to teaching mathematics with mathematical practices in mind. *School Science and Mathematics*, 118(8), 385–395.
- Zee, M., & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research*, 86(4), 981–1015. <https://doi.org/10.3102/0034654315626801> (Original work published 2016).

Implementation

- Hill, C. J., Scher, L., Haimson, J., & Granito, K. (2023). Conducting implementation research in impact studies of education interventions: A guide for researchers. (NCEE 2023-005). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from <http://ies.ed.gov/ncee>.
- Manley, M., Knechtel, V., Place, K., Gothro, A., & Ruggiero, R. (2023). *Guide to measuring implementation: The 3 Us—Usability, usefulness, utilization*. Mathematica.