

The Effective and Appropriate Use of Large-Scale Assessments in Mathematics Education to Guide Systemic Improvement and Equitable Student Learning

A Position of the National Council of Teachers of Mathematics

Large-scale assessments are assessments given to a population of students to measure student achievement for the purpose of educational accountability, including identifying gaps in curriculum, recognizing groups of students who are being underserved, and setting and achieving system-wide goals. Data from large-scale assessments must be used with caution when making judgments about individual student performance because of the narrow constraints of the types of knowledge and skills that can be assessed through such a format.

Large-Scale Assessments Can Support Systemic Improvement

A vital role for large-scale mathematics assessments in schools and districts is to identify systemic deficiencies and successes in addressing the needs of students. In the last paragraph of a joint press release, the Leadership Conference on Civil and Human Rights (2015) stated, “We cannot fix what we cannot measure. . . . Abolishing the tests or sabotaging the validity of their results only makes it harder to identify and fix the deep-seated problems in our schools.” Although we must be critical of their misuse and mindful of their limitations, large-scale mathematics assessments serve an important role in illuminating inequities and signaling success in addressing these. Two questions to keep in mind when interpreting group data from large-scale assessments are the following:

- How are the data from large-scale assessments being used to identify mathematics education deficiencies and highlight successes in your organization?
- How can the data from large-scale assessments be used to inform plans for systemic improvements?

Inferences From Large-Scale Assessment Data Must Be Used With Caution

Results from large-scale assessments can be overanalyzed and misinterpreted (National Council of Teachers of Mathematics [NCTM], 2014) and can lead to invalid inferences. Each assessment makes certain assumptions about which tasks, skills, or important information accurately captures and demonstrates knowledge of the material (Pellegrino et al., 2001). Often large-scale assessment results are used for purposes for which they were not intended and for which there can be negative consequences for students and teachers. Thus, the social consequences should be considered when examining how test results are used (Messick, 1989; Shepard, 1993). These consequences might be positive, such as increased attention to areas of need, or negative, such as inequitable tracking

of students or unsubstantiated teacher evaluation. To guard against making invalid inferences from assessment data, two questions to keep in mind are these:

- What are valid inferences and interpretations that can be made from the data?
- What are the limitations of the data with respect to how they can validly be used?

Large-Scale Assessment Must Always Be Viewed Through an Equity Lens

Large-scale assessment should not contribute to labeling and sorting students but must be designed and used to reduce inequities in educational systems. Ideally, assessments should be bias-free and have relevance to a wide range of students so they provide opportunities for all students to show what they know and can do; this is often not the case with large-scale assessments (Berry et al., 2014; Randall et al., 2022). In reality, “discrepancies in scores on standardized achievement tests mirror discrepancies in opportunities and life chances that students from different backgrounds experience in their everyday lives” (Gutiérrez, 2008). When selecting, designing, administering, or using results from large-scale assessments with attention to equity, consider two questions:

- How does your system acknowledge and work to lessen the impact of biases that might be inherent within the large-scale assessments that students are required to take?
- What protocols and processes are in place to ensure that results of large-scale assessments are used productively and not punitively to inform efforts to strengthen teaching, curriculum, and support for all students?

Large-Scale Assessments Must Align with Curriculum and Goals

It is important that the assessments used are aligned with the mathematics curriculum being taught and are coherent with the goals of sound mathematics teaching and learning within the system. Students must be assessed on the mathematics that is most important to learn rather than merely the mathematics that is easiest to assess. Assessment outcomes may illuminate areas of curriculum that need greater attention. Therefore, large-scale assessment results should be used to inform educators not only about the progress students have made but also about adjustments to curriculum and instruction that are needed (Jimenez & Modaffari, 2021).

To ensure large-scale assessments are better aligned with curriculum and pedagogical goals, here are two questions to consider:

- What safeguards are in place to ensure large-scale assessments not only focus on mathematics content knowledge but also reinforce the importance of mathematical processes and practices that engage students in learning (NCTM 1995, 2014; Suurtamm et al., 2016)?
- How are assessments used to encourage and inform pedagogical practices that support strong mathematics teaching and learning rather than promoting a focus on merely procedural learning and rote instruction?

Major Decisions About Student Placement, Promotion, or Graduation Must Be Based on Valid Inferences From Multiple Data Sources

Decisions about students must be informed by and based on evidence from multiple data sources and

not only one large-scale assessment. Results from large-scale mathematics assessments are one source of information and provide a snapshot of student mathematical knowledge on a particular set of problems on a particular day. By contrast, ongoing formative and summative assessments provide a moving picture of student understanding and offer evidence of students' progress toward established learning goals. Formative assessments, in particular, help to show the progress of learning or stumbling blocks in learning and offer feedback to students and teachers about areas for improvement.

In making inferences about individual student learning, the following are two questions to consider:

- How do large-scale assessment outcomes interact with information derived from ongoing classroom-based sources to offer guidance about student progress toward learning goals?
- How are high-stakes decisions about students informed by multiple data sources and transparent and accessible to caregivers and students?

References

- Berry, R. Q., III, Ellis, M., & Hughes, S. (2014). Examining a history of failed reforms and recent stories of success: Mathematics education and Black learners of mathematics in the United States. *Race Ethnicity and Education, 17*(4), 540–568.
- Gutiérrez, R. (2008). Research commentary: A gap-gazing fetish in mathematics education? Problematizing research on the Achievement Gap. *Journal for Research in Mathematics Education, 39*(4), 357–364. <https://doi.org/10.5951/jresmetheduc.39.4.0357>
- Jimenez, L., & Modaffari, J. (2021). *Future of testing in education: Effective and equitable assessment systems*. Center for American Progress. <https://www.americanprogress.org/wp-content/uploads/sites/2/2021/09/EquitableAssesmentSystems1.pdf>
- Leadership Conference on Civil and Human Rights. (2015, May 5). Civil rights groups: “We oppose anti-testing efforts” [Press release]. <https://civilrights.org/2015/05/05/civil-rights-groups-we-oppose-anti-testing-efforts/>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). Macmillan.
- National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*.
- National Council of Teachers of Mathematics, (2014). *Principles to actions: Ensuring mathematical success for all*.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. National Academies Press.
- Randall, J., Slomp, D., Poe, M., & Oliveri, M. E. (2022). Disrupting White supremacy in assessment: toward a justice-oriented, antiracist validity framework. *Educational Assessment, 1*–9.
- Shepard, L. A. (1993). Evaluating test validity. *Review of Research in Education, 19*, 405–450.
- Suurtamm, C., Thompson, D. R., Young Kim, R., Diaz Moreno, L., Sayac, N., Schukajlow, S., Silver, E., Ufer, S., & Vos, P. (2016). *Assessment in mathematics education: Large-scale assessment and classroom assessment* (1st ed.). Springer Nature.