Contents

The **Mathematics Empowerment through Standards-based Grading (ME-SBG)** project aims to address barriers to scaling an evidence-based intervention that has shown significant positive results, increasing math learning by nearly half a year—36% for geometry students and 45% for algebra students ([redacted], et al., 2023). Building on this prior success, ME-SBG will support implementation in over 70 schools across two states (Pennsylvania and Oregon), impacting more than 14,000 students in grades 6-9, including with more than 5,700 from historically underserved (primarily Black and Hispanic/Latino) student groups. We define high-needs students as those from historically underserved groups, specifically students of color or members of federally recognized Indian tribes, which is aligned with the goals of our school partners (Freeman & Bamat, 2022; Portland Public Schools, 2022),

ME-SBG distinguishes itself by integrating the crucial component of **reassessment for full credit after further study** into classroom grading systems, thus supporting student self-regulation and enabling more learning. This approach, bolstered by targeted teacher professional development and a curated library of resources, empowers students to take charge of their own learning, recognizing what they need to learn and understanding what actions to take to learn it.

ME-SBG is supported by moderate evidence (Absolute Priority 1), and addresses mathematics education (Absolute Priority 3, STEM). It also addresses the impacts of COVID-19 (Competitive Preference Priority 2). The COVID-19 pandemic has exacerbated an existing problem: gaps in prerequisite knowledge needed to master math standards, especially among historically underserved students (Deitz & Freyman, 2024). ME-SBG will perform asset mapping, collaborating with pilot school teachers to identify existing resources and pinpoint knowledge gaps intensified by the pandemic. Informed by what we learn, ME-SBG will offer: 1) **Diagnostic Tools** that identify gaps in prerequisite knowledge essential for understanding specific mathematical standards; 2) **Just-in-Time Support,** i.e., Targeted instructional materials to address these foundational gaps, enabling effective student progression; and 3)**Professional Development training** for educators to utilize diagnostic tools and integrate just-in-time support into their teaching.

HISTORY OF ME-SBG

Formative assessment uses assessments periodically to provide feedback about each student's progress, enabling *teachers* and *students* to adapt their teaching and learning strategies to help each student progress based on their individual needs. Numerous literature reviews have reported that formative assessment has important positive implications for student engagement, learning, and performance, especially for previously low-performing students (e.g., Black & Wiliam, 1998, 2010; Kingston & Nash, 2011).

Between 2004 and 2007, the Mathematics and Science Partnership of Greater Philadelphia (MSP-GP)—a consortium of 46 school districts and 13 institutions of higher education working to improve STEM education in grades 6-12—collaborated with formative assessment expert Dylan Wiliam to find effective ways to use formative assessment to improve adolescent STEM learning. Among various innovations, the most impactful was a program that empowered students to use assessment evidence to adapt their own learning strategies, integrating standards-based grading with reassessment for full credit after further learning(Clymer & Wiliam, 2006).

In 2007, the MSP-GP established the nonprofit **21st Century Partnership for STEM Education (21PSTEM)**, the lead applicant for this proposal, to continue its mission. The National Science Foundation (NSF) supported 21PSTEM in conducting a cluster Randomized Control Trial (RCT) to evaluate the impact of the formative assessment program described by

Clymer and Wiliam (2006) on eighth and ninth graders' algebra and geometry learning. This project, known as **Proficiency-based Assessment and Reassessment of Learning Outcomes (PARLO)** and described in more detail below, yielded positive results, with students in the treatment group achieving an additional 45% of a year's worth of algebra learning ([redacted] et al., 2023). Despite these positive outcomes, the PARLO program faced barriers to scale after NSF support concluded, hindering its growth and sustainability. The current ME-SBG project seeks to overcome these barriers, ensuring the successful, scalable implementation of especially effective, student-centered formative assessment practices in grades 6-9 mathematics classrooms.

A. SIGNIFICANCE

In 2022, only 26% of U.S. eighth graders performed at or above proficiency on the math portion of the National Assessment of Educational Progress, a decline of 7% from pre-pandemic levels in 2019. This drop erased almost half of the incremental gains made since 1990 (The Nation's Report Card, n.d.). State and district test scores mirrored this trend, with only a slight rebound in 2023. For instance, in Portland, Oregon, only 41% of eighth graders met state standards in mathematics in the spring of 2023, and statewide in Pennsylvania, just 26.1% achieved proficiency. Alarmingly, the proficiency rates were even lower for historically underserved groups. In Portland, only 8% of Black eighth graders and 19% of Hispanic/Latino eighth graders were proficient in mathematics. Similarly, in Pennsylvania, proficiency rates were just 13% for Black eighth graders and 18.7% for Hispanic/Latino eighth graders (Commonwealth Foundation, 2023; Oregon Department of Education, 2023).

These statistics highlight a critical and daunting challenge: if we aim for the majority of students from all groups to meet proficiency standards in mathematics, we must not only counteract the educational setbacks caused by COVID-19 but also significantly advance mathematics learning for American adolescents beyond historical norms.

Our proposed project, **ME-SBG**, tackles this challenge by empowering students in grades 6-9 to take active control of their mathematics learning. This age group is crucial because students begin to have a heightened need for status and respect (Yeager, 2017). A lack of empowerment in math can erode trust in teachers and lead to long-term negative impacts on motivation and learning, especially for students from historically underserved groups, who may already feel marginalized by the educational system (Liang et al., 2020). However, a program like ME-SBG, which focuses on formative feedback, high expectations, and scaffolding success, can address these challenges (Yeager et al., 2014b).

Our previous research in algebra classrooms indicates that the classroom assessment changes we propose can lead to an improvement of 0.33 standard deviations, equating to 45% of an additional year's growth in algebra knowledge ([redacted] et al., 2023). If such improvements accumulate over several years, we could enhance average student achievement by more than a full standard deviation, potentially increasing math proficiency rates in Pennsylvania to 60-75% (derived from Pennsylvania State Department of Education, 2023, Table 17-2).

ME-SBG aims to refine and simplify the implementation of our previously tested **PARLO** assessment system ([redacted] et al., 2023). Under ME-SBG, while students will receive summative scores on quizzes and assignments, such scores will not be used in a weighted average that determines final grades. Rather, the teacher will use such assessment evidence both to evaluate the extent to which a student is proficient in each of the course's learning outcomes *at that particular moment in time*, and then to provide students with personalized feedback designed to guide further learning. Students will then be given opportunities to do further work, at home or in school, and to be reassessed for full credit. In other words, summative assessments become formative tools designed to promote further learning, instead of merely yardsticks to measure what they have learned already.

Components of ME-SBG

1. Standards-based grading. In ME-SBG classrooms, teachers assign report card grades based on student proficiency and high performance on specific standards. ME-SBG classrooms in our study will include both of the following approaches to reporting these proficiencies, ensuring the program's adaptability to diverse grading systems:

- Direct Proficiency Reporting: Schools record standard proficiency levels directly on students' report cards, without converting them to traditional letter grades.
- Algorithmic Conversion: Schools use an algorithm to translate proficiency levels into traditional letter grades (A/B/C/D/F) on report cards.This approach calculates letter grades based on the number of standards on which the student is proficient and the number of standards on which the student demonstrates high performance, converting these into an overall course grade.

Schools using Direct Proficiency Reporting will include those from Portland and those recruited with the Aurora Institute (40-45 schools), while schools using Algorithmic Conversion will include those from Pennsylvania (39-44 schools). ME-SBG is designed to function effectively in both scenarios.

