Amplify Desmos Math

Grades K–8

RESEARCH

mCLASS Beginning-of-Year Screener

In education's rapidly evolving landscape, it's more critical than ever to ensure all students receive the support they need to succeed in mathematics. And as more and more schools rely on data to inform this support, the importance of accurate, effective screening measures has grown exponentially. Reliable screeners play a fundamental role in understanding students' capabilities and providing educators with insights they can use to support grade-level learning.

At the beginning of the school year, screeners are used to evaluate students' foundational knowledge, essential for grade-level learning. Screeners are designed to efficiently identify those who may benefit from additional instructional support (Clemens et al., 2015; Nelson et al., 2016). By proactively identifying students' strengths and areas for growth early in the year, screeners give educators data they can use to intervene and make instructional decisions that respond to all students' learning needs (Datnow & Hubbard, 2015).

But implementing efficient and diagnostically useful screeners can be challenging. Research has found that many screeners focus on number concepts (e.g., counting and arithmetic), often neglecting other mathematical constructs that predict later success (Clements & Sarama, 2008). They may also fail to align to the range of concepts addressed in the Common Core State Standards in mathematics (Brendefur et al., 2015, 2018). While screeners can identify students at risk for mathematics difficulties, they are often too narrow in scope or too time-consuming for large-scale use, particularly in the early grades (Brendefur et al., 2015). Moreover, when screeners only assess accuracy, students have limited opportunities to display the depth of their mathematical knowledge in their responses. This narrow interpretation of student thinking, coupled with the limited scope of many existing screeners, can result in data that misrepresents students' needs and abilities, leading to interventions that do not fully address their learning needs.

Unlike traditional screeners that narrowly focus on number concepts or require extensive administration time, the Amplify Desmos Math Beginning-of-Year Screener (ADM BOY Screener) provides an evaluation of essential mathematics concepts and skills in each mathematical domain while being efficient enough for large-scale use. Educators can obtain information about their students' thinking without sacrificing valuable instructional time as part of their core curriculum. This paper begins with a description of the 2023–24 Item Calibration Study, which was conducted to evaluate the quality of the individual items in the screener and how they functioned together before being incorporated into the ADM BOY Screener. It then provides a detailed explanation of the purposes and design of the ADM BOY Screener. Finally, it discusses the principles Amplify applied to define and measure what constitutes an effective screener. Altogether, the ADM BOY Screener was developed with a robust underlying framework that aims to address the challenges of early identification and intervention while paving the way for more personalized instruction and student success.

Item Calibration Study

The Item Calibration (IC) Study was an important first step in developing the ADM BOY Screener. During the 2023–24 school year, this study examined the quality of all K–3 items at beginning of year (BOY) and K–8 items at end of year (EOY). The sample included 2,459 students in grades K–5 and 551 students in grades 6–8 at EOY. The EOY results were used to develop alternate, equivalent benchmark forms for grades K–8 for use at beginning, middle, and end of year. This helped ensure the items on each benchmark form were developmentally appropriate and aligned with the expected progression of students' mathematical knowledge and skills at the respective point in the school year.

Classical test theory (CTT), item response theory (IRT), and differential item functioning (DIF) were employed to provide information on how each individual item functioned, the level of difficulty, and whether or not the items discriminated between students of varying performance levels. The following research questions were examined:

1. To what extent do items exhibit appropriate difficulty levels for their respective grade? What is the overall distribution of item difficulty within each grade?
2. How effective is each item in distinguishing between high-performing and low-performing students, as evidenced by their overall score on the assessment? How does item performance vary between students with different levels of achievement?
3. Does each item show differential item functioning (DIF)? How does item difficulty vary among specified groups of students (e.g., ethnicity and gender) with equivalent performance levels?

The first research question examined the extent to which items exhibited appropriate difficulty levels for their respective grade and the overall distribution of item difficulty within each grade as measured by the p-value. The results show the majority of K–5 items exhibited appropriate difficulty levels ($0.2 < p\text{-value} < 0.9$) at BOY and EOY (63%–80.3% at BOY and 80%–98.2% at EOY). The higher grade levels (grades 2 and 3 at BOY and grades 4 and 5 at EOY) had a higher proportion of items flagged as difficult. Between BOY and EOY, the proportion of items flagged as difficult across grades K–3 decreased substantially following revisions by internal and external content experts. The proportion of items flagged as easy increased from BOY to EOY (1.5% at BOY, 3.2%–8.3% at EOY); however, the proportion of easy items at EOY remained small. Results revealed that there were no items flagged as easy in grades 6–8. Similarly, very few items were flagged as difficult, with 0.0% flagged in grade 6, 9.9% in grade 7, and 6.3% in grade 8.

