CCSSO
Council of Chief State School Officers

State End-of-Course Testing Programs:

A Policy Brief

Produced by:
Christopher Domaleski (NCIEA)
Technical Issues in Large-Scale Assessment (TILSA)
State Collaborative on Assessment and Student Standards (SCASS)

February 2011

THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

The Council of Chief State School Officers (CCSSO) is a nonpartisan, nationwide, nonprofit organization of public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, and five U.S. extra-state jurisdictions. CCSSO provides leadership, advocacy, and technical assistance on major educational issues. The Council seeks member consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress, and the public.

STATE END-OF-COURSE TESTING PROGRAMS: A Policy Brief

A paper commissioned by the
Technical Issues in Large-Scale Assessment Collaborative
Council of Chief State School Officers

Prepared By:
Christopher Domaleski, National Center for the Improvement of Educational Assessment

Supported By TILSA Advisors:
Duncan MacQuarrie, Douglas Rindone, and Charlene Tucker

Council of Chief State School Officers
One Massachusetts Avenue, NW, Suite 700
Washington, DC 20001-1431
Phone (202) 336-7000
Fax (202) 408-8072
www.ccsso.org

State End of Course Testing Programs: A Policy Brief