Regardless of the reporting method, effective implementation of ME-SBG requires adherence to the following three principles: i) **Grades Based on Standards Proficiencies**: Grades are determined by student proficiency and high performance on the defined standards. ii) **Transparency in Grading Procedures**: Grading methods must be clear and understandable to students and their parents or guardians. iii) **Separation of Non-Academic Factors**: Factors such as attendance, attitude, and homework completion are not averaged into the final grade. Instead, they are viewed as contributing to the end goal of understanding course content and are reported separately if necessary.

2. **Teachers share clear standards and rubrics for proficient and high performance.** Instruction under ME-SBG begins by clearly defining approximately 10 learning outcomes per semester, derived from state standards like the Common Core. These outcomes are communicated to students and parents to ensure a shared understanding of educational goals. For example, sixth-grade standards might include: "Fluently add, subtract, multiply, and divide multi-digit decimals," and "Understand and use the concept of a unit rate."

Teachers provide rubrics that outline criteria for both proficiency and high performance on each standard. Typically, there are two ways to achieve high performance: to solve challenging problems that require strategic thinking; or to demonstrate the ability to teach the material by helping a classmate move from not-yet proficient to proficient. While many modern standards-based grading programs establish criteria for both proficient and high performance (e.g., Marzano, 2010), this approach sets **ME-SBG** apart from older models, such as Mastery Learning (Bloom, 1968), which primarily aimed for proficiency (not high-performance) thresholds.

3. **Student-focused formative feedback. ME-SBG** emphasizes student-focused feedback, using various formative assessment techniques such as quizzes, exit tickets, and observational notes. This feedback is designed to guide students toward achieving proficiency and high performance on each learning outcome, thus fostering continuous improvement.

4. **Reassessment for full credit after further learning**. Student grades are not averaged over the semester; instead, students are rated on each learning outcome as not-yet-proficient,

proficient, or high-performance based on the best work they can show by the end of the semester. While students can reassess for full credit, before reassessment they must first engage in further learning, completing activities such as:

a. Error logs: Students create detailed explanations of their errors on problems, rework the problems, and articulate the strategies they to avoid similar mistakes in the future.
b. Remediation Plans: Students develop personalized contracts outlining specific activities and resources they will use to prepare for reassessment.
c. Flashback days: Scheduled in-class sessions allow students to revisit learning outcomes individually or in groups. These sessions provide structured opportunities to reinforce concepts and work toward achieving proficiency or high performance, fostering a collaborative learning environment.

By allowing reassessment after these learning activities, **ME-SBG** encourages persistent effort, self-reflection, and mastery. This approach transforms setbacks into learning opportunities, enabling students to continuously improve and achieve their best potential on each learning outcome/standard.

5. **Professional development**. The key idea behind ME-SBG is to use assessment and formative feedback to empower all students to take charge of their own mathematics learning; however, standards-based grading alone—merely assigning grades based on standards proficiencies—is insufficient for helping students to take on this charge. Rather, teachers must implement all the components listed above, and in addition they must help their students learn how to respond differently to assessment.

This assessment paradigm differs significantly from traditional classroom practices and from what most teachers experienced during their own education. To support teachers in making this shift, our program ensures comprehensive professional development. The program includes three days of summer training before the first implementation year, followed by monthly Professional Learning Community (PLC) meetings throughout the school year. Additionally, there is one day of refresher training in the second summer and continued monthly PLC meetings in the second school year. Administrator support is crucial; therefore, one day of summer professional development will be provided for principals and administrators from each participating school to enhance their understanding and support.

In case of teacher turnover, we will provide new teachers with the full three-day introductory training before their second year of implementation. Supported by experienced colleagues and ongoing PLC meetings,we expect new teachers to join their colleagues in implementing ME-SBG successfully.

A Promising New Strategy that Builds on and Moves Beyond Existing Strategies

ME-SBG integrates and builds upon established strategies for promoting mathematics achievement, including:

Mastery Learning. Mastery Learning, introduced by Benjamin Bloom (1968), laid the foundation for a system where students achieve mastery through formative assessments followed by targeted remediation and reassessment. Bloom envisioned that this approach could enable 90% of students to reach levels previously attained only by the top 10%. Despite its potential, Mastery Learning fell short in practical application due to the lack of detailed implementation guidelines, adequate teacher training, quality assessments, and remedial materials (Horton, 1979). Additionally, traditional Mastery Learning often imposed rigid progression rules, requiring students to achieve a specific procedural standard (e.g., 80% accuracy on a procedural task) before moving on, which limited student autonomy (Goodman, 1985).

ME-SBG builds on Bloom's principles by overcoming these limitations with comprehensive professional development and high-quality support materials. ME-SBG shifts from procedural mastery to an empowering framework where students advance through formative feedback, moving beyond proficiency to high performance and strategic thinking. Students actively participate in their learning, deciding when to revisit or progress. This focus on student empowerment marks a significant advancement over traditional Mastery Learning.

Formative Assessment. Formative assessment information can be used either by teachers or by students themselves. To date, most formative assessment interventions, especially in mathematics, have been **teacher-focused**. These interventions provide feedback to teachers about their students' current proficiency levels, enabling them to adapt their teaching strategies to meet individual student needs (e.g., Supovitz et al., 2018). In contrast, the ME-SBG program is **student-focused**, where assessments give *students* feedback about their own proficiency and errors, thereby empowering them to self-regulate and enhance their learning.

Other than our project, we are aware of only one extensively researched student-focused formative assessment intervention in mathematics: the **Assistments** program. **Assistments** provides students with immediate feedback on the accuracy of homework problems and has demonstrated a positive impact on mathematics learning, particularly for students with lower prior math scores (Rochelle et al., 2016; Murphy et al., 2020). This success underscores the potential of student-focused formative assessment to enhance mathematics learning and reduce achievement gaps. However, while **Assistments** focuses on homework feedback, the **ME-SBG** project differentiates itself by emphasizing in-class processes.

Standards-Based Grading. Standards-based grading (e.g. Marzano, 2010) is a necessary component of ME-SBG. However, our project is unique in providing teachers the professional development and support materials they need to implement the critical next step: *reassessment for full credit after further learning.* As explained in our Conceptual Framework, below, it is this step that scaffolds self-regulation and empowers students to take charge of their own learning.

B. STRATEGY TO SCALE

B.1 Strategies That Address Barriers To Scale

The main goal of this project is to overcome barriers to scaling our successful PARLO assessment system, creating ME-SBG, a new version that is positioned for sustainability and growth in grades 6-9 mathematics classrooms. We have identified seven barriers:

Barrier 1 - Lack of Software Integration. Our previous NSF-funded grant developed an online software product called *Tracker*, which participating teachers used to communicate with students and parents/guardians, reporting assessment evidence and proficiency for each standard. (For a screenshot, see Appendix J.1.) However, *Tracker* was not integrated into local Student Information Systems (SIS) or Learning Management Systems (LMS). This required teachers to double-enter assessment data, an unnecessary burden on teachers' time.

Solution Strategy. ME-SBG will integrate Tracker with district IT systems, enabling single sign-on and automatic data transfer. We will link Tracker with Canvas LMS (used by Portland Public Schools) and Schoology LMS (used by our Pennsylvania partners). Our software team will design adaptable interfaces for broader compatibility with other LMS/SIS systems, and create documentation and a manual to facilitate integration for newly recruited schools.

Barrier 2: Time Required to Develop Resources for "Reassessment for Full Credit After Further Study." The key feature that distinguishes our assessment system both from traditional classroom assessment and from commonly implemented standards-based report card systems is the emphasis on student empowerment through reassessment for full credit after additional study. While theoretically, many standards-based grading (SBG) systems advocate for this ideal, practical implementation often falls short due to the substantial resources required. Teachers need to provide clear rubrics for each standard and ensure high-quality assessment resources are available for repeated assessments. They also need to find or create study materials and supports for students' targeted learning needs, such as worked-out problems, peer tutoring guides, explanatory videos, or lesson plans for reteaching difficult concepts to small groups. In our previous study, highly committed volunteer teachers supported by financial incentives were able to develop the necessary resources; however, new teachers without similar support, experience, or time would struggle to replicate this effort.