Table of contents

Item response theory (IRT) was also used to calibrate the items. This allowed difficulty estimates of the items to be placed on a common scale. These results aligned with the p-value statistics, indicating that the majority of items were neither too easy nor too difficult.

The second research question examined how effective each item was in distinguishing between high-performing and low-performing students and the variation in item performance between students with different levels of achievement. The ability of an item to discriminate between high- and low-performing students was evaluated with the point-biserial correlation between the item score and the sum of the remaining item scores. Higher point-biserial correlation values indicate the item does a better job of discriminating between high- and low-performing students. At BOY and EOY, the findings demonstrate the majority of items distinguished between high-performing and low-performing K–5 students well (75%–90.9% at BOY and 90.3%–97.9% at EOY). There was a slightly higher proportion of items that received biserial flags in grades 2 and 3 at BOY, but not at EOY. Again, the proportion of items flagged as discriminating poorly between students with different performance levels decreased substantially after the items underwent revisions between BOY and EOY. In grades 6, 7, and 8, 15.4%, 9.1%, and 12.5% of items received biserial flags, respectively. Any items that received difficulty or biserial flags were revised and re-reviewed by internal and external content experts.

The third research question identified whether each item displayed differential item functioning (DIF) and how the item difficulty varied among specified subgroups of students with equivalent performance levels. Overall, most K–2 items did not indicate DIF at BOY or EOY. The proportion of items with little to no indication of DIF was similar in regards to Gender (Male/Female) and Race (White/Other). In BOY, 1.7%–6.5% of items exhibited Category C DIF. For the purposes of this paper, items classified as Category C displayed significant DIF and were rejected or flagged for review. This range remained similar at EOY, in which 2.2%–4.2% of items exhibited Category C DIF. Items in grades 3–8 were not evaluated for DIF because the sample sizes were too small, potentially providing insufficient information to reliably identify items with differing functions. Additional data collection is planned for the fall of 2025 to further analyze items for DIF.

Background on Amplify Desmos Math Beginning-of-Year Screener

Unlike traditional screeners that focus too narrowly on number concepts or require extensive administration time, the ADM BOY Screener provides an evaluation of essential mathematics concepts and skills in each mathematical domain, while time-being efficient enough for large-scale use. This allows educators to obtain information about their students' thinking without sacrificing valuable instructional time.

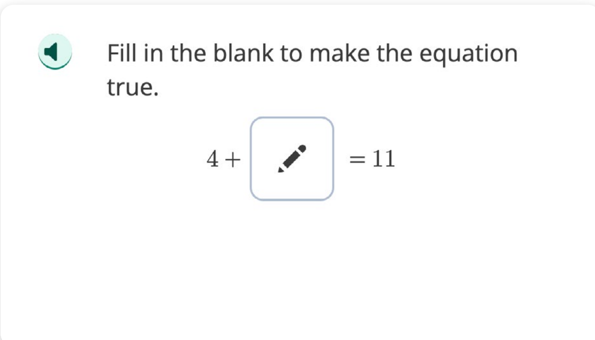
The ADM BOY Screener is a research-based, standardized assessment designed to evaluate content that is predictive of students' success in mathematics. To begin the development process, the research team, along with mathematics content experts, extensively reviewed a variety of resources to identify the mathematics concepts and skills to include in the screener. These resources included:

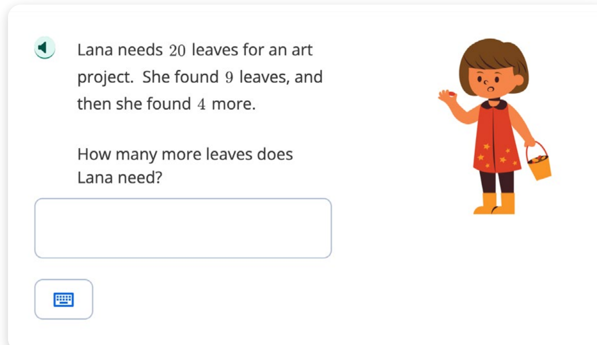
- Achieve the Core: Focus by Grade Level.
- Common Core State Standards for Mathematics.
- National Council of Teachers of Mathematics (NCTM).
- National Assessment of Education Progress.
- Trends in International Mathematics and Science Study.
- High-impact academic journals, published books, and research from prolific researchers across cognitive science and education.

