

FairTest

The National Center for Fair & Open Testing

Interim, Formative and Through-Year Assessments— Keeping Assessments Authentic and Useful (Assessments for Education!)

How Do We Assess Student Learning?

Classroom teachers assess student learning every day. They do this by monitoring student in-class performance, assigning and reviewing homework, giving quizzes, assigning and grading in-class tests, essays and projects. These assessments are inextricably bound with curriculum and classroom subject matter. They provide educators with information about students that allow for tailoring material and approaches to facilitate student learning. Assessments that cover a range of material could be described as summative; those along the way could be characterized as formative or interim. These are kinds of “interim” or “formative” assessments that have been a part of schooling since the development of the common school and comprehensive high school. They were not the purview of testing companies.

Summative Test-Based Accountability

Because of disparities in education quality across schools and districts, there was a feeling that more common standardized measures were needed by which to judge student and school performance. Driven by the *Nation at Risk* report of 1983, the United States has been on a 40-year course of using summative standardized tests as the core of student and school accountability measures. By summative tests we mean those that assess learning over a semester or year reflective of a set of state learning standards, but were not necessarily tied to discrete curricular elements. This system reached its apex with the passage of No Child Left Behind (NCLB) in 2001 with the requirement that students take end-of-year tests with schools required to make Adequate Yearly Progress or be deemed failing and slated for remedial measures, including state takeover. The data for each school was to be disaggregated by subgroup, notably race, to demonstrate the achievement disparities and allow resources to be targeted where they were most needed. The noble, if ultimately misguided, goal of the testing was to make sure that kids weren't being left behind and force schools to address student learning needs.

Punishing schools for failing to meet summative testing score goals failed to produce the improvement desired. Crucially, it came with a price—crowding out of non-tested subjects, excessive test prep rather than deep authentic learning, and

FairTest

The National Center for Fair & Open Testing

stigmatization of schools and districts largely with socioeconomically disadvantaged populations. Exasperation with testing mandates prompted a large opt-out movement in the early-mid 2010s. Congress's response to the shift in public mood and the collateral damage of NCLB was to pass the Every Student Succeeds Act (ESSA) in 2016. While test results no longer carry the mandated consequences of NCLB, every student still has to be tested in ELA and Math every year in grades 3-8 and once in high school. They must also be tested in Science in high school.

Summative tests developed by states or, more frequently, large testing companies for states and administered at the local level, remain the main element of school accountability systems.

What is Happening: The Drive to Shorter, More Frequent Testing

The decades of NCLB and ESSA testing and the backlash against those excesses and negative consequences have left the field of mandated assessments in flux. On the one hand, teachers in schools across the country continue to assess students on their knowledge, skills and mastery of classroom content with an eye towards a final course grade. Layered over and pressuring that system, however, are accountability measures imposed by the federal, state and local governments. Those testing systems add layers to classroom requirements and inevitably make classroom teachers gear their approach and sequencing to prepare students for those external tests, rather than tasks and assessments devised to meet the needs of the students before them.

The trend on the part of state and local authorities is to try to get more actionable information from mandated tests. The rallying cry of “learning loss” from the COVID pandemic provided an opening for testing companies to sell their wares to districts desperate for solutions to address the negative impacts of remote schooling. The good intentions of having tests that are instructionally useful for individual students are nullified by the external origins and inherent misfit of standardized tests with the actual learning process. Rather than leaving granular decisions about assessment to the classroom and school level, the window has opened for commercially generated products divorced from the classroom experience to dominate school in an even more unproductively intrusive way than the summative accountability assessments giving stakeholders pause.

FairTest

The National Center for Fair & Open Testing

Interim, Formative or Benchmark Assessments

Benchmark or interim assessments are assessments administered to measure student learning against academic standards to support decision-making at the classroom, school, or district levels. To actually assist with student learning, however, they have to be tied to the curriculum, rhythm and pacing of the classroom. By imposing external interim assessments, the teacher's professional function is usurped. The assessment tasks of the student become divorced from what they are authentically experiencing in the classroom setting. Less room is available for demonstrations of student learning that are performance and project based, have greater connection to the real world, and that enable students to have greater say in and control of their learning.

The proliferation of commercial produced district/school purchased interim assessments has been staggering. According to a Rand Corporation study¹, in the 2021-22 school year 87% of K-8 Schools surveyed reported using commercially produced interim assessments (the largest being iReady, MAP, and Star); 77% reported using locally created assessments (state, local or school). 62% of High Schools reported using commercially produced interim assessments, while 84% used locally created assessments.