Solution Strategy. We will develop a comprehensive resource library for each state standard, including rubrics, assessment tools, and study materials, aligned with the Common Core State Standards for Mathematics (CCSS-M). This library will be integrated into the Tracker program for easy access by teachers and students. Using a user-centered design process, our development team will work with pilot school teachers from Jan, 2025 until the RCT starts in fall 2026 to ensure these resources effectively supplement existing curricula.

Barrier 3 - Student difficulty understanding their new role. The new assessment system shifted expectations: students could no longer pass by merely attending, behaving, and completing homework; they actually had to demonstrate understanding of the material. Instead of moving on after failure, they could now relearn and improve. While some students excelled in this system, others found it confusing or stressful.

Solution Strategy. We will provide introductory materials/lessons for ME-SBG teachers to help students adapt. To highlight student empowerment and partnership in learning, these materials will be branded as “Professional Development for Students.”

Barrier 4 - Confusion and anxiety about grading. In schools using an A/B/C/D/F report card, students found it difficult to understand how their proficiency ratings would be translated into letter grades.

Solution Strategy. We will design, user-test, and perfect a student and family dashboard, integrated into our *Tracker* system. In districts that use a traditional (vs. standards-based) report card it will transparently calculate a student's letter grade based on the percent of rated standards that are "not-yet-met" , "proficient" or "high-performance" and let the student and their parents/guardians know precisely what they can work on to increase their letter grade.

Barrier 5 - Homework completion and class participation. Grades were no longer an extrinsic reward for class participation, homework completion, etc., resulting in some students being less willing to engage in these learning activities.

Solution Strategy. In the pilot year, we will test and refine two potential solutions by implementing them in pilot schools. Collaborating with participating teachers, we will evaluate their effectiveness and adjust based on teacher and student feedback:

- **Participation Grade**: Following Marzano's (2010) suggestion, some pilot teachers will assign a separate "participation" grade that reflects students' non-academic behaviors, such as class participation and homework completion.
- **Vetting Assignments:** Ensure homework and classwork genuinely support learning by consulting students on their usefulness and adjusting based on their input.

Barrier 6 - Community resistance. Some parents/guardians were resistant to a grading system that did not resemble the one they were accustomed to.

Solution Strategy. Previously we had not supported teachers in communicating with parents/guardians. We will develop new materials and resources teachers can use for community communication.

Barrier 7 - Inconsistent administrator support. In our previous work, administrative support for ME-SBG varied widely. Some principals hindered implementation by requiring teachers to include external summative test scores in student grades, even when students demonstrated mastery in later assessments. Others did not adequately support teachers when facing objections from community members about the grading changes. However, principals who understood and supported ME-SBG facilitated smoother implementation.

Solution Strategy. We will create a 1-day professional development session for principals and administrators to deepen their understanding of ME-SBG and gain their support for effective implementation.

B.2. Management Plan: Responsibilities, Milestones, and Timelines

The ME-SBG project will be managed by 21PSTEM in close collaboration with Portland Public Schools, Westat, Lincoln County PA Intermediate Unit (LCIU), Jim Thorpe Area School District, Interboro School District, and the Aurora Institute. The project will be structured in three sequential phases to achieve our objectives:

Phase 1 (Jan 2025 - Aug 2026): Development and Pilot Testing. Work with teachers from 3 schools in Portland, and 2 schools each in Jim Thorpe and Interboro school districts to develop materials and test them in classrooms. **Phase 2 (2026-27 School Year and 2027-28 School Year): Implementation and Evaluation.** Test program effects in a cluster RCT involving 70-80 schools, including the remaining 25 middle schools in Portland, 3 middle schools from Jim Thorpe and Interboro, plus additional schools recruited with assistance from LCIU and the Aurora institute. **Phase 3 (Post-RCT): Analysis, Dissemination, and Broader Implementation**. Analyze data, disseminate findings, and expand implementation to include former control schools.

Figure 1 displays a high-level project timeline and responsibilities by organization. On each row, the first organization listed is the lead. Appendix J3 provides a much more detailed management plan, with specific activities and responsible individuals identified.

Figure 1. High Level Project Timeline & Responsibilities

<table>
<tr><th colspan="4">Year 1: 2025</th><th colspan="4">Year 2: 2026</th><th colspan="4">Year 3: 2027</th><th colspan="4">Year 4: 2028</th><th colspan="4">Year 5: 2029</th><th rowspan="2">Responsible Organization</th></tr>
<tr><th>Q1</th><th>Q2</th><th>Q3</th><th>Q4</th><th>Q1</th><th>Q2</th><th>Q3</th><th>Q4</th><th>Q1</th><th>Q2</th><th>Q3</th><th>Q4</th><th>Q1</th><th>Q2</th><th>Q3</th><th>Q4</th><th>Q1</th><th>Q2</th><th>Q3</th><th>Q4</th></tr>
<tr><td colspan="21">Objective 1: Develop the training, digital tools, and other supports necessary to overcome adoption barriers for ME-SBG in Grades 6-9 math and Algebra I classrooms.</td></tr>
<tr><td colspan="6">Develop Support Library</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>21,D,P,A</td></tr>
<tr><td colspan="6">Develop software</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>21</td></tr>
<tr><td colspan="5">Plan PD</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>21,PA,P</td></tr>
<tr><td colspan="21">Objective 2: Implement ME-SBG with high fidelity across diverse schools in at least 2 states & evaluate its effectiveness for all students, particularly students from historically underserved communities.</td></tr>
<tr><td colspan="4">Develop Instruments</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>W</td></tr>
<tr><td colspan="6">Recruit&Randomize</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>W,P, A, PA, 21</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="8">Support teachers in implementation</td><td></td><td></td><td></td><td></td><td></td><td></td><td>21,PA,P</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="8">Collect Data</td><td></td><td></td><td></td><td></td><td></td><td></td><td>W</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="4">Analyze data</td><td></td><td></td><td>W</td></tr>
<tr><td colspan="21">Objective 3: Broaden the project's impact through dissemination and further implementation</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="6">Implement at Control Sch.</td><td>21,P,PA,D</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="6">Create & Distribute Info</td><td>21,PA,A</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="6">Plan next steps</td><td>21,W,P,PA,A</td></tr>
</table>

FIGURE 1 KEY (responible organizations)
21: 21PSTEM; A: Aurora Institute; D: Districts (Interboro and Jim Thorpe); P: Portland; PA: Pennsyvania IU 12; W: Westat

B.3 Applicant's Capacity to Bring the Project to Scale

Our partnership has the capacity to bring ME-SBG to scale at a regional level. The lead applicant, 21PSTEM, managed the IES-funded National Research & Development Center on Cognition in Science, recruiting, providing treatment for, and studying effects in a cluster RCT involving 180 middle schools in Pennsylvania and Arizona ([redacted], et al., 2024). Our independent evaluator, Westat, is a national leader in educational research, data collection and analysis, evaluation, and communications. Our partners include suburban and rural districts in Pennsylvania (Interboro and Jim Thorpe). The urban Portland Public Schools in Oregon has committed to having 28 schools participate in our project. Lincoln County IU will lead our collaboration with the Pennsylvania Department of Education to recruit 35-40 additional Pennsylvania schools, and the Aurora institute will leverage its national network to recruit an additional 15-20 schools nationwide.