The ADM BOY Screener is aligned to rigorous state standards across the Common Core State Standards for Mathematics domains in K–8. As part of the Amplify Desmos Math curriculum, data gathered from this assessment supports differentiation within core instruction, helping educators identify key instructional targets and resources to support students' diverse learning needs.

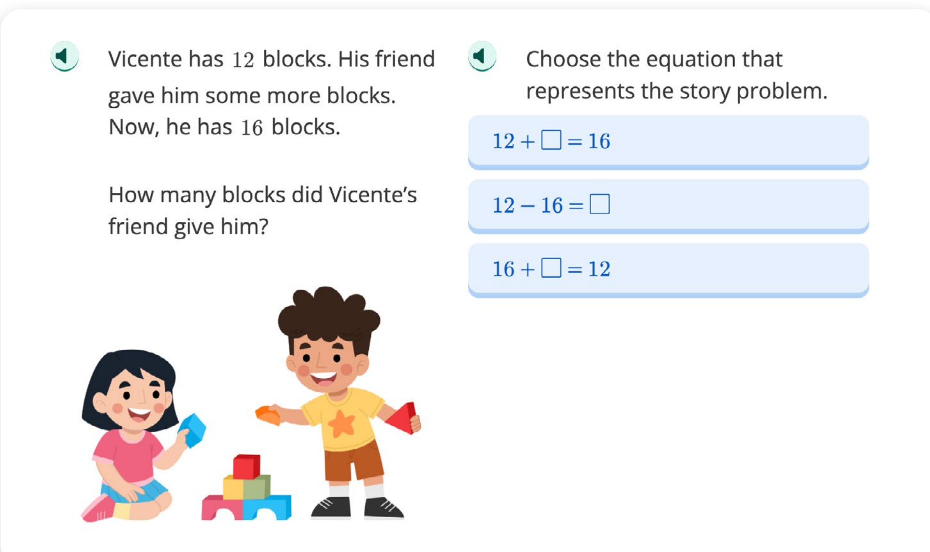
The ADM BOY Screener is designed to not only assess a range of essential concepts and skills, but also to inform instruction at a granular level. It consists of 12–15 items, which vary by grade level. This is a group-administered assessment and takes students between 10–15 minutes to complete. Each grade-level assessment contains items that determine students' readiness to learn content that will be taught in the upcoming school year. These items allow students to draw on prior knowledge and previously learned strategies when problem solving, making the ADM BOY Screener accessible for all students. The format of K–2 items includes math input, multiple choice, and constructed response (e.g., drag-and-drop and click-to-place) items.

