



WHAT'S WRONG WITH HIGH SCHOOL TESTING AND WHAT CAN WE DO ABOUT IT?

April 24, 2018

Scott Marion | Executive Director | Center for Assessment



National Center for the Improvement
of Educational Assessment¹
Dover, New Hampshire

CENTER FOR ASSESSMENT

WHAT'S WRONG WITH HIGH SCHOOL TESTING AND WHAT CAN WE DO ABOUT IT?

What's wrong with asking a high school student to try as hard as they can on a test that does not count for her grades, cannot be used for college admission or placement, and is barely related to things she is learning in her current classes? Just about everything! It seems like every few weeks we hear about a new twist on high school testing in the U.S.; testing that is, at least in part, mandated by federal education law. As someone who grew up taking the New York State Regents exams, I've been thinking about high school testing for a long time. I offer some thoughts about the current state of high school testing, exploring the tradeoffs with some of the common approaches. I conclude with some thoughts about how we might improve the current state of affairs.

¹ I acknowledge the terrific feedback on previous drafts from my colleagues at the Center for Assessment and to Randy Bennett (Educational Testing Service) for his insightful comments. Any errors and omissions are my own.

² This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



PURPOSES AND USES

We can't talk about high school testing or any testing for that matter, without first clarifying some requirements as well as intended the purposes of the testing and the intended uses of the test scores. The primary U.S. federal education law—the Every Student Succeeds Act (ESSA); the

The primary U.S. federal education law—the Every Student Succeeds Act (ESSA); the latest instantiation of the Elementary and Secondary Education Act of 1965—requires that states test all students in grades 3-8 and at least once in high school in English language arts (ELA) and mathematics.

latest instantiation of the Elementary and Secondary Education Act of 1965—requires that states test all students in grades 3-8 and at least once in high school in English language arts (ELA) and mathematics. Additionally, states are required to test students in science at least once in each grade span. The ELA and mathematics test scores must be used in states' school accountability systems.

Therefore, a key use of at least some (or all) high school test scores is to support school accountability determinations. A limited number of states use high school test scores as part of educator evaluation systems and approximately a dozen states require students to pass a single or a set of exams in order to be eligible to graduate from high school. Finally, almost all states included a postsecondary readiness indicator in their ESSA school accountability systems of which exams such as the SAT and ACT play a prominent role. Even before the ESSA became the law of the land, many states had instituted census testing of the ACT or SAT to support college readiness initiatives. With so many potential purposes and uses, a discussion of high school

testing can quickly become unwieldy; therefore I focus on testing for the purposes of measuring student achievement for use in states' school accountability systems.

It is important to identify these nominal purposes, but to truly understand the testing landscape, particularly at the high school level, we need to dig a little deeper to both get more specific about the purposes and uses as well as to articulate the "claims" we want to make based on the assessment results. A student scoring proficient on the 11th grade math test has demonstrated competence on the required high school math standards and will likely be successful in postsecondary mathematics courses is an example of a claim that state assessment leaders might want to make regarding their high school assessment. Simply saying the purpose is school accountability does not offer enough information to guide test design. What is the purpose of the accountability system? The Elementary and Secondary Education Act, of which ESSA is the latest version, has a clear focus on evaluating and enhancing equality of educational opportunity. Therefore, an assessment system in support of such an accountability system must provide information about the extent to which students have an opportunity to meet the intended learning goals. An accountability system focused on prioritizing excellence would likely lead to a different sort of assessment system. Many states are trying to promote both excellence and equity, an admirable endeavor, but this means that state leaders have to be exceptionally thoughtful about the assessments they employ.

CURRENT APPROACHES

There are two basic types of high school testing: Single grade (e.g., 11th grade) survey tests or end-of-course tests. As the name implies, end-of-course tests are tied to specific high school courses (e.g., American Literature, Life Science, Algebra) where only those students participating in the course sit for the exam. A survey test is administered to all students in a given grade designed to broadly cover the grade level or grade span content standards in that subject area. The use of a “nationally-recognized college entrance exam,” when used as the achievement

A survey test is administered to all students in a given grade designed to broadly cover the grade level or grade span content standards in that subject area. The use of a “nationally-recognized college entrance exam,” when used as the achievement indicator, is a special case of the survey test.

indicator, is a special case of the survey test but given the rapidly increasing use of such tests, I discuss it separately below.

