We Should Listen to the Canaries

How alternate assessments for students with disabilities and English learners can point us toward better systems for all.

Ellen Forte

Students with disabilities and English learners rank high among those facing the greatest academic challenges in public education. Like canaries in the coal mine, these students are often the first to reveal how conditions in schools are not adequately supporting students' academic growth.[1] Their mandated participation in alternate statewide assessments offers insights into why that might be.

First, the testing requirement for these groups incentivizes state and local educators to teach these students appropriate academic content and keep them in classrooms with their age-appropriate peers. Keeping students with disabilities in these classrooms, rather than segregating them in isolated classrooms, helps them and their nondisabled peers in a variety of ways, including improved academic achievement and self-regulation.[2] Similarly, English learners and their English-proficient peers both benefit from academic and social interactions in ways that enhance their achievement and linguistic skills.[3]

Second, high-quality alternate academic assessment and English language proficiency assessments involve some degree of individualized administration: Students taking the test interact directly with their teachers as it is administered. There is no secret about what the tests are asking students to do, and teachers see and evaluate students' performance in the moment. As a result, teachers gain insights into what their students know and can do directly from the process of testing rather than having to wait until scores become available.[4] They can use those insights in their classrooms that very day, and students can reap immediate benefits from their assessment-drive interactions with their teacher.

A Plethora of Tests

The break from testing that the pandemic instigated in spring 2020 only renewed old questions about the purpose of the entire assessment system and qualms about the number of summative state tests. While the full repertoire varies to some degree across states, some components are common to all. All states administer assessments in English language arts and mathematics in grades 3 through 8 and at least once in high school as well as science assessments in at least one grade in each of the 3 to 5, 6 to 8, and high school grade ranges. All states also administer English language proficiency assessments to K-12 English learners each year.

Every one of these general assessments also has an alternate version for students with significant cognitive disabilities who cannot access the general assessments.[5] Some states also offer alternate academic assessments in languages other than English to allow English learners who are not yet proficient in English to demonstrate their knowledge and skills. In sum, a state must administer sixty assessments annually, with individual students taking as many as four in a given year.

A state must administer sixty assessments annually, with individual students taking as many as four in a given year.

History of Alternate Assessments

The federal government has required academic assessment only since the fall of 1994, when the Elementary and Secondary Education Act (ESEA) was reauthorized as the Improving America's Schools Act (IASA). It required states to administer annual, statewide English language arts and mathematics assessments to all students in specific grades. And it took nearly a decade for all states to comply with that mandate.[6]

Through two subsequent reauthorizations—No Child Left Behind and the Every Student Succeeds Act—states have been required to set their own standards, create their own assessments, and design their own accountability systems within a set of parameters. No federal or national sets of standards, accountability

assessments, or accountability systems have ever been imposed.

Before IASA and the Individuals with Disabilities Education Act (IDEA) of 1997, there was no explicit requirement for educators to apply the same academic expectations for students with disabilities as for all other students. From the early 1970s until 1994, ESEA required assessments only in schools receiving Title I funding, required only testing of students directly receiving services paid for under Title I of ESEA, and allowed districts to choose which assessments to administer.

In those days, the options were mostly limited to tests yielding norm-referenced scores as opposed to scores associated with standards-based achievement expectations. If a district were using a large-scale assessment for accountability or program evaluation purposes, it could choose to use that assessment or an entirely different one for the Title I students in its Title I schools.

Alignment to standards was not required, and if some students—say, those with cognitive disabilities or without English language proficiency—happened to miss out on testing, c'est la vie. There were no negative consequences for not testing students, including those whose absence was essentially a blanket exemption based on their disability, language learning status, or even the likelihood that they would not perform well on the tests. Thus the very students who ESEA was meant to support, with the billions of dollars it directed to education agencies each year, could easily go undetected in evaluations of school performance.

IASA addressed this deficit by requiring states to test all students, regardless of whether they were receiving services paid for by Title I or attending schools receiving Title I funds. The message was that states, districts, and schools must hold the same high expectations for all students, including those who face added challenges to achieving academic success. IASA explicitly included English learners under the all-means-all umbrella.

To allow for all students to participate, a state must offer a general assessment, a general assessment with accommodations, and alternate assessments. Under IASA and its successors, alternate assessments must be based on the same academic content standards as the general assessments but have different performance or academic achievement standards, which define what constitutes proficient performance, performance that exceeds grade-level expectations, and performance that does not yet meet those expectations.