Item Formats: Grades K–2

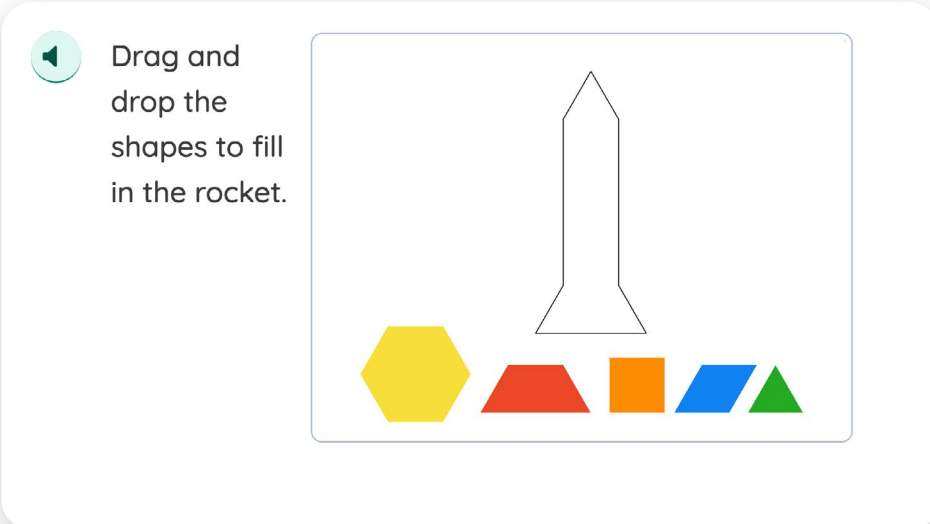




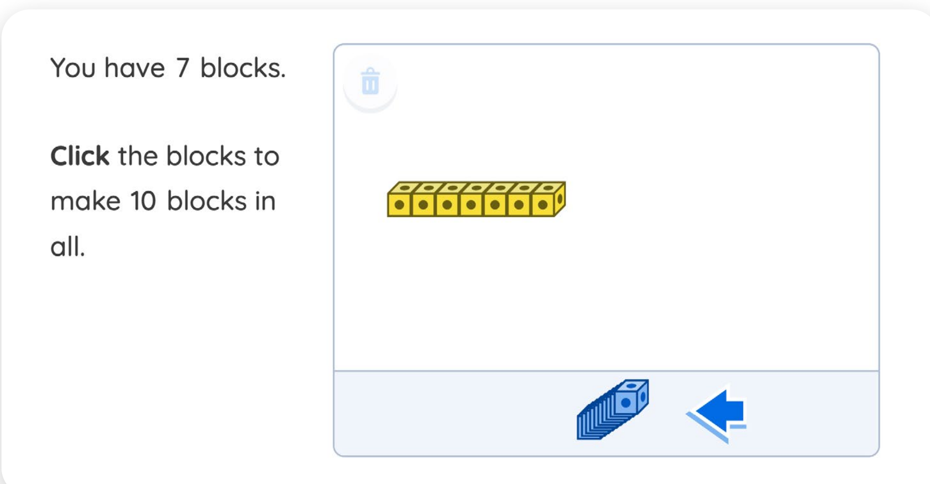
Math input items in the Grade 2 BOY Screener: The left item is inline math input; students determine the unknown addend that makes an equation true. The right item is math input; students solve a two-step word problem involving addition and subtraction within 100.



Multiple choice item in the Grade 2 BOY Screener: Students identify the equation that represents a change unknown word problem.



Constructed response: Drag-and-drop item in the Grade 1 BOY Screener. Students solve a shape puzzle by filling in the outline of a rocket.

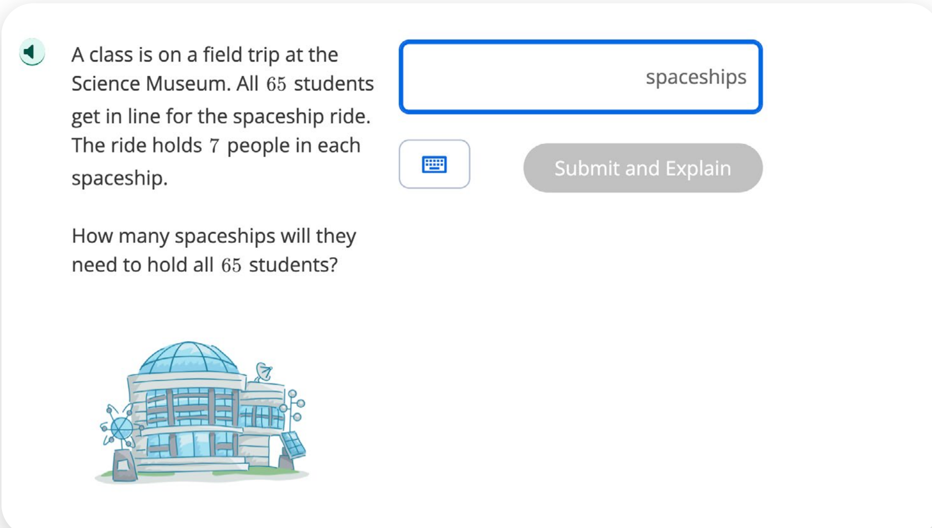


Constructed response: Click to place item in the Grade K BOY Screener. Students make 10 by clicking the blocks on the bottom to add them to the block tower.

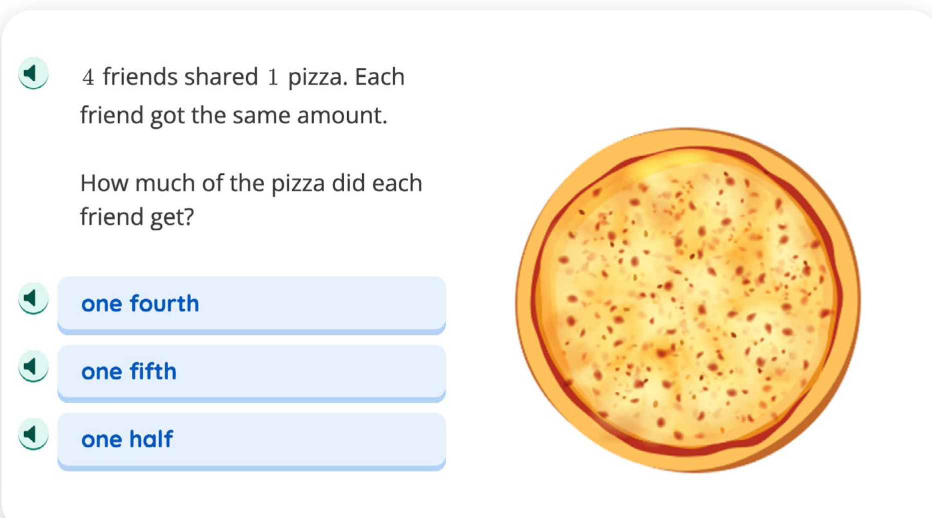
Table of contents

Item formats: Grades 3–5

The format of items in grades 3–8 includes math input, written explanations (text input), multiple choice, table input, and constructed response (drag-and-drop) items.



