

“Interim” Assessments and ESSA

A Great Opportunity

Stuart Kahl | January 2017

The Every Student Succeeds Act (ESSA) places the details of accountability systems back in the hands of the states. It also includes some assessment-related provisions that allow the states to implement innovative solutions to longstanding issues regarding state testing programs. Two major issues have to do with (1) the nature of the content and skills that have been tested via the efficient end-of-year No Child Left Behind (NCLB) assessments (and resultant negative impacts on instruction, classroom testing, and student learning) and (2) the over-testing spawned by NCLB and the USDOE’s Race to the Top program.

Efficient Testing under NCLB

The extensive testing and quick turnaround of results required by NCLB encouraged the use of efficient test instruments dominated by the multiple-choice format. These tests addressed what might be called “foundational knowledge,” but not the deeper learning that can be demonstrated by the application of lower-level knowledge and skills to the solution of real-world problems. Of course as history has shown, the focus of high-stakes testing becomes the focus of instruction and classroom assessment—what is measured gets taught. An original intent of the Smarter Balanced and PARCC state assessment consortia, funded by Race to the Top, was to include significant performance assessment components that tapped deeper learning. In the end, however, efficiency ruled the day. Both consortia scaled back their plans for performance assessment, using short-term, on-demand, secure exercises. Although these components are of high quality, they’re not significantly different from the non-multiple-choice components some states have used in the past.

Over-Testing

The high stakes associated with the results of NCLB testing were raised even higher by Race to the Top requirements pertaining to the use of results of

student achievement tests for teacher evaluations. These high stakes were almost certainly a factor in the increased local use of interim tests from publishers and from the states themselves for purposes of preparation for end-of-year testing and early warning—i.e., the identification of students or curricular topics requiring extra attention before the high-stakes summative testing.

Of course, interim assessments of different types have legitimate and valuable uses. Concerns about over testing should be raised when some tests are unnecessary or duplicative. The main focus of this paper, however, is on the ESSA-allowed use of interim assessments for purposes of accountability.

ESSA and Multiple Measures

Both NCLB and ESSA stated that the assessments “shall...include multiple up-to-date measures of student academic achievement, including measures that assess higher-order thinking and understanding” To some extent, this requirement of NCLB was ignored. States that administered all or predominately multiple-choice tests were probably not living up to the spirit of the provision. ESSA added to the statement, “... which may include measures of student academic growth and may be

partially delivered in the form of portfolios, projects, or extended performance tasks.” Such approaches to performance assessment were not prohibited before; but by identifying them the new law clarifies the intent of the multiple-measures requirement.

ESSA and “Interim” Assessments

ESSA allows the accountability assessment requirement to be met via a “single summative assessment . . . or through multiple statewide interim assessments during the course of the academic year that result in a single summative score. . . .” Given this flexibility and the specific mention of performance assessment alternatives, a state or consortium of states could be truly innovative by implementing a two-component assessment system consisting of, for example, curriculum-embedded performance assessments and a shortened end-of-year summative assessment. Such an approach could address two common concerns about accountability assessment under NCLB—negative instructional impact and too much time spent on testing. Curriculum-embedded performance assessment is just one of several approaches to interim assessment. In designing an interim component for an accountability assessment program, many issues associated with the different approaches need to be understood.

“ . . . a state or consortium of states could be truly innovative by implementing a two-component assessment system . . . ”

Types of Through-Course Assessments

The term “interim assessment” is often used as a general term for various types of through-course assessments. For the purposes of this paper, the term “through-course assessment” is used for the general term, and “benchmark assessment” and “interim general achievement measure” are used for two particular types of through-course assessments. A *benchmark assessment* covers content and skills recently covered by instruction. By this definition, teachers’ end-of-unit or marking period tests are benchmark tests, and the teachers create these themselves, sometimes making use of items provided by vendors or the states. In any case, several benchmark assessments administered at different times during the school year would be needed to measure understanding of a full year’s content. An *interim general achievement measure*, on the other hand, is made up of a sampling of items or tasks spanning a full year’s worth of curricular content. These assessments can be used as through-course tests for *monitoring growth* and/or as *early warning* tools.

Appropriate Through-Course Assessments for ESSA Accountability

It would make no sense to use several general achievement measures as through-course assessments for annual accountability assessment purposes, aggregating their results to produce summative scores, for a number of reasons.

- Student performance at several points during the school year would be lower than at the end of the year, and an average of a student’s scores on such interim tests would not accurately reflect where the student stands at the end of the year.
- For purposes of monitoring growth, one would not want to use too many general achievement measures during the year. If there is a short time between successive tests, actual growth would be minimal and often overshadowed by measurement error. Student scores could actually fluctuate even when there is real growth.