The project team is uniquely qualified to carry out the proposed work. The team brings extensive expertise and experience in mathematics curriculum and instruction, formative assessment, standards based grading, professional learning, software development, and rigorous mixed-methods research and evaluation. Appendix J2 provides a detailed description of key personnel and their responsibilities.

B.4 Dissemination for Further Development and Replication

Our partnership will employ a multi-faceted approach to broadly disseminate project information and support the development and replication of ME-SBG.

Conferences and Convenings: We will present ME-SBG research and implementation experiences at major national conferences such as AERA, NCTM, the National Conference on Student Assessment by CCSSO, and the ASCD-ISTE joint conference. Additionally, regional conferences like the Pennsylvania Educational Technology Expo and the Aurora Institute's annual Symposium, which attracts over 1,000 educational leaders, will feature sessions on ME-SBG findings.

Journal Articles: Our team, including [redacted], [redacted], and the Westat evaluation team, will publish at least two articles in peer-reviewed journals to share insights and outcomes from ME-SBG, contributing to the academic discourse on standard-based grading, student empowerment and self-regulation, and their impact on mathematics education.

Social Media Engagement: 21PSTEM will create a project-dedicated online platform for dissemination. This will include infographics, podcasts, and webinars. Additionally, the

Aurora Institute will leverage its digital presence—reaching nearly 30,000 social media followers and 10,000 email subscribers—to distribute project findings. The Aurora Institute's highly regarded weekly CompetencyWorks blog, with 57,000 annual visitors, and the Institute's website, with 140,000 annual visitors, will also feature project updates.

Earned Media: The Aurora Institute will leverage its earned-media expertise (with established journalist relationships and over 50 earned media placements annually) to implement a strategic earned media campaign, including securing op-eds and thought pieces, amplifying the project's impact.

Direct Scale-Up: After completing the RCT in the spring of 2028, we will support the scale-up of ME-SBG to the 35-40 control group schools, encompassing approximately 7,000 students. These schools will receive the same support as the treatment group, including access to materials, professional development, and monthly Professional Learning Community sessions through December 2029.

B.5 Utility of Products that will Result from the Proposed Project

The project will produce six essential products to facilitate the implementation of ME-SBG:

1. **Library of Classroom Support Materials:** Comprehensive resources aligned with each Common Core standard, featuring rubrics for proficiency and high performance, just-in-time instruction for prerequisite skills, assessment and reassessment materials, and supplementary resources such as videos, peer teaching guides, worked sample problems, and teacher-led lessons.
2. **Updated Tracker Program:** A user-friendly tool linking standards to appropriate library materials. For schools that need to do so, the program will also transparently translate standards-based ratings into report card letter grades.

3. **LMS Integration:** Integration of Tracker into widely used LMS (Canvas and Schoology), plus software and a user manual designed to facilitate integration into other LMS platforms.
4. **Student and Parent Support Materials:** Engaging introductory lessons for students ("Professional Development for Students") and resources for communicating with parents and guardians, emphasizing their roles as active partners in students' educational journeys.
5. **Professional Development Manuals:** Detailed guides for conducting summer professional development for teachers and administrators and facilitating Professional Learning Communities (PLCs) to support teachers in implementing ME-SBG effectively.
6. **Evaluation and Research Documentation:** An independent evaluation of the effectiveness of ME-SBG, alongside articles published in peer-reviewed journals documenting the program's impacts and outcomes.

All materials will be accessible to the public free of charge on the project-dedicated online platform outlined in our dissemination strategy. These resources empower school districts to assess the impact of ME-SBG and enable districts following the Common Core State Standards for Mathematics (CCSS-M) to implement the program effectively.

After analyzing impact findings, we plan to seek additional funding to expand the impact of the ME-SBG program:

- Adapting the Library of Support Materials for states that did not adopt CCSS-M, partnering with districts to secure funding for customization.
- Exploring extensions of ME-SBG to older student cohorts and/or other subject areas, building on successful pilots in areas like seventh-grade science (Clymer and Wiliam,

2006).

- Investigating the potential benefits of integrating ME-SBG with interventions designed to enhance student motivation (e.g., Yeager, et al., 2014a) leveraging insights from educational psychology to amplify positive educational outcomes.

These initiatives aim to maximize the utility of ME-SBG across diverse educational contexts and enhance its long-term sustainability and impact.

C. QUALITY OF THE PROJECT DESIGN

C.1 Conceptual Framework

Our logic model (Appendix G) presents ME-SBG's conceptual framework. This framework includes key program components: a library of classroom materials, the updated Tracker program, integration with Learning Management Systems (LMS), and support materials for students and parents. Combined with professional development, these components enable teachers to implement ME-SBG effectively, overcoming barriers identified in our previous work.

Teachers using ME-SBG share clear standards and rubrics for proficient and advanced performance, provide formative feedback, and encourage students to respond through further study and reassessment for full credit. Standards-based grading reflects what students have learned by semester's end, without penalizing early struggles.

This system restructures classroom assessment so that it promotes and scaffolds self-regulated learning. Effective self-regulation involves self-reflection, where students identify goals, monitor and evaluate their performance, understand the reasons behind their results, and determine steps for improvement (Zimmerman et al., 2017). This process of ongoing monitoring and adjustment has been shown to significantly enhance student achievement (Cleary, 2020).

ME-SBG supports self-regulation by providing a clear pathway: learning goals are clearly defined, formative feedback informs students about their current proficiency levels, and reassessment opportunities allow them to address gaps in understanding and demonstrate improved performance. This structured process of goal setting, feedback, further learning, and reassessment enables students to continuously monitor their progress, set goals based on standards, and adapt their learning strategies. Our logic model reflects that through self-regulation, students can learn more with ME-SBG.

Additionally, ME-SBG positively impacts student motivation, a second mediator of program effects. Qualitative evidence ([redacted], 2023) suggests the program enhances students' relatedness, trust in teachers, peer relations, and sense of competence in math, thus boosting their motivation to learn (Ryan & Deci, 2009). It also promotes growth mindsets—the belief that mental ability is not fixed, but instead grows through hard work (Dweck, 2007)—and mastery goals, the desire to learn as much as possible from schoolwork (Senko, 2016). Motivated students with growth mindsets and mastery goals are more likely to leverage self-regulation opportunities provided by ME-SBG, amplifying the positive effects on learning.

In summary, ME-SBG’s emphasis on reassessment for full credit after further learning directly supports self-regulated learning by creating a structured environment where students can reflect on their progress, address gaps, and continually improve. This approach, coupled with enhanced motivation and a growth mindset, leads to better academic outcomes.

C.2 Goals, Objectives, and Outcomes

The goal of the project is to enhance mathematics proficiency among grades 6-9 students, particularly those from historically underserved communities, through implementing the ME-SBG system. This involves addressing and overcoming previously identified barriers to scale by providing comprehensive teacher training, deploying effective digital supports, and evaluating program impact. Appendix J.4 provides a summary of the project's objectives, strategies, outcomes, and measures.

C.3. How the Design Addresses the Target Population's Needs

Support for Students

As discussed in Section A, adolescents, especially those from historically underserved groups, have a critical need for recognition and respect within educational environments. Traditional approaches often fall short in meeting these needs, leading to significant learning challenges for these students. Additionally gaps in prerequisite knowledge, exacerbated by the Covid-19 pandemic, affect all students, especially those from historically underserved groups. The ME-SBG project directly addresses these needs. Our previous rigorous cluster RCT with 29 schools and over 2,500 algebra and geometry students demonstrated the success of this intervention ([redacted] et al., 2023).

Empowerment Through Reassessment for Full Credit: ME-SBG empowers students by allowing them to reassess for full credit after additional learning, fostering a growth mindset and self-regulation–thus enabling them to take charge of their own mathematics learning.