Math input/text input item in the Grade 5 BOY Screener: Students first solve a division with a remainder word problem and enter the answer in the math input box, then explain their thinking in the text input box that appears below.



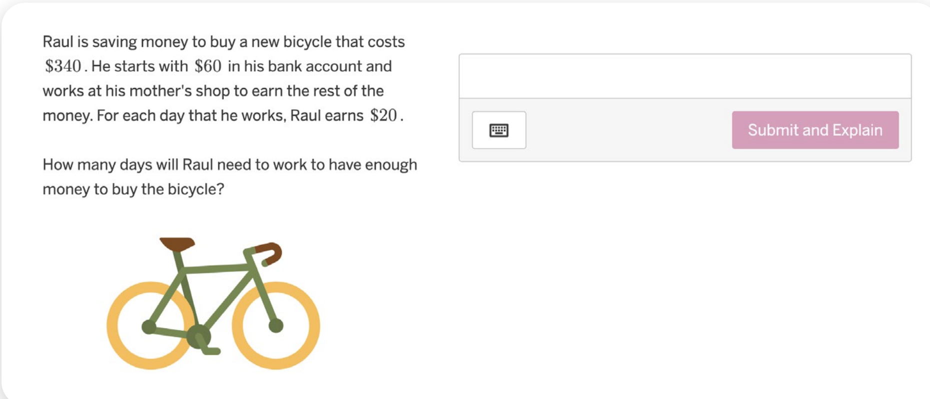
Multiple choice item in the Grade 3 BOY Screener: Students select an option to describe the size of one share in an equally partitioned shape.

Determine which of the following descriptions apply to each shape by clicking the boxes to mark them.

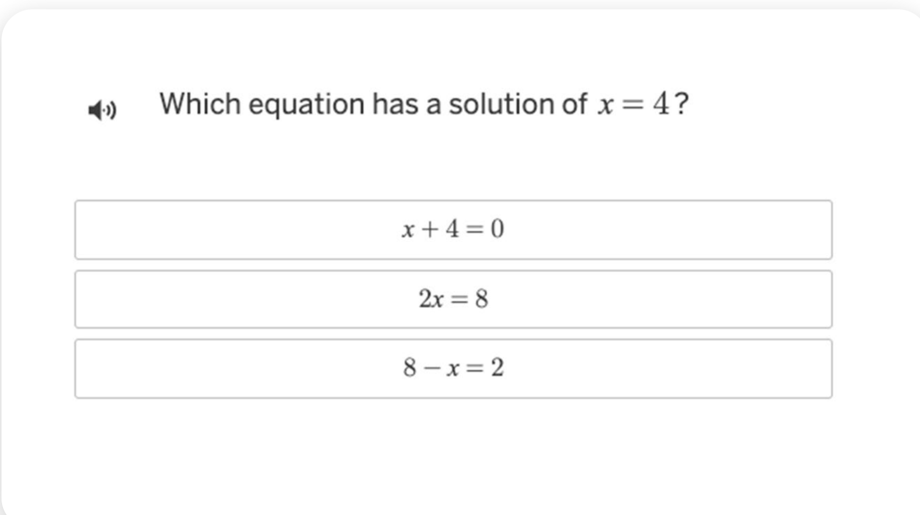
	Hexagon	Rectangle
The shape has only four sides.		
All of its sides have equal lengths.		
The shape is two-dimensional.		
The shape is a pentagon.		
The shape is a quadrilateral.		

Table input items involve click-to-place interactions in grades 3–5. In the Grade 4 BOY Screener, students identify attributes of two shapes by clicking cells within the table to mark them.

Item formats: Grades 6–8



Math input/text input item in the Grade 6 BOY Screener: Students first solve a multi-step word problem and enter the answer in the math input box, then explain their thinking in the text input box that appears below.



Multiple choice item in the Grade 8 BOY Screener: Students select an option to identify the shape that results from the rotation of another shape.