While IASA introduced the requirement for testing all students, it did not impose consequences for schools' failure to do so. States simply had to report aggregated assessment scores for all students as well as scores that were disaggregated for several student groups, including students with disabilities and English learners, on state and local report cards. NCLB attached accountability to these reporting requirements by mandating that at least 95 percent of the students in each student group, such as students with disabilities and English learners, had to participate in the assessments. In addition, each student group—rather than just the total number of students tested—had to meet performance expectations for adequate yearly progress. Under NCLB, all meant all in terms of both participation and performance.

Students with Disabilities. Meanwhile, the 1970 reauthorization of ESEA included as its Title VI grants program the Education of the Handicapped Act, subsequently established as the stand-alone Education for All Handicapped Children Act of 1975. The act required local education agencies to provide a "free and appropriate education" for students with disabilities, which included individualized education programs (IEPs) for each student as well as access to the least restrictive environment (LRE) for their instruction. These requirements were extended and expanded when the act was reauthorized as IDEA in 1990.

In 1997 and 2004, IDEA was reauthorized, in part to align with IASA and NCLB requirements, respectively. Both reauthorizations required states and local school districts to provide alternate assessments suitable for students with significant cognitive disabilities for every test they require their students to take.

English Learners. Assessment requirements for English learners stem from a combination of legislation, memoranda from the Office of Civil Rights in the U.S. Department of Education, and federal case law.[7] Title VI of the Civil Rights Act of 1964 specifically prohibited "discrimination on the basis of race, color, or national origin" in federal or federally funded programs.

While that legislation did not explicitly mention English language proficiency, Congress subsequently passed the Bilingual Education Act of 1968 as Title VII of the 1968 ESEA reauthorization, which provided federal funding "for programs to help language-minority students overcome linguistic barriers that prevent them from meaningfully participating in their educational program."

The NCLB reauthorization of ESEA introduced in its Title III a state education agency role for supporting English learners (then referred to as "limited English proficient" or LEP students) that extended the standards, assessments, and accountability components of the systemic reform model to English language proficiency. All states had to establish English language proficiency standards, assessments aligned with those standards, and an accountability model that used scores from those assessments to identify districts where English learners were not acquiring, or making progress toward acquiring, English language proficiency. ESSA retained these requirements but pulled them into Title I, which applies to all schools and school districts rather than only those receiving Title III funding for English language support services.

Why Offer Alternate Tests?

The logic behind all these testing mandates stems from the model of systemic school reform that emerged in the 1980s and early 1990s.[8] Systemic reform seeks to shift the focus away from mere compliance with rules for the use of school funding and toward evidence that this funding was being used effectively to improve student achievement.[9] Implementing systemic reform means

- articulating expectations for student achievement in each grade in the form of academic content and academic performance or achievement standards,
- assessing student achievement using tests aligned with these standards, and
- evaluating schools based on their students' performance on these tests.

From the beginning and across the reauthorizations of the federal legislation, three key words have been at the core of the testing mandate: accountability, alignment, and all. Accountability means that these tests must yield scores that states use to hold schools accountable for student learning. While the scores may also inform some curriculum or programming decisions, accountability uses are required. That is, states must use the scores to identify schools where students appear to be underperforming and provide resources to support students' achievement in those schools.

Alignment means that the assessments must measure the range of knowledge and skill expectations defined in the state's academic content and performance standards—not just some of them and not some random list of knowledge and skills. Tests that do not align with those clearly articulated academic expectations, which also should be used to guide classroom instruction, cannot yield information of use to teachers, administrators, or parents.

And all means all. All students in the tested grades must participate in the statewide assessment system. Some may despise large-scale testing or at least large-scale accountability testing of the type ESEA requires, often for justifiable reasons. But it is often the "all" requirement that gives many educators, parents, and policymakers pause. All includes every student in every classroom in every school in every neighborhood. All includes every student with disabilities and every English learner, even those who are not yet proficient in English. It is an uncompromising requirement bent on equity and remains a fundamental part of the federal policy logic.

As these long legislative histories reflect, one answer to why states keep testing students is that federal law and dozens of other legal memoranda and legal precedents—some of which are grounded in the Equal Protection Clause of the U.S. Constitution—demand it. ESEA provides significant funding to states and school districts, and student assessment has been a key element in holding state and local agencies accountable for how they use the millions of dollars they get each year. In return for the funding, these agencies agree to be held accountable for the achievement of all of their students.

But the more compelling answer is that alternate assessments benefit students and give incentives to schools to make sure that students with disabilities and English learners

Three key words have been at the core of the testing mandate: accountability, alignment, and all.

Instead of debating how many tests are appropriate, state leaders might instead ask why there are not more tests that involve teacher-student interactions.

have opportunities to learn and meet rigorous expectations.