High Expectations and Respect: Crucially, ME-SBG respects all students by maintaining high expectations. The program encourages students not only to achieve proficiency in each standard but to strive for high performance. By setting high standards and providing the tools to meet them, ME-SBG fosters a culture of respect and achievement, particularly benefiting historically underserved groups.

Addressing Gaps in Prerequisite Knowledge: For each standard ME-SBG will include diagnostic tools to identify gaps and provide just-in-time support tailored to the specific needs of students. Professional development will focus on strategies to effectively integrate these tools into everyday teaching practices.

Support for Teachers

To scale ME-SBG effectively and meet student needs, we must also address teacher needs. Through extensive surveys, interviews, and informal discussions with teachers implementing our intervention, we identified key barriers to implementation as detailed in Section B1. This project aims to overcome these barriers by providing necessary support and resources to teachers, thus enabling them to better support their students.

E. PROJECT EVALUATION

E.1. Methods Designed to Meet WWC Standards Without Reservations

Westat will conduct an independent impact study to address the research questions (RQ) shown in Table 1 that are aligned with the ME-SBG logic model (see Appendix G). This mixed-methods study will assess the program's impact on grade 6-9 students' math achievement, and the extent to which student self-regulation and motivation-related mindsets mediate this impact. There are five confirmatory (indicated by an asterisk) and five exploratory research questions. The exploratory questions are geared to understand subgroup differences, mediation effects, moderation effects, and what items are pertinent for replication or further study. The table also shows the primary outcome of interest (see Appendix J.8) and the WWC domain for outcomes that are eligible for review (WWC, 2023).

Table 1. Research questions (RQ) and type

Research question	Primary Outcome	WWC Domain
1. ***Intervention and implementation fidelity:** Did the ME-SBG project operate as intended (i.e., meet the threshold of acceptable implementation)? a. Did the actual implementation adhere to the planned implementation? b. Did teachers participate in the intervention at expected levels?	Met fidelity thresholds	N/A
2. ***Treatment contrast:** As implemented, did the ME-SBG program features	Met fidelity thresholds	N/A

Research question	Primary Outcome	WWC Domain
represent a substantially different experience from the status quo?		
3. ***Impact for historically underserved student groups:** What is the impact of ME-SBG on academic performance in mathematics for students from historically underserved groups?	State-provided standardized test scores	Mathematics Achievement
4. Impact for all student groups: What is the impact of ME-SBG on all students' academic performance in mathematics class?	State-provided standardized test scores	Mathematics Achievement
5. Impact by grade level: What is the impact of ME-SBG on students' academic performance in mathematics class by grade level?	State-provided standardized test scores	Mathematics Achievement
6. **Mediation:** How is the impact of the ME-SBG project on students' academic performance in mathematics class mediated by self-regulation and motivation-related mindsets (i.e., culture of peer support, trust in teacher, growth mindset and mastery goals; sense of competence)?	State-provided standardized test scores	Mathematics Achievement
7. **Moderation:** How is the impact of the ME-SBG project on student achievement in mathematics moderated by school, teacher, and student characteristics?	State-provided standardized test scores	Mathematics Achievement
8. ***Costs:** What are the costs of operating the ME-SBG project?	Incremental cost	N/A
9. ***Cost effectiveness:** Is ME-SBG more cost effective than the status quo?	Cost effectiveness ratio	N/A
10. **Replication:** What, if any, strategies are suitable for replication or testing in other settings?	Varies	N/A

The study will use a school-level, clustered RCT design to meet What Works Clearinghouse Standards Without Reservations (WWC, 2022). Westat will obtain IRB approval of the study design and all protocols prior to use (see Appendix J.3. for the evaluation timeline).We will propose to the IRB that we obtain informed consent from parents for participation in the student survey through an opt-out process. No students will be identified in any reports.

Study Conditions and Random Assignment. We propose to conduct a cluster RCT with schools as the unit of assignment because ME-SBG entails teacher changes to practice under coordinated instructional leadership. Participating schools will include a variety of grade-level organizations (some grades 6-8; others grades 9-12). Schools will be randomized (by blocking on geography and grade level organization) to either implement ME-SBG and receive associated professional development or engage in business-as-usual instructional activity. For the ME-SBG intervention, all teachers instructing grade 6-8 mathematics or 9th grade algebra will implement ME-SBG for two years; teachers in comparison schools will be introduced to ME-SBG after the intervention is concluded in treatment schools. Westat staff will perform random assignment using a 50/50 assignment ratio, collect all outcome data, and conduct impact analyses.

Sample size and power calculations. A two-level model will be used to assess the effect of ME-SBG on student outcomes. For the primary confirmatory impact analysis, we assume standard power analysis thresholds (power [1-β] = 80%, $p < 0.05$, two-tailed test), 72 schools in the study, an average of 80 students from historically underserved groups per school (assuming a full population of 200 students per school), 12 schools per block, and 6 blocks. Assuming an intra-class correlation of 0.15, an equal number of ME-SBG and comparison schools, a level-one R^2 of 0.49 (based on prior spring state standardized student assessment scores), and a level-two R^2 blocking covariate of 0.12, the minimum detectable effect size would be 0.214. In contrast, [redacted] et al. (2023) detected a 0.33 effect using a variant of ME-SBG. Appendix J.5 provides additional power analysis details and scenarios that demonstrate a sample size of approximately 72 schools, each having an average of 80 students, will be adequate under several conservative assumptions. We expect more than 14,000 students across treatment and control conditions, with about 5,700 of these students representing historically underserved groups.

Strategies to guard against sample attrition. Low attrition rates (especially differential attrition between treatment and control groups) are important to ensure that the evaluation meets WWC Standards Without Reservations. We will use these strategies to guard against **cluster (school) attrition**:

- During recruitment, we will be explicit in communicating expectations to obtain administrator and teacher buy-in.

- We will over-recruit schools by 4 to 11 percent (or 3 to 8 additional schools) to allow for unanticipated school recruitment challenges and school attrition.
- The wait-list control group study design allows all schools in the study to receive the intervention, which will help with sample retention of the control schools.

Under the assumption of 4-11 percent school attrition and given that the intervention is not expected to affect school attrition, differential attrition should be well below the optimistic boundary of 10.4-10.9 percentage points (WWC, 2022). ***Teacher attrition.*** The use of a two-year teacher treatment window raises the prospect that teacher turnover between school years could present a study threat. However, as discussed in the program section, the intervention is not contingent on having the same teacher over the course of the study period and training will be provided to new teachers. Furthermore, teacher turnover should not result in differential school attrition in the context of this RCT (and student state mathematics assessment data would still be available for the ITT impact analysis). ***Student attrition.*** Similarly, we do not expect differential student attrition based on 21PSTEM's prior RCT experience of the PARLO intervention ([redacted] et al. 2023). Furthermore, the ITT analysis will be conducted with math scores of students who leave schools but remain within the state.

Missing data. We will use maximum likelihood procedures following requirements specified in the WWC Handbook (WWC, 2022) to supplement complete case analysis.

Joiners. Our primary impact analysis will focus on the student level as we test the impact of a professional development program in a school-level RCT. WWC standards establish that the risk of bias due to student joiners is low (WWC, 2022).

Contamination. ME-SBG's PD is not easily transmitted without the summer training and monthly PLCs. The teacher PD literature has consistently found that professional learning without sufficient duration and collective participation is unlikely to have any impact on instructional practice (Garet et al., 2016; Yoon et al., 2007). ME-SBG is also supported by an online (password protected) library of resources (e.g., rubrics, assessment materials, teaching resources). Furthermore, contamination concerns were not detected in the prior PARLO study ([redacted] et al. 2023). Nevertheless, we will (1) instruct school staff not to share ME-SBG resources, (2) have study staff monitor schools for any use of ME-SBG resources by teachers in control schools, and (3) examine teacher surveys and select teacher interviews (to be determined through criterion-based sampling) for evidence of contamination. The study team will work with schools and 21PSTEM team to mitigate any concerns.