Not surprisingly, according to RAND, teachers who used self-created benchmark assessments, served students performing at or above grade level, and spent more professional learning time on analysis or use of student assessments were more likely to perceive alignment between their benchmark assessments and curriculum materials. Teachers' perceptions about the alignment between their benchmark assessments and curricula were linked to how easy or difficult it was for them to identify tasks or activities within their curriculum materials to address students' unmet learning needs.

In other words, when teachers create their own interim or benchmark assessments (frequently done in grade or subject matter teams), those assessments are more **instructionally useful** and **educationally relevant** than externally imposed commercial products. They are also much more likely to provide students with greater agency in their learning.

¹ Doan, Sy, Joshua Eagan, David Grant, Julia H. Kaufman, and Claude Messan Setodji, American Instructional Resources Surveys: 2022 Technical Documentation and Survey Results. Santa Monica, CA: RAND Corporation, 2022.
https://www.rand.org/pubs/research_reports/RRA134-14.html.

FairTest

The National Center for Fair & Open Testing

Additionally, the standardized interim or formative assessments do not produce data that allows stakeholders to make accountability related comparisons. Data from formative assessments are not designed to be aggregated up beyond the classroom level and thus are not well suited to make comparisons in student achievement within and across schools.

Commercial interim assessments, in trying to merge two discrete functions, do neither well. They are both poor learning devices and accountability tools.

Through-Year Assessments: Hoax not Savior

Through-year assessment models administer multiple tests throughout the school year as part of an assessment system designed to produce a single summative score meeting federal and state accountability requirements. The claim of proponents is that through year assessments connect what students learn to what they are tested on, providing timely feedback to teachers, students and families. The reality is that through-year assessments, given their attenuation from actual classroom learning, are hard pressed to be instructionally relevant or educationally useful. They also suffer from the psychometric challenges of validity and comparability when they are aggregated to form a single summative score. Their utility for either purpose is highly questionable.

Thirteen states are currently undertaking experiments in through-year assessments for ESSA compliance purposes.² The Center for Assessment in its report, “Through Year Assessments: Ten Key Considerations”³ noted several consequences of a through-year system that undermine its intended impact.

1. Through year systems will impact the substance and timing of local curricular decisions. Even with semi-flexible testing windows, schools and districts will have pacing and content decisions essentially imposed on them in a way that summative accountability tests do not.
2. The current reporting of through-year assessments make them less than useful for instruction and student improvement. A simple numerical

²<https://www.education-first.com/insights/our-resources/innovations-in-assessment/through-year-curriculum-connected-assessments/>

³<https://www.nciea.org/wp-content/uploads/2023/03/Ten-Key-Considerations-Through-Year-Assessment-Report-March2023-F.pdf>

FairTest

The National Center for Fair & Open Testing

score, even with section breakdown, does not give adequate information to students and teachers about what the class and individual need to do in order to facilitate authentic student learning.

3. Through-year assessments may compound existing tensions between accountability and instruction: more perceived high-stakes tests will have a negative effect on instruction.
4. Overall testing time will go up, unless state through-year assessments replace existing interim assessments.

Florida is an empirical example of the impact of through-year testing on overall testing time, which proponents of the practice argue would be reduced. In 2021-22, the total estimated time for state testing in Florida was: 4,545-4,550 minutes for K-12, plus an average of 177.5 minutes for the School Day SAT/ACT. In 2023-24, the total estimated time for state testing is: 6,580-6775 minutes for VPK-12, plus an average of 143 minutes for the School Day SAT/ACT/CLT. **With F.A.S.T. testing, the estimated time for state testing has increased by over 2,000 minutes; that is more than 33 hours of increased state testing over a student's school career.**

If possible, chunking up a summative assessment used for federal accountability may be worse for education than the single end of year instrument.

Assessment for Education: Addressing The Needs of Students and Teachers

Trying to use mass produced commercial testing instruments to drive learning, whether they are administered annually or in smaller doses, does not produce the education stakeholders want and need. The assessments that give teacher and student instructionally useful information about what's going on in real time are those developed and administered at the school and teacher level. Summative assessments arrive too late to be genuinely useful; commercial interim assessments do not evaluate authentic tasks or actual classroom practice.

Schools would do well to engage in a balanced approach with performance-based assessments and classroom evaluations (quizzes, exams, tests, projects) to drive educational goals and outcomes, and standardized tests to provide a low-stakes thermometer for systemic recalibration. Districts, schools and teachers should develop

FairTest

The National Center for Fair & Open Testing

and use assessments that are aligned with curriculum, engaging to students, and that allow them to demonstrate learning through authentic tasks. Standardized tests can be a periodic check to see if schools and systems are broadly serving students, are meeting the needs of disadvantaged populations, and meeting some external standards. But they are external to the daily process of education. How student learning is managed and improved should not be the purview of external standardized instruments.