There are legitimate concerns about the quality of some of the assessments for these groups. It's true that traditional methods for assessing these students' knowledge and skills often fall far short of what is necessary to elicit valid, reliable information about what they can do. Scores for these individual students may reflect more errors in targeting knowledge and skills than scores for other students. However, the benefit that students, their families, and their communities accrue through the use of inclusive strategies and support for their academic achievement outweighs concerns about testing precision.

High-quality alternate academic assessments have been recently developed, such as those administered by the Multi-State Alternate Assessment and the Dynamic Learning Maps consortia.[10] Recently developed academic English language proficiency assessments, such as Access for ELLs and ELPA21, are far more technically sound and rigorous than their predecessors.[11] Such progress in testing rigor would not have occurred without the ESEA policy mandates.

What State Boards Should Take Away

Students with disabilities and English learners offer a window into what large-scale testing could look like for all students if state and federal policies loosen their footing in traditional testing models and allow for more innovative, student-centered approaches. The tenets of accountability, alignment, and "all" would still apply—and likely with greater success—if state leaders would embrace and expand the lessons of alternate assessment.

In fact, high-quality alternate assessment and English language proficiency assessment formats should prompt state boards of education to rethink the nature of general assessments. Instead of debating how many tests are appropriate, state leaders might instead ask why there are not more tests that involve teacher-student interactions, administered at times and in ways that directly support students' learning. Certainly, there are time and cost implications for shifting to more student-centered, performance-based assessment formats. But if students, their teachers, and their parents can actually learn from the testing process itself, that seems an investment worth exploring. Especially when the techniques necessary for building such tests would also demand stronger alignment with learning expectations and deep attention to the equity and accessibility of learning and assessment contexts.

The models on which current statewide accountability tests are based reflect tradition more than best practice. Innovative, student-centered assessments that are grounded in instruction and elicit rich evidence of student thinking appear more likely to effect real, sustainable changes in student learning than assessments that are distal to and separate from classroom interactions.[12] ■

[1]Charlene Tucker and Ellen Forte Fast, "Autopsy of a Canary: Searching for a Theoretical Framework to Understand the Underperformance of Students from Poor Families and Students of Color," paper presented at the CCSSO Annual Large-Scale Assessment Conference, Snowbird, Utah, 1999.

[2]Thomas Hehir and Lauren Katzman, *Effective Inclusive Schools: Designing Successful Schoolwide Programs* (Indianapolis: Jossey-Bass, 2012).

[3]Diane August et al., "Recent Research on English Language and Literacy Instruction: What We Have Learned to Guide Practice for English Language Learners in the 21st Century," in Marilyn Shatz and Louise C. Wilkinson, eds., *The Education of English Language Learners: Research to Practice* (New York: Guilford Press, 2010).

[4]Ellen Forte, Rachel Quenemoen, and Martha Thurlow, "NCSC's Theory of Action and Validity Evaluation Approach," NCSC Brief #9 (Minneapolis: University of Minnesota, National Center and State Collaborative, 2016); Rachel Quenemoen, Claudia Flowers, and Ellen Forte, "The Curriculum, Instruction, and Assessment Pieces of the Student Achievement Puzzle," in Fred Spooner, ed., *More Language Arts, Math, and Science for Students with Severe Disabilities* (Baltimore: Brookes Publishing, 2014).

[5]General assessments are those academic or English language proficiency tests taken by most students and are distinguished from "alternate" assessments.

[6]William Erpenbach, Ellen Forte Fast, and Abigail Potts, "Statewide Educational Accountability under NCLB: Central Issues Arising from an Examination of State Accountability Workbooks and ED Reviews under the NCLB Act of 2001" (Washington, DC: CCSSO, 2003).

[7]Ellen Forte, *The Administrator's Guide to Federal Programs for English Learners*, 2nd ed. (Washington, DC: LRP, 2017).

[8]Michael Smith and Jennifer O'Day, "Systemic School Reform," *Journal of Education Policy* 5, no. 5 (1990).

[9]David Cohen, "What Is the System in Systemic Reform?" *Educational Researcher* 24, no. 9 (1995): 11–17, 31.

[10]Arizona Department of Education, "Assessment" website, https://www.azed.gov/assessment/msaa/; Dynamic Learning Maps, website, https://dynamiclearningmaps.org/.

[11]WIDA, "Access for ELLs," website, https://wida.wisc.edu/assess/access; ELPA21, "The ELP Assessment System," website, https://www.elpa21.org/.

[12]Linda Darling-Hammond, "Developing and Measuring Higher Order Skills: Models for State Performance Assessment Systems" (Washington, DC: CCSSO, 2017).

Ellen Forte, Ph.D., is CEO and chief scientist at edCount, LLC.