Survey Test

Most states employing a survey test approach typically administer tests in ELA and mathematics to all students in grade 11. A few states do so in 10th grade, but I do not know of any states that administer a survey test to 12th grade students. If the survey test is designed to evaluate what students have learned relative to the high school standards, it might make sense administer the test in 12th grade to provide a more complete picture of high school achievement. Are we worried that seniors would not take seriously a test that did not “count” for them? If so, wouldn't we have the same concern about juniors? Thus, motivation is a key consideration in evaluating the validity and usefulness of the scores for school

accountability purposes because the test results, in almost all cases, do not count for students and because the test content is not closely related to what students are learning in their courses at the time. Importantly, motivation is influenced by many other factors such as cultural background and gender so trying to understand motivational effects at some average level will be incredibly misleading. This is especially problematic if the tests ask students to draw on factual and/or basic procedural knowledge from memory rather than applying things they have learned over the years at some deeper level. A related problem is that students taking the exam vary considerably in their course-taking patterns and prior preparation. Without exaggeration, some students taking an 11th grade math survey test will be enrolled in an AP Calculus class while others might still be enrolled in remedial algebra. Besides the inferential challenges of trying to make sense of scores for students so differentially prepared, this poses a considerable motivational challenge for students for whom the test is far too easy or far too challenging. These motivational and instructional sensitivity shortcomings limit the potential utility of these assessments because educators trying to use the results to evaluate curricular and instructional programs have to either ignore these threats or try to figure out how to account for them in their evaluations.

In spite of these considerable challenges, survey tests offer a convenient way to meet federal and state requirements. State leaders can easily document the alignment of these tests to the body of knowledge and skills students are supposed to learn. This is no small thing. In a standards-based accountability system, alignment is a promissory note that says to educators “if

you teach the content standards well, your students will have a fair opportunity to demonstrate their knowledge of the standards on the assessment.” Another critical advantage for using a survey test approach to meet federal requirements is that ESSA, like the original ESEA, is focused on ensuring that typically underserved students receive appropriate educational opportunities and that these opportunities need to be evaluated by testing all students. Having all students in

Another critical advantage for using a survey test approach to meet federal requirements is that ESSA, like the original ESEA, is focused on ensuring that typically underserved students receive appropriate educational opportunities and that these opportunities need to be evaluated by testing all students.

a specific grade take the same test at the same time prevents certain students or groups of students from being “hidden” in the assessment and accountability systems thereby providing a comparable evaluation of equality of opportunity to learn. This is played out through the statewide assessment and accountability system by disaggregating and monitoring the results for each of the demographic subgroups both within year and over time. Such monitoring can occur with the end-of-course model discussed below, but it is much more straightforward to evaluate the performance of the various subgroups of students when assessed using a common assessment for all students administered at roughly the same time in their educational careers.

End-of-course testing

End-of-course (EOC) tests are common in approximately one-half of the states. In certain states, the EOC test results are required to be incorporated into course grades, while in other states they are prohibited from counting toward student grades. If the assessments are high-quality and aligned to the specific course content, then the

results should be allowed to count in student grades, depending on the wishes of the local school leaders. A major challenge with EOC tests is determining which courses to test. Anyone who has looked recently at a comprehensive high school course catalogue knows that there are hundreds of courses. It would be a financial and logistical nightmare to try to have an EOC testing system that covers most courses. Therefore, states have to prioritize which courses they want to include in its EOC testing system. States with EOC testing systems generally test in courses that are required for all students—especially when the results must support school accountability systems—such as Algebra 1, Geometry, English 9, English 10, Life Science, and perhaps one of the physical sciences. Some states also include EOC exams in commonly-required courses like U.S. History, World History, U.S. Government, and perhaps Economics.

There are many benefits of a high-quality EOC exam system, including potentially raising and creating shared expectations across the state and ensuring that students are evaluated using exams that are generally higher-quality than those created locally. However, there are some challenges associated with an EOC exam system. The first, discussed already, is prioritizing which courses will be tested and determining how the results should be used. The second, which is the converse of shared expectations, is that EOC tests, like what is observed with Advanced Placement (AP) exams, tend to shape course content and instruction, which will

reduce local control. Some might consider this a benefit. The costs and capacity necessary to maintain a high-quality EOC system can be considerable. It costs about as much to develop a single 11th grade survey test as it does to develop one EOC test. Therefore, every additional test employed multiplies the cost of high school testing. Additionally, every test requires direct supervision by state personnel to ensure that the state is getting what has been promised and at the level of quality negotiated. Therefore, more testing means more money to hire more state personnel. It goes without saying that money is far from unlimited and what is spent on high school testing could come at the cost of other assessment opportunities. Since all students do