Non-compliance with randomization. It is theoretically possible that schools assigned to the control condition may determine that they want to implement ME-SBG, nonetheless. While this will be difficult to do without the support components and core activities as discussed above, we are prepared to monitor for and address noncompliance. If noncompliance is suspected, we will estimate complier average causal effects (CACE) as a sensitivity check, as this represents the average effect of taking up ME-SBG among compliers.

Baseline equivalence. We expect random assignment will generate equated groups; we will assess baseline equivalence and account for baseline differences in impact estimation if needed.

Treatment contrast. Treatment contrast refers to the **difference in receipt** of ME-SBG components given access to the program (for schools in the program group) and without access to the program (for schools in the control group) (Weiss et al, 2014). The greater the treatment contrast, the more likely it is that the intervention differs from the status quo condition. To answer RQ 2, we will assess the conditions at both program and control schools through teacher and student surveys; and through semi-structured interviews with a small number of teachers each year (the same teachers selected to contribute to the implementation fidelity analyses).

Quantitative data analyses. We will use a two-level multilevel hierarchical model that factors in student- and school-level covariates and school-level random effects to estimate the impact of ME-SBG on student outcomes. Appendix J.6 specifies the statistical models for RQs 3, 4, 5, 6, and 7. We will conduct an ITT analysis as the primary approach supplemented by treatment-on-the treated analyses as exploratory analyses. We will examine state end-of-year standardized mathematics assessment scores (obtained from the state) as our primary outcome measure, using standardized scores in impact analyses (see May et al., 2009). This measure of mathematics achievement is part of the academic readiness, knowledge, or skills domain eligible for review by the WWC according to the Study Review Protocol, v5.0 (WWC, 2023). RQs 4 and 5 focus on mediators and moderators, respectively, of the impact of ME-SBG and will involve multilevel modeling with cross-level interactions (Krull & MacKinnon, 2001). Appendix J.6 provides the estimation models for the multilevel mediation and multilevel moderation analyses.

Cost and Cost-effectiveness analyses. To answer RQ 8 on costs, we will record all resources used to implement ME-SBG in order to 1) further describe what is needed to implement the intervention, 2) complement the fidelity of implementation research, 3) help potential adopters/adapters decide whether resources are better spent going to ME-SBG program or another endeavor; and to 4) help others determine what is needed to replicate the intervention (Levin & Belfield, 2015; Hollands & Levin, 2017). To estimate costs, we will use the ingredients method (Levin et al., 2018) to document the costs of ME-SBG implementation (i.e., cost per teacher, cost per student), as well as how costs vary across sites. One of the benefits of a RCT is the attainment of causal impact estimates, which can represent a measure of effectiveness. To answer RQ 9, we will conduct a cost-effectiveness analysis by dividing costs by a measure of effectiveness. The resulting cost-effectiveness ratio (CER) will allow practitioners and policymakers to compare programs by comparing CERs, enabling the selection of an alternative that provides the best results relative to costs. See Appendix J.7 for additional details.

Replication analyses. To answer RQ 10, we will conduct a series of exploratory analyses geared to understand what, if any, strategies are suitable for replication or testing in other settings. For example, the ME-SBG theory of change posits that an important component is having student grades be transparent and based on students' proficiency/advanced status on standards. However, it could be important that the report card itself is A/B/C/D/F or standards based. Since we expect to have about half the schools in the study providing A/B/C/D/F report cards and the other half doing standards-based report cards, we will explore whether ME-SBG is equally effective in both types of report-card settings. This may warrant future replication.

E.2. Generation of Guidance About Effective Strategies Suitable for Replication

The impact study will primarily work with schools located in Portland, Oregon and Pennsylvania. The schools will vary in size and diversity with respect to faculty and student populations. Having a **large and diverse evaluation sample** representative of multiple states across middle school grades 6 to 8 and grade 9 will generate information that can guide schools and districts on whether, and how, to implement ME-SBG. Guidance to be generated will cover 1) how ME-SBG fits into different school contexts and what supports are necessary, 2) how ME-SBG can be implemented with fidelity, 3) the impact of ME-SBG in different settings and for different populations, and 4) the costs to implement ME-SBG.

Westat will use qualitative data to understand whether ME-SBG was implemented by schools as 21PSTEM staff intended (i.e., whether schools met the thresholds for acceptable implementation; see Appendix J.9 for thresholds). Data collected from document reviews, semi-structured interviews with about 10 percent of teachers (selected using criterion-based sampling to have representation by grades, geographic areas, and current report card type), and observations of training will be analyzed to assess whether actual implementation mirrors the planned implementation (Weiss et al, 2014). RQ 1 will be examined using a qualitative lens and descriptive and multivariate analyses, ranging from correlations and cross tabulations to ordinary least squares regressions and multilevel analyses, to understand differences in patterns of ME-SBG operation.

In addition to assessing the implementation fidelity (RQ 1 and 2) of the ME-SBG training component, both the quantitative and qualitative **implementation data** will help refine the training and the resources/materials used to support teachers. Our exploratory impact analyses examining the impacts on all students (RQ 4) and students by grade (RQ 5), as well as our **moderator** analyses (RQ 7) will help us describe any **differential impacts** of ME-SBG across populations and settings. Examples of other moderators include student motivation, teacher buy-in, administrator support for changed grading practice, and teacher turnover. The **cost effectiveness analysis** (RQ 9) will help inform districts and schools about the cost of ME-SBG to achieve certain effects for students and how those costs compare to the expense of the business-as-usual condition PD. In **developing guidance about effective scale-up strategies** (RQ 8), we will frame our findings and discussion to address four conceptualizations of scale: 1) adoption, or the potential for widespread use of ME-SBG, 2) adaptation, or the widespread use of a modification of ME-SBG to meet the needs of local users, 3) replication, or widespread replication of ME-SBG with high fidelity that produces expected outcomes, and 4) reinvention, or the creation of a new intervention that mixes parts of ME-SBG with other components (Morel et al, 2019). This framework will allow us to take a more nuanced and reform-focused perspective on what it takes to promote the depth, sustainability, spread, and local ownership of ME-SBG implementation.

E.3. Clear Articulation of Components, Mediators, Outcomes, and Measurable Thresholds

The ME-SBG logic model (see Appendix G) informs the design of the impact evaluation. Appendix J.8 provides additional detail about all measures we intend to use, including the role, the target respondent, the measurement approach, when we will assess the measure during the study, and the member of the team that is responsible for developing the measure. **Key components of ME-SBG** include 1) standards-based grading, 2) teachers share clear standards and rubrics for proficient and high performance, 3) students are provided formative feedback, 4) students are reassessed for full credit after further learning, and 5) teachers receive initial and ongoing PD. Several measures will be examined to assess implementation fidelity including principal support for ME-SBG, implementation compliance, summer training attendance, and PLC attendance. **Minimum thresholds for acceptable implementation** are: 1) principals agree to survey statements about support (see Appendix Table J.9.1 for statements); 2) 75% of math teachers in the targeted grades agree that they provide students time and resources for further study in response to feedback and 75% agree that they encourage reassessment for full credit after further study; 3) 75% of math teachers in target grades at the school attend all three days of summer PD training; and 4) 75% of math teachers in target grades at the school attend 4 or more PLC sessions.