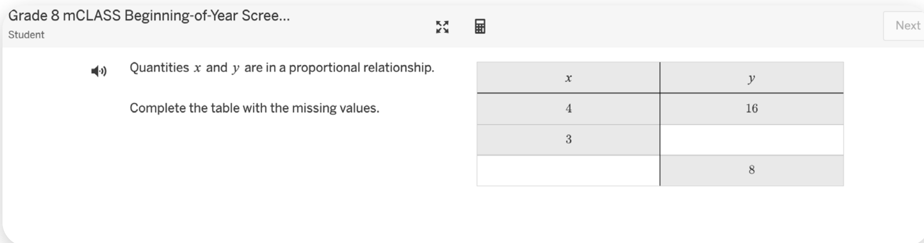
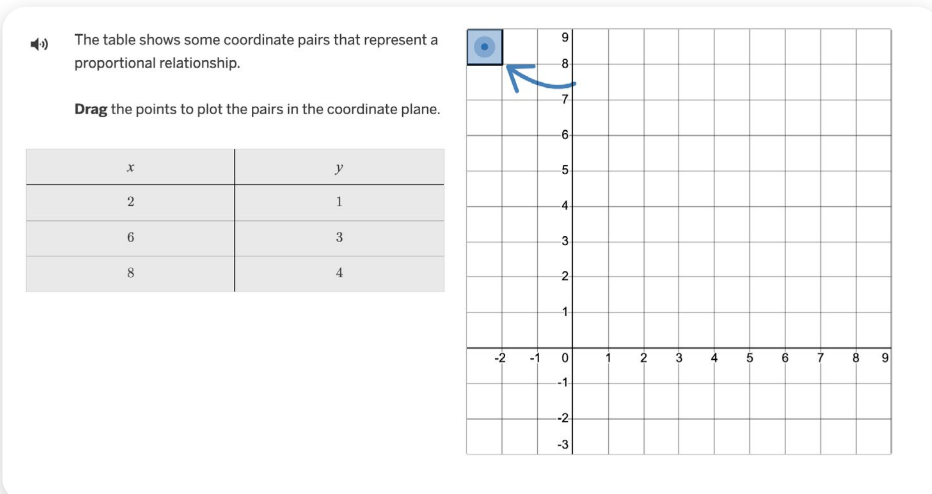


Table input items involve math input interactions in grades 6–8. In the Grade 8 BOY Screener, students fill in cells within the table to represent a proportional relationship between two quantities.



Constructed response: Drag-and-drop item in the Grade 7 BOY Screener. Students drag and drop to plot ordered pairs representing a proportional relationship in the coordinate plane.

A student's overall score is used to determine their level of performance. Following the completion of the 2024–25 field study, three performance levels (Above Benchmark, Benchmark, Below Benchmark) determined by two cut scores will be established. These scores will indicate risk and on-grade-level proficiency for students. Each level provides an overview of students' mathematical knowledge and skills, allowing educators to quickly identify where each student falls in terms of readiness for grade-level content. This tiered approach helps teachers target instruction more effectively, addressing the specific needs of students at different performance levels. In addition, teachers gain detailed insight into students' diverse ways of thinking, which can directly inform instructional next steps.

Applied principles

Three principles were used to inform the design of the ADM BOY Screener. Principle 1 emphasizes the need for screeners to be reliable, valid, and research based. Principle 2 describes the ADM BOY Screener's asset-based approach to assessment. Principle 3 highlights the importance of balancing efficiency with accuracy to maximize the impact of data-driven instruction.

Principle 1: Screeners should be grounded in research, ensuring reliability, validity, and technical soundness.

The American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education highlight the importance of rigorous research and testing to validate the effectiveness and accuracy of assessments (2014). The items within the ADM BOY Screener are under evaluation in a large-scale field study, conducted across four regions of the United States with thousands of students. This field study will ensure that the screener is reliable and valid, reinforcing its potential as a robust screening measure that aligns to these rigorous standards.

Researchers assert that "screening tools must be technically sound" (VanDerHeyden et al., 2017, pp. 65–66). This means that the screener provides strong validity on important outcome measures, such as high-stakes assessments (Salvia et al., 2007); demonstrates that core instruction is effective; and identifies students who may be at risk for difficulty with mathematics to enable targeted instruction (Johnson et al., 2010). With reliable and valid data, schools and districts can make confident decisions when identifying students in need of additional support, ensuring these decisions are both timely and appropriate. Moreover, adhering to these standards of technical adequacy helps ensure that assessments do more than just measure performance—they provide actionable insights that drive meaningful change in the classroom.

A well-designed screener must consistently produce stable results so educators can trust the assessment is consistently measuring essential concepts and skills. This reliability minimizes the risk of misclassification of students and supports accurate, data-driven instructional decisions. A successful screener must also be valid, referring to how accurately an assessment measures what it claims to measure. Two types of validity are

Table of contents

important when considering this assessment: concurrent and predictive validity. Concurrent validity is tested by administering the assessment alongside an external criterion measure around the same time. This helps educators understand how accurately the screener reflects students' current abilities in relation to other established measures. Predictive validity is assessed by comparing assessment scores to external assessment performance at later points, ensuring that the measure can predict future performance. For instance, a well-designed kindergarten screening measure should predict potential difficulties by the end of first grade or later, allowing educators to allocate resources early for intervention in regular classroom settings (Gersten et al., 2012). This helps guide educators toward implementing interventions that can ameliorate long-term difficulties.

Principle 2: Screeners should reveal what students know and can do, providing more accurate and actionable data than scoring of correct or incorrect can provide by itself.

An asset-based approach to mathematics instruction focuses on students' thinking, allowing teachers to build on their students' strengths and diverse ways of problem solving (Celedón-Pattichis et al., 2018). An asset-based approach highlights students' existing knowledge to point the way toward deeper learning. It also supports teachers in implementing instruction that draws on students' funds of knowledge (Gonzalez et al., 2001), positions them as capable, and promotes academic success (Bartell et al., 2017). Indeed, recognizing that assessment is a central component of teaching and learning, NCTM (2020) asserts that assessments should gather evidence of students' mathematical thinking, understanding, and reasoning in their responses, even if they are incorrect on an item.

The ADM BOY Screener uncovers how students think mathematically and gives a meaningful picture of their readiness for grade-level mathematics instruction. Each item is thoughtfully designed to elicit detailed information about students' understanding of concepts and skills in foundational mathematics, including their use of strategies and common response patterns. This data is synthesized into Student Thinking Reports that empower teachers to respond effectively, intervene as necessary, and tailor instruction to meet students' learning needs. In doing so, the Student Thinking Reports teachers leverage students' strengths to drive learning forward.

Table of contents

Principle 3: Screeners must be efficient and inform instructional decisions within a core curriculum.

It's important for screening tools to balance accuracy and sensitivity with efficiency so that they are able to provide valuable insights to educators while minimizing the time required for administration (VanDerHeyden et al., 2017). The ADM BOY Screener has been designed to focus on the critical areas that yield valid and reliable information about students' progress in grade-level learning. With the use of technology, assessments can be administered to large groups of students and scoring can be done almost automatically (Gersten et al., 2012). This reduces the burden of manual testing while making recommendations that are tied to instructional targets readily available to improve teachers' instructional decision-making process (Ketterlin-Geller et al., 2018)

The ADM BOY Screener is designed to streamline the assessment process and reduce the overall amount of time spent on assessments. The ADM BOY Screener is recommended to replace the Pre-Unit Check in the first and second unit of ADM; it covers a wider range of content in addition to the same concepts and skills. This provides educators a more comprehensive understanding of students' mathematical knowledge at the start of the school year. While the Pre-Unit Check typically includes 2–5 items that assess the skills relevant to the upcoming unit, the ADM BOY Screener goes beyond the targeted skills of the Pre-Unit Check by including a broader range of concepts both within a domain and across domains. This helps illustrate the connectivity of mathematics thus giving further insight into student thinking.

The screener also allows for both individual item analysis and domain-level analysis of student responses, helping teachers identify patterns in students' thinking. The different levels of data provide a comprehensive picture of a student's knowledge, from a single item level to an overall conceptual understanding of concepts and skills that are foundational to grade-level learning. A detailed analysis of student responses offers greater opportunities for asset-based recommendations grounded in student thinking, supporting continued progress in grade-level standards and domains. For instance, when looking at students' understanding of a given essential math concept or skill, the Student Thinking Report analyzes and synthesizes students' ways of thinking, then provides tailored recommendations for small-group mini-lessons and individualized practice lessons. These recommendations provide a road map into core curriculum planning, linking to specific moments in the curriculum and resources for small-group instruction and practice. Emphasizing asset-based and personalized instruction, the ADM BOY Screener supports all students in achieving grade-level success in mathematics.