- Aggregating “growth” over short periods of time within a year would be worthless because of the compounding of measurement error associated with many growth measures. What meaningful growth measures used for annual accountability need to show is growth over a full year—i.e., the difference between comparable achievement scores from the end of the year and the beginning of the year (or end of the previous year).

Thus, if results of assessments administered “during the course of the academic year” are to be aggregated for purposes of ESSA accountability, some form of benchmark testing would be most appropriate. However, this type of testing is not without its own issues.

Formative Assessment Is Not Frequent Through-Course Assessment

It is not unusual to see policy makers and some educators refer to interim or benchmark assessments as “formative assessments” and call for the use of scores from them in a state’s accountability system. The formative assessment that research has shown to be particularly effective in advancing student learning is a multi-step instructional process, in which one step is evidence gathering relative to student learning of a specific learning target addressed by instruction at the time. Next steps involve descriptive feedback and instructional adjustments to address learning gaps identified through the evidence-gathering process.

Evidence can be gathered by any of a number of formal and informal measures. This step is accomplished **during the learning**—likely before the student has reached the level of proficiency with respect to the learning target that he or she will reach by the end of the unit of instruction. So it is often not appropriate to grade the evidence (student work) for summative purposes. (There are certainly gray areas between formative and summative assessment, such that performance on summative measures can inform future performance, but further discussion of formative assessment is beyond the scope of this

paper.) The significant point is that benchmark or interim assessment, whether in the form of teachers’ unit or marking-period tests or external (publisher- or state-provided) measures, is a form of summative assessment. Certainly this is the case for through-course measures to be used as components of a state accountability assessment system.

“What through-course assessment approaches can provide timely and useful information to teachers and at the same time, add meaningfully to a state’s accountability assessment system?”

Benchmark Measures of Foundational Knowledge and Skills

Teachers often complain that end-of-year state tests include items assessing content and skills that are taught many months before the tests are administered. Many would like to see their states rely on benchmark assessments that cover material recently taught. This approach presents several potential problems. (Remember, the ESSA refers to **statewide** interim assessments.)

First, concerns about the security of high stakes tests would mean the states would likely want to maintain control over the testing schedule and testing material (or online access to the tests). This would mean much more “external” testing and test prep in the schools, as well as added intrusion into the scope and sequence of schools’ curricula. While one might suggest that teachers’ own tests be used, that approach would not be acceptable for ESSA accountability because the states would have no

control over the technical quality of the instruments and there would not be the required comparability of results across schools within states.

Another problem with the use of benchmark testing as described above is that the results on tests administered right after instruction would not reflect how well students retain their understanding at the end of the year. And retention of knowledge and skills should be a goal of any instructional program.

“Unfortunately, immediacy of results has become the expectation of all testing, under the unfounded assumption that all tests . . . should inform teachers’ daily decisions in real time.”

Efficient state tests focus primarily on foundational knowledge and skills—which are knowledge and skills deemed important by committees of teachers and curriculum specialists who work with the state departments of education and their contractors to develop the tests. While people may be concerned about time spent on test prep activities associated with end-of-year testing, such efforts do provide for reinforcement of concepts and skills so important to the learning process. Considerable prep time is provided for a school’s semester and final exams, but that does not seem to be a concern. Review periods for final exams are common and good practice.

The Match between a Test and Its Use

Unfortunately, immediacy of results has become the expectation of all testing, under the unfounded assumption that all tests, including general achievement measures used in statewide

assessment programs, should inform teachers’ daily decisions in real time. It’s important to understand how each kind of test is best used and to be precise in our terms when discussing various test types.

End-of-year summative tests from external sources (e.g., the state) provide useful information to guide programmatic improvements to benefit the next class of students passing through a tested grade. However, such a general achievement measure, with its “thin” sampling of a whole year’s worth of curricular content, can in no way provide the kind of timely diagnostic information needed for teachers’ daily decision making.

Interim general achievement measures administered part way through the academic year can be useful for early warning and growth monitoring purposes, but as suggested earlier, one wouldn’t want to use too many of them.

Intact benchmark tests (measuring a subset of a content domain) from external sources are of value if their content is a good match to what the teachers have actually been teaching. However, many externally available tests are not well matched to local curriculum, so their value is limited. Teacher-made benchmark (e.g., unit) tests, of course, are well matched, although item quality could be poor. Instead of using intact tests from external sources, teachers can access item banks from external sources to select high quality test items and tasks that closely match their instruction. They can use these resources for formative evidence gathering or in lieu of totally self-developed unit or marking period tests.