The **key student outcome** is **student mathematics achievement** (RQs 3, 4, and 5) as measured by state assessments, such as Pennsylvania's System of School Assessment (grades 3 to 8) and Keystone Exam (grade 9) and Oregon's Statewide Assessment System.

There are three **key moderators**. These include **school report card type** (the report card itself is A/B/C/D/F or standards based) as measured by direct observation with low inference or based on school records; **student demographics** as measured by school records; and student-level **prior year test scores**, which will be obtained through the state. Other moderators may be examined, but not be reported in the main text of the evaluation report and included in the report's appendix.

Key mediating variables for student outcomes are students' self-regulation and motivation-related mindsets. **Student self-regulation** (RQ 6) will be assessed by self-report, using an adaptation of the Likert-scaled items from the Maladaptive Regulatory Behavior Scale (Cronbach α=0.72) of the Self-regulation Strategy Inventory (Cleary, 2006). Validated for ninth and tenth grade students taking science classes in urban high schools, the scale measures the use of ineffective regulatory behaviors such as the avoidance of situations in which learning is difficult. The questions for this scale will be adapted by the project leadership team and tested as described in Appendix J. 8. **Motivation-related mindsets** (RQ 7) will be measured using five separate scales (in accordance with the logic model in Appendix G). *Culture of peer support* will be measured using the Student Social Support Index (Cronbach α=0.84), which quantifies belonging and classmates' interest in caring (Middleton et al, 2022). *Trust in teacher* will be measured using the School Trust Scale (Cronbach α =0.80) (Yeager et al., 2014). *Growth mindset* will be measured using the Field-Specific Ability Beliefs scale (Cronbach α=0.87-0.90) (Leslie et al, 2015). *Mastery goals* will be measured using the Mastery-Approach subscale of the Achievement Goal Questionnaire-Revised (Cronbach α=0.84) (Leslie et al, 2015). Finally, *Sense of competence* will be measured using the Mathematics Expectance subscale of the Expectancy-Value-Cost (EVC) scale (coefficient omega=0.88) (Kosovich et al, 2015).

References

Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society. Series A (General)*, *149*(1), 1–43. http://doi.org/10.2307/2981882

Baron, R. and Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6):1173–1182.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods, 11(*2), 142–163.

Belfield, C. R. (2014). Cost-benefit analysis and cost-effectiveness analysis. In *Handbook of Research in Education Finance and Policy* (pp. 141-156). Routledge.

Black, Paul, and Dylan Wiliam. *Inside the black box: Raising standards through classroom assessment*. Granada Learning, 1998.

Black, P., & Wiliam, D. (2010). A pleasant surprise. *Phi Delta Kappan*, *92*(1), 47-48

Blackwell, L. S., Trzesniewski, K. H. & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78(1), 246 – 263.

Bloom, B. S. (1968). Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation Comment*, *1*(2), 1–12. https://eric.ed.gov/?id=ED053419

Cleary, T. J. (2006). The development and validation of the self-regulation strategy inventory—self-report. *Journal of School Psychology*, 44(4), 307-322.

Cleary, T. J. (2020). Core Components and Empirical Foundation of the Self-Regulation Empowerment Program (SREP) in School-Based Contexts. *Student Engagement: Effective Academic, Behavioral, Cognitive, and Affective Interventions at School*, 281-292.

Clymer, J. B., & Wiliam, D. (2006). Improving the way we grade science. *Educational Leadership*, *64*(4), 36.

Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32*(6), 3–12. Common Core Standards Writing Team. (2011). Progressions for the common core state standards in mathematics. *Tucson, AZ: University of Arizona Institute for Mathematics and Education. Downloaded from* *https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://achievethecore.org/file/1173&ved=2ahUKEwjhkMnslPiGAxVrD1kFHccaDMgQFnoECBQQAQ&usg=AOvVaw3pJKf539BOblmwfiLEBkyu* *on June*, *25*, 2024.

Commonwealth Foundation (2023). *2023 PSSA Scores Highlight the Need for Educational Opportunity*. Downloaded June 23, 2024 from https://www.commonwealthfoundation.org/research/2023-pssa-scores/ .

De Leeuw, J., & Kreft, I. G. (1998). *Introducing multilevel modeling*. Sage.

Dietz, S. & Freyman, C. (2024) *The State of U.S. Science and Engineering 2024*. Alexandria, VA: National Science Board. Download July 1, 2024 from https://ncses.nsf.gov/pubs/nsb20243/key-takeaways

Dong, N. and Maynard, R. A. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and sample size requirements for experimental and quasi-experimental designs. *Journal of Research on Educational Effectiveness*, 6(1), 24-67.

Dong, N., Kelcey, B., Spybrook, J., & Maynard, R. A. (2017). PowerUp!-Mediator: A tool for calculating statistical power for causally-defined mediation in cluster randomized trials. (Beta Version 1.0) [Software]. Available from http://www.causalevaluation.org/

Dong, N., Kelcey, B., Spybrook, J., & Maynard, R. A. (2023). PowerUp!-Moderator-MRTs: A tool for calculating statistical power and minimum detectable effect size difference of the moderator effects in multisite randomized trials (Version 0.7) [Software].

Dweck, C. S. (2007). *Mindsets: The New Psychology of Success.* New York: Ballantine. ISBN: 978-0-345-47232-8

Eccles, J. S. (1999). The development of children ages 6 to 14. *The Future of Children*, 30-44. Freeman, J., & Bamat, D. (2022). Persistent Unequal Access to Educational Opportunity in Pennsylvania: For K-12 Students. *Research for Action*.

Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., et al. (2016). Focusing on mathematical knowledge: The impact of content-intensive teacher professional development (NCEE 2016-4010). 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/pubs/20164010/pdf/20164010.pdf.

Goldstein, H. (1987). *Multilevel models in educational and social research*. Griffin.

Goldstein, H. (2003). *Multilevel statistical models*. (3rd, ed.). Arnold.

Goodman, K.S. (1985). Chicago mastery learning reading: A program with three left feet. *Education Week*, *5(6)*, 18-20 https://www.edweek.org/teaching-learning/opinion-chicago-mastery-learning-reading-a-program-with-three-left-feet/1985/10

Horton, L. (1979). Mastery learning: Sound in theory, but... *Educational Leadership,* 37(2), 154-156.

Guo, G., & Zhao, H. (2000). Multilevel modeling for binary data. *Annual Review of Sociology*, *26*, 441–462. http://doi.org/10.1146/annurev.soc.26.1.441

Hayes, A. F. (2015). An Index and Test of Linear Moderated Mediation. *Multivariate Behavioral Research*, 50(1), 1–22.

Hedberg, E. C., & Hedges, L. V. (2014). Reference values of within-district intraclass correlations of academic achievement by district characteristics: results from a meta-analysis of district-specific values. *Evaluation review, 38*(6), 546–582. https://doi.org/10.1177/0193841X14554212

Hedges, L. V., & Hedberg, E. C. (2013). Intraclass Correlations and Covariate Outcome Correlations for Planning Two- and Three-Level Cluster-Randomized Experiments in Education. *Evaluation Review*, *37*(6), 445-489. DOI: 10.1177/0193841X14529126

Hollands, F.M., Hanisch-Cerda, B., Levin, H. M., Belfield, C.R., Menon, A., Shand, R., Pan, Y., Bakir, I., Cheng, H. & Kushner, A. (2015-22). *E$timator.* Teachers College, Columbia University. www.costtoolkit.org

Hollands, F. & Levin, H. (2017). The critical importance of costs for education decisions (REL 2017-274). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education and Evaluation and Regional Assistance, Analytic Technical Assistance and Development. Retrieved from https://ies.ed.gov/ncee/pubs/REL_2017274/pdf/REL_2017274.pdf

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Institute of Education Sciences. (2020). *Cost Analysis: A Toolkit (IES 2020-001).* U.S. Department of Education. Washington, DC: Institute of Education Sciences. Retrieved from https://ies.ed.gov/seer/pdf/IES_Cost_Analysis_Starter_Kit_V1.pdf

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of*

Applied Psychology, 69(2), 307–321.