Conclusion

As a critical component of Amplify Desmos Math's core curriculum, the ADM BOY Screener is a powerful tool for understanding students' mathematical thinking. Guided by three principles—technical soundness, an asset-based approach to assessment, and efficiency—the ADM BOY Screener is able to supplant the additional ADM Pre-Unit Checks for Units 1 and 2 of each grade level, streamlining assessment moments and providing timely and actionable insights for teachers. Rooted in rigorous research on foundational mathematics concepts and skills, this screener is a valid and reliable resource that aligns with state standards and addresses the need for efficient, data-driven instructional decisions. By leveraging an asset-based approach and attending to students' capabilities, the ADM BOY Screener extends beyond measuring accuracy, empowering educators to understand students' unique and valid mathematical thinking and uncovering ways to build on their knowledge to improve learning outcomes. This assessment provides an opportunity to gather meaningful data at the start of the school year to inform differentiation and support all students' success in mathematics. In this way, the ADM BOY Screener transforms assessment into an opportunity for empowerment, using data not just for identification, but for growth among all students.

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational & psychological testing*. AERA.

Bartell, T., Wager, A., Edwards, A., Battey, D., Foote, M., & Spencer, J. (2017). Toward a framework for research linking equitable teaching with the standards for mathematical practice. *Journal for Research in Mathematics Education, 48*(1), 7–21.

Brendefur, J., Strother, S., Thiede, K., & Appleton, S. (2015). Developing multiplication fact fluency. *Advances in Social Sciences Research Journal*, *2*(8), 142–154. https://doi.org/10.14738/assrj.28.1396

Brendefur, J. L., Johnson, E. S., Thiede, K. W., Strother, S., & Severson, H. H. (2018). Developing a multi-dimensional early elementary mathematics screener and diagnostic tool: The primary mathematics assessment. *Early Childhood Education Journal, 46*(2), 153–157. 10.1007/s10643-017-0854-x

Celedón-Pattichis, S., Borden, L., Pape, S., Clements, D., Peters, S. A., Males, J. R., Chapman, O., & Leonard, J. (2018). Asset-based approaches to equitable mathematics education research and practice. *Journal for Research in Mathematics Education*, *49*(4), 373–389. 10.5951/jresematheduc.49.4.0373

Clemens, N. H., Keller-Marguilis, M. A., Scholten, T., & Yoon, M. (2015). Screening assessment within a multi-tiered system of support: Current practices, advances, and next steps. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (pp. 187–213). Springer. https://doi.org/10.1007/978-1-4899-7568-3_12

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, *45*(2), 443–494. https://doi.org/10.3102/0002831207312908

Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record*, *117*, 040302. www.tcrecord.org/?ContentId=17848

Gersten, R. M., Clarke, B., Jordan, N., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for the primary grades: Beginnings of a research base. *Exceptional Children*, *78*, 423–445. **doi.org/10.1177/00144029120780**

Gonzalez, N., Andrade, R., Civil, M., & Moll, L. (2001). Bridging funds of distributed knowledge: Creating zones of practices in mathematics. *Journal of Education for Students Placed at Risk (Jespar), 6*(1–2), 115–132. **doi.org/10.1207/S15327671ESPR0601-2_7**

Johnson, E. S., Jenkins, J. R., & Petscher, Y. (2010). Improving the accuracy of a direct route screening process. *Assessment for Effective Intervention*, *35*, 131–140. doi:10.1177/1534508409348375

Ketterlin-Geller, L., Shivraj, P., Basaraba, D., & Schielack, J. (2018). Universal screening for algebra readiness in middle school: Why, what, and does it work? *Investigations in Mathematics Learning, 11*(2), 120–133. **10.1080/19477503.2017.1401033**

National Council of Teachers of Mathematics (NCTM). (2020). *Catalyzing change in early childhood and elementary mathematics: Initiating critical conversations*. NCTM.

Nelson, P., Van Norman, E., & Lackner, S. (2016). Research into practice: A comparison of methods to screen middle school students for reading and math difficulties. *School Psychology Review*, *45*(3), 327–342. **10.17105/SPR45-3.327-342**

Salvia, J., Ysseldyke, J. E., & Bolt, S. (2007). *Assessment in Special and inclusive education* (10th ed.). Houghton Mifflin.

VanDerHeyden, A., Codding, R., & Martin, R. (2017). Relative value of common screening measures in mathematics. *School Psychology Review*, *46*(1), 65–87. **doi.org/10.17105/SPR46-1.65-87**

Visit us online for more information on Amplify's products and services.

amplify.com

Amplify.

© 2025 Amplify Education, Inc.
All trademarks and copyrights are the property of Amplify or its licensors.