Through-course assessments from external sources have limitations relative to teachers’ day-to-day decision making—and can contribute to over-testing and disruption of school schedules. And benchmark tests don’t assess the retention of important knowledge and skills. So the logical question is, “What through-course assessment approaches can provide timely and useful information to teachers and at the same time, add meaningfully to a state’s accountability assessment system?” In other words, “What’s missing from a state’s end-of-year assessments that can address those two needs?”

Curriculum-Embedded Performance Assessment

Part of the answer could lie in the use of curriculum-embedded performance assessments. Hofman, Goodwin and Kahl (2015)¹ define a curriculum-embedded performance assessment (CEPA) as an instructional module or unit that involves a series of instructional and assessment activities, the latter sometimes yielding student work (evidence) that can be used for formative purposes and at other times leading to student work that can be evaluated for summative purposes. Marion and Shepard (2010)² refer to “replacement units,” which are similar to CEPAs although those authors focus primarily on formative uses.

Performance assessment can address two needs that efficient end-of-year state tests can’t meet: It can inform teachers’ day-to-day instructional decisions, and it can assess deeper learning. However, many past efforts at performance assessment in statewide programs were extremely burdensome on teachers and schools—they just added more and cumbersome testing that wasn’t tied to local curricula.

ESSA’s interim assessment provisions provide the opportunity to implement performance assessment that is truly curriculum-embedded. CEPAs strike a needed balance—between formative and summative assessment, between attention to foundational knowledge/skills and deeper learning, and between local and state components of an accountability assessment system, without adding burden to teachers.

- How much additional testing time is required by the use of CEPAs? Absolutely none! These are high quality instructional units that can replace others that teachers have used in the past.
- The performance assessments within CEPAs test deeper learning—what the efficient state tests have neglected.
- As the Hofman et al (2015) paper proposes, the CEPAs can be developed by teachers, but submitted to the state for revision and piloting just as other state assessment components are.

- Results are immediate because teachers score their own students’ work themselves and use those results however they choose—for feedback to students and/or grades.
- The state can implement a scoring audit process to determine necessary score adjustments and assure CEPA summative scores are comparable across schools.

“Performance assessment can address two needs that efficient end-of-year tests can’t meet: It can inform teachers’ day-to-day instructional decisions, and it can assess deeper learning.”

Thus, immediacy and utility of results, performance task quality, and technical/measurement quality can all be accomplished by the CEPA approach as described in Hofman et al (2015).

Two-Component Accountability Assessment

ESSA requirements allow a system of accountability assessment that merges the results from an efficient end-of-year state test with those from through-course performance assessments. The efficient end-of-year measure should include a variety of item types in addition to selected-response, but can still be relatively short since it comprises only part (perhaps half) of the total student assessment score. The least intrusive approach to incorporating periodic assessment would be truly curriculum embedded.

Significant performance tasks within instructional units would provide the student work to be scored both for immediate classroom use and, with auditing, for accountability purposes.

During the course of the school year, a teacher could choose, say, three state-approved CEPAs. These would be high quality instructional units initially developed by teachers, but revised appropriately and field-tested just as other state test components are. Teachers would choose to use CEPAs in place of other units to address the same content standards. .

Over time, large numbers of such units can be made available to teachers. Scheduling of testing “windows” (time spans) and security concerns would be irrelevant. For the tasks within CEPAs used for summative purposes, of course, teachers would follow administration directions just as they are expected to do for traditional assessments

This potential approach to through-course assessment for accountability has several advantages. Why use the “interim” assessment option of the ESSA to do what efficient end-of-the-year state tests can do better? CEPAs tap the deeper learning that the state tests do not measure adequately, and they can do so in a non-intrusive way that doesn’t simply add testing.

The passing of ESSA has provided an opportunity for states to design programs that address major concerns that educators and non-educators have expressed over the years about state testing programs. Many approaches to through-course testing for accountability purposes are flawed and would not offer remedies to those concerns. The two-component approach described here can work. Now’s our chance to *get it right*.

¹ Hofman, P., Goodwin, B., & Kahl, S. (2015). *Re-balancing assessment: Placing formative and performance assessment at the heart of learning and accountability*. Denver, CO: McREL International.

² Marion, S. & Shepard, L. (2010). *Let’s not forget about opportunity to learn: Curricular supports for innovative assessments*. Dover, NH: Center for Assessment.