Jung, S. (2021). Introduction to Mediation Analysis and Examples of Its Application to Real-world Data. *Journal of Preventive Medicine and Public Health*, 54(3):166-172.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. Educational Measurement: Issues and Practice, 30(4), 28–37. https://doi.org/10.1111/j.1745-3992.2011.00220.x

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.

[redacted]

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36*(2)*,* 249–277.

Leslie, S., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219), 262–265.

Levin, H., McEwan, P., Belfield, C., Bowden, B., and Stand, R. (2018). *Economic Evaluation in Education: Cost-Effectiveness and Benefit-Cost Analysis.* Thousand Oaks, CA: Sage

Publications

Levin, H. & Belfield, C. (2015). Guiding the Development and Use of Cost-Effectiveness Analysis in Education, *Journal of Research on Educational Effectiveness*, 8:3, 400-418

Liang, C. T., Rocchino, G. H., Gutekunst, M. H., Paulvin, C., Melo Li, K., & Elam-Snowden, T. (2020). Perspectives of respect, teacher–student relationships, and school climate among boys of color: A multifocus group study. *Psychology of Men & Masculinities*, *21*(3), 345.

Longford, N. T. (1993). *Random coefficient models*. Clarendon Press.

Marzano, Robert J. (2010). *Formative assessment & standards-based grading.* Bloomington, IN. Marzano Research Laboratory.

May, H., Perez-Johnson, I., Haimson, J., Sattar, S., & Gleason, P. (2009). *Using state tests in education experiments: A discussion of the issues (NCEE 2009-013)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Lawrence Erlbaum Associates.

Middleton, J. A., Wiezel, A., Riske, A., Jansen, A., & Smith, E. (2022). Relationships between perceived teacher and peer support on motivation and achievement in high school mathematics. In C. Fernández, S. Llinares, A. Gutiérrez, & N. Planas (Eds.). *Proceedings of the 45th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 195-202). https://rua.ua.es/dspace/bitstream/10045/126628/1/proceedings-pme-45-vol3-25.pdf

Morel, R. P., Coburn, C., Catterson, A. K., & Higgs, J. (2019). The multiple meanings of scale: Implications for researchers and practitioners. *Educational Researcher*, 48(6), 369-377.

https://sim.ku.edu/sites/sim/files/files/Events/2023%20SIMposium/paquinmorel%20coburn%20et%20al%202019%5B8%5D.pdf

Murphy, R., Roschelle, J., Feng, M., & Mason, C. A. (2020). Investigating efficacy, moderators and mediators for an online mathematics homework intervention. *Journal of Research on Educational Effectiveness*, *13*(2), 235-270.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards.* Washington D.C: Authors.

The Nations's Report Card (nd). *NAEP Report Card—Mathematics.* Retrieved June 23, 2024 from https://www.nationsreportcard.gov/mathematics/nation/achievement/?grade=8

Oregon Department of Education (2023). *OREGON AT-A-GLANCE DISTRICT PROFILE 2022-23 Portland SD 1J*. Retrieved June 23, 2024 from https://www.ode.state.or.us/data/ReportCard/Reports/Index

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). Harcourt Brace College Publishers.

Pennsylvania State Department of Education (2023). *2023 Pennsylvania system of school assessment technical report mathematics, English language arts, and Science.* Retrieved June 23, 2024 from https://www.education.pa.gov/K-12/Assessment%20and%20Accountability/PSSA/Pages/PSSA-Technical-Reports.aspx#

Portland Public Schools (2022). 2*022-2027 Board Goals.* Downloaded May 1, 2024 from https://www.pps.net/Page/1789

Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, *59*(1), 1–17. http://doi.org/10.2307/211248221

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data* *analysis methods*. Sage. Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA open*, *2*(4), 2332858416673968.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology*, 57(5), 749–761.

Ryan, R. M., & Deci, E. L. (2009). Promoting self-determined school engagement: Motivation, learning, and well-being. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook on motivation at school* (pp. 171-196). New York: Routledge.

Senko, C. (2016). Achievement goal theory: A story of early promises, eventual discords, and future possibilities. In K. R. Wentzel & D. B. Miele (Eds.), Handbook of motivation at school (2nd ed., pp. 75–95). Routledge. https://www.researchgate.net/profile/Corwin-Senko/publication/ 312591307

Snijders, T. A., & Bosker, R. J. (1999). *Multilevel analysis*. Sage.

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). Routledge.

Supovitz, J. A., Ebby, C. B., Remillard, J., & Nathenson, R. A. (2018). Experimental Impacts of the Ongoing Assessment Project on Teachers and Students. CPRE Research Report # RR 2018-1. In *Consortium for Policy Research in Education*. Consortium for Policy Research in Education. https://eric.ed.gov/?id=ED593465

U.S. News & World Report: More States Seek to Repeal Common Core Archived 2018-01-18 at the Wayback Machine. January 31, 2014.

Walton, G. M., & Yeager, D. S. (2020). Seed and soil: Psychological affordances in contexts help to explain where wise interventions succeed or fail. *Current Directions in Psychological*

Science, 29(3), 219–226.

Webb, N. L. (2002). Depth-of-knowledge levels for four content areas [Unpublished manuscript]. Retrieved from: http://facstaff.wcer.wisc.edu/normw/All content areas DOK levels 2032802.pdf.

Weiss, M. J., Bloom, H. S., & Brock, T. (2014). A conceptual framework for studying the sources of variation in program effects. *Journal of Policy Analysis and Management*, 33(3), 778-808.

What Works Clearinghouse (WWC). (2022). *What Works Clearinghouse Procedures and Standards Handbook, Version 5.0.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Handbooks

What Works Clearinghouse (WWC). (2023). *Study Review Protocol, Version 5.0.* Study Review Protocol Version 5.0 (March 2023) (ed.gov)

Yeager, D. (2017). Social-emotional learning programs for adolescents. The Future of Children, 27(1), 73–94. https://doi.org/10.1353/foc.2017.0004

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., ... & Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*(7774), 364-369.

Yeager, D. S., Henderson, M. D., Paunesku, D., Walton, G. M., D'Mello, S., Spitzer, B. J., & Duckworth, A. L. (2014a). Boring but important: a self-transcendent purpose for learning fosters academic self-regulation. *Journal of personality and social psychology*, *107*(4), 559.

Yeager, D. S., Lee, H. Y., & Dahl, R. E. (2017). Competence and motivation during adolescence. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), Handbook of competence and motivation: Theory and application (pp. 431–448). The Guilford Press. https://www.researchgate.net/publication/337856583

Yeager, D.S., Purdie-Vaughns, V.P., Garcia, J., Apfel, N., Brzustoski, P., Master, A., Hassert, W.T., Williams, M.E., & Cohen, G. H. (2014b) Breaking the cycle of mistrust: Wise interventions to provide critical feedback across the racial divide. *Journal of Experimental Psychology*, 143(20), 804-824.

Yoon, K. S., Duncan, T., Lee, S. W-Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007-No. 033). Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods, 12*(4), 695–719.

Zimmerman, B. J., Schunk, D. H., & DiBenedetto, M. K. (2017). The role of self-efficacy and related beliefs in self-regulation of learning and performance. *Handbook of competence and motivation: Theory and application*, *313*, 41-50.