Depending on the number of courses used in the school accountability determinations, ensuring that all students are “counted” appropriately requires careful attention of state and district leaders.

not complete the EOC tests at the same time—by design—the state needs an efficient data system in order to track student course-taking and maintain test performance records to aggregate results according to well-conceptualized business rules. Depending on the number of courses used in the school accountability determinations, ensuring that all students are “counted” appropriately requires careful attention of state and district leaders. Algebra I represents a good case study of some of the more serious school accountability challenges with an EOC system. Students typically complete Algebra I in 9th grade, but many higher performing students do so in 8th grade (or even 7th) while some lower-achieving students may take two years to complete Algebra I, finishing the course in grade 10. Many states have “banked” the 8th grade

Algebra I score for use as part of the high school accountability system. But does it make sense to “reward” high schools for the generally higher performance of those students who take Algebra I in middle school? Conversely, if these scores are not counted for high schools, then it is easy to imagine an unintended negative consequence of districts limiting the number of students permitted to take algebra in middle school. Admittedly, algebra is a considerable challenge, but there are many related challenges when trying to use EOC test scores to produce a picture of high school achievement. For example, depending on the number of EOC exams used, school personnel might see the multiple EOCs as just more ways to “fail.” The EOC approach attempts to balance the equity and excellence demands, but to do so requires common required courses across the state.

Nationally-recognized college entrance exam

The Every Student Succeeds Act invites states to consider using a “nationally-recognized college entrance exam” for its required high school assessment in ELA and mathematics. The regulations defined this phrase to apply to the ACT and SAT almost exclusively. The law and regulations allow districts to choose among the college readiness assessments if the state implements a quality control process to allow such choice while maintaining comparability. For example, both Florida and Georgia passed legislation to allow districts to potentially substitute ACT or SAT scores for end-of-course test results. Both of these potential use cases have significant practical and comparability challenges and will likely have few, if any, districts pursuing such an option if it is even allowed. In most cases, states using the ACT or SAT are doing so on a statewide level. Because of the additional complexity introduced when considering a district choice approach, I focus this discussion solely on the case of a statewide adoption of

either the ACT or SAT for use as the single high school achievement indicator. Achieve recently released a [report](#) strongly opposing the use of the ACT or SAT as the measure of high school achievement. I agreed with much of the Achieve report, but I was concerned that they did not appear to recognize the context within which many of state assessment systems operate.

In spite of claims made by the companies, the few independent alignment studies conducted to evaluate the relationship between the ACT/SAT and state content standards have questioned the match between the tests and the standards students are expected to learn. Assuming these alignment studies are generally accurate, this would mean that only part of the standards would be tested and many worry, justifiably, about the narrowing of the curriculum because of the “what gets tested, gets taught” phenomena. This is a legitimate concern and violates the promissory note discussed earlier. Further, under a standards-based approach, the use of non-aligned tests may challenge the validity of school accountability inferences.

I argue that, in many cases, states are making a rational decision to use the ACT or SAT as the achievement indicator. While many state policy makers do not understand or do not care about the alignment concerns, they are happy to reduce some testing and provide a visible benefit to many of their constituents.

However, we must keep in mind that this use is not being compared to the “perfect” testing system. In almost cases, using the ACT or SAT as the high school achievement indicator replaces a single survey test, therefore the validity threats associated with the ACT or SAT must be considered in light of the validity threats of the single survey test discussed above such as lack of motivation and a weak connection to the students’ actual course-taking patterns. Having received first-hand many of the complaints about high school testing such as “students are just drawing Christmas trees on the answer sheets,” I argue that, in many cases, states are making a rational decision to use the ACT or SAT as the achievement indicator. While many state policy makers do not understand or do not care about the alignment concerns, they are happy to reduce some testing and provide a visible benefit to many of their constituents. Further, there is some evidence and many anecdotes supporting “diamond in the rough” stories where a few students with poor grades and lack of school performance score surprisingly well on the college

entrance exams. School and district leaders are more aware of the alignment issues, but if they are going to be held accountable for the achievement of their students, they would rather have the 50-70% of students considering attending college take the test seriously than not. As someone who has spent much of his professional life worried about unintended negative consequences associated with various testing and accountability policies, I am surprised to find myself supportive of the use of ACT or SAT as the high school achievement test IF it is replacing a single survey test because it will be no worse in terms of match to the specific courses students are taking than an 11th grade survey test and it will likely improve the motivational concerns. I have a serious concern about the lack of available test accommodations for students with disabilities and English learners on the ACT and SAT compared to most state assessments. However, there are some signs that things are improving on this front.

