



---

## FOUR BIG TAKEAWAYS ABOUT TESTING



Despite its pervasiveness in educational settings, testing is a topic that is often misunderstood by the students, teachers, policymakers and members of the public who consume or produce the results. A recent edition of the *ANNALS of the American Academy of Political and Social Science* aims to explore and maybe even reduce some of this confusion, by taking a step back and asking psychometricians and other testing experts to ponder a core question: “What use is educational assessment?” Specifically, the edition delves into uses of assessment in real-world settings such as classrooms, legislatures, and admissions committees. Contributors included NEPC Fellows [William Penuel](#), [Benjamin Shear](#), and [Lorrie Shepard](#), all professors at the University of Colorado Boulder. Here are four broad takeaways from their articles.

- 1. Grading is a problem for learning.** Grades play an important role in schooling, as the most common form of assessment. As [Shepard writes](#), a downside of grading is that it elicits comparisons with others that dampen student motivation. For some students, grading can send a message that a temporary failure to reach a goal is a permanent result of insufficient levels of ability. Summarizing the literature in this area, Shepard notes that students learn more from written comments alone than from written comments plus grades. Yet educators are typically required to assign grades. With this in mind, Shepard offers several research-based suggestions for ameliorating grades’ detrimental impact on student learning, including (a) allowing students the opportunity to learn from feedback and improve their work prior to receiving a grade and (b) permitting students to replace early efforts that fall short with subsequent evidence of mastery. While colleges may require high school grades for admissions purposes, Shepard suggests replacing younger students’ grades with evidence about

developmental milestones and advice on what parents can do to support their children's growth.

- 2. We should stop assuming that the same state assessment can and should be used both to hold school systems accountable and to formatively shape classroom instruction.** Some critics of state testing programs have suggested tweaks to help make accountability tests more relevant to classroom instruction (e.g., by getting teachers the students' test scores before the school year ends, or by adding more questions so that individual students' scores are more reliable). This perspective erroneously assumes that it is both possible and desirable to have a complex assessment system in which a standard set of formative assessments is aligned with interim and accountability-based exams, [Shepard explains](#). The reality is that state accountability tests and formative classroom tasks are just too different. They are designed differently because they have different objectives. Accountability tests sample a wide variety of curricula because they aim to monitor school systems for policy and programmatic purposes. The purpose of formative assessment, by contrast, is to help individual teachers adapt instruction for the individual students in their classrooms. As such, formative assessment—which may take the form of a conversation rather than a written test—should be aligned to ambitious teaching practices that support deeper learning. Moreover, state tests must be “curriculum neutral,” meaning they should not favor one curriculum over another. Curriculum varies substantially between districts, schools and even classrooms. And, while state exams aim to align to standards, standards are by definition much more general than curricula.
- 3. The National Assessment of Educational Progress (NAEP) might not be perfect, but it's pretty much the best we have.** The NAEP testing program has been administered to representative student samples since 1969. Its “quality, longevity, stability, and comparability” stand out as superior to any alternative for “monitoring how student achievement has changed over time and varies among states or subgroups,” [Shear writes, along with his coauthors Erin Fahle and Kenneth Shores](#). However, they also echo Shepard's point that large-scale assessments are useful for some purposes (e.g., monitoring trends) but not necessarily for others (e.g., formative assessment). They warn that “many valued educational outcomes cannot be measured by standardized tests.” Along the same lines, if large-scale assessments are to be used for monitoring (which we see as distinct from high-stakes accountability), they ought to be designed and evaluated specifically for such purposes. The NAEP is a good example of an assessment system that is intentionally designed for large-scale, low-stakes monitoring purposes. They suggest that future approaches to monitoring might combine multiple data sources.
- 4. Assessments can promote equity, rather than contribute to inequity.** [Penuel and his co-author, Douglas Watkins](#), write about their work with Denver Public Schools on classroom assessments embedded in curricula designed to promote both equity and justice. The approach privileges the student point of view on students' experience of class. For instance, at the beginning of one unit, students watched a video of a girl who'd been hospitalized with antibiotic-resistant infections. They then worked alone and in groups to generate and prioritize a set of questions they would

need to ask in order to determine what's going on with the sick girl. This gives students agency because they are coming up with their own guiding questions rather than responding to questions written by testing companies or even by their teachers. At the end of each lesson, students also respond to brief surveys about their experiences in the classroom. Survey results are then disaggregated by categories such as student ethnicity and race in order to determine whether certain subgroups are having different experiences than others. For instance, are females or African Americans less likely to perceive that their perspectives are important during whole-class discussions? Researchers and administrators review the results, using them as the basis to design professional development opportunities for teachers (e.g., workshops on creating inclusive classroom cultures). "The classroom is a community where students pursue questions together and evolve standards for what counts as a question worth pursuing as well as for how answers to questions are to be warranted."

### NEPC Resources on Assessment

This newsletter is made possible in part by support provided by the Great Lakes Center for Education Research and Practice: <http://www.greatlakescenter.org>

The National Education Policy Center (NEPC), a university research center housed at the University of Colorado Boulder School of Education, produces and disseminates high-quality, peer-reviewed research to inform education policy discussions. Visit us at: <http://nepc.edu>