NOW WHAT?

State leaders and their technical advisors find themselves in these seemingly no-win situations because of accountability policies that appear to consider high schools as big elementary schools with older students. Yes, the requirement to test only once in high school is some recognition of the difference, but barely. So what would I advise state leaders when it comes to high school testing? I suggest two potential directions. The first works within the existing status quo, while the other is a break from current practice.

Working Within the Existing System

- **Grades 9 and/or 10 survey tests to accompany the SAT or ACT**

Many states have decided to administer grade-level tests in grades 9 and 10 tied to the state's ELA and mathematics standards. In fact, several states that have adopted this approach are also administering the ACT or SAT in grade 11. This affords the state several opportunities. The state can measure student learning of the state's own standards in these two grades and it can use these test results as the achievement indicator for high school accountability, while limiting the SAT or ACT to its validated use as a college readiness indicator. If focused on the content standards that most students were expected to learn in grades 9 and 10, the assessments will have a better chance of providing a common measure (to serve equity purposes) of what students should have learned compared to waiting until 11th grade when the differences will be exacerbated. Employing grades 9 and 10 tests also provides the opportunity to compute student longitudinal growth from middle school through grade 11. On the other hand, grade 9 and 10 survey tests suffer from some of the same challenges as a single survey test in 11th grade in that students participating in either or both the grades 9 and 10 assessments may be in very different courses, leading to motivation and interpretation challenges with these tests. This may be more of an issue in mathematics where "tracking" is quite common but it is perhaps less of an issue in early high school ELA where students often take the same core classes before moving to electives.

- **End-of-Course Tests**

First, end-of-course testing offers considerable potential for supporting valid inferences about student achievement. However, such tests must be rigorous, high-quality, and incorporate the types of items and tasks that we would like to see used in instruction (because they will be mimicked in instruction). While not necessarily endorsing AP and IB, these exams generally meet this vision, but they are not cheap! Therefore, if states are going to go down the EOC road, they should limit the courses tested to those that they can do very well because of the likely direct effect on instruction. The compromise suggested above with using grades 9 and/or 10 survey tests to supplement the use of the ACT/SAT can also apply to an EOC model. The state can identify a very limited number (e.g., one in each content area) of key classes that almost all high school (not middle school) students complete to augment the use of the college readiness test as the achievement indicator.

Bending the Rules

Many of us have been touting the potential of balanced assessment systems in order to serve multiple users and multiple purposes of high school testing. It is impossible to do this from the state capitol alone. While I understand the equity imperative associated with school

High school accountability should be focused at the student level in ways that foster meaningful personalization once some limited set of knowledge and skills is secure.

accountability systems, high school accountability should be focused at the student level in ways that foster meaningful personalization once some limited set of knowledge and skills is secure. Such an approach could include a 10th grade survey test or a very limited set of EOC exams and could even include the use of a college readiness assessment as well. But to move toward more meaningful and deeper learning opportunities, students should be expected to engage in opportunities such as senior exhibitions, pursue internships or other extended learning opportunities, and/or complete rich performance tasks characteristic of New Hampshire's Performance Assessment of

Competency Education initiative and Wyoming's former Body of Evidence System. Of course, there is nothing stopping states and districts from pursuing such approaches now, but if such activities are not included in the ways in which high schools are held accountable, it is easy for such ambitious efforts to fall by the wayside.

No matter which approach that a state pursues, it should not do so to simply meet the minimal legal requirements imposed by a state or federal law. States should be clear about their goals, especially related to equity and excellence aspirations and related questions regarding commonality and specialization. State leaders should then outline a theory of action that very specifically describes how the proposed instructional and assessment system will best help to realize the intended outcomes. But that's not enough. Almost all policy initiatives suffer from unintended negative consequences. As test designers and policy leaders, we need to try to root out as many of these unintended negative consequences as possible and design our high school assessment system accordingly.



National Center for the Improvement
of Educational Assessment¹
Dover, New Hampshire

www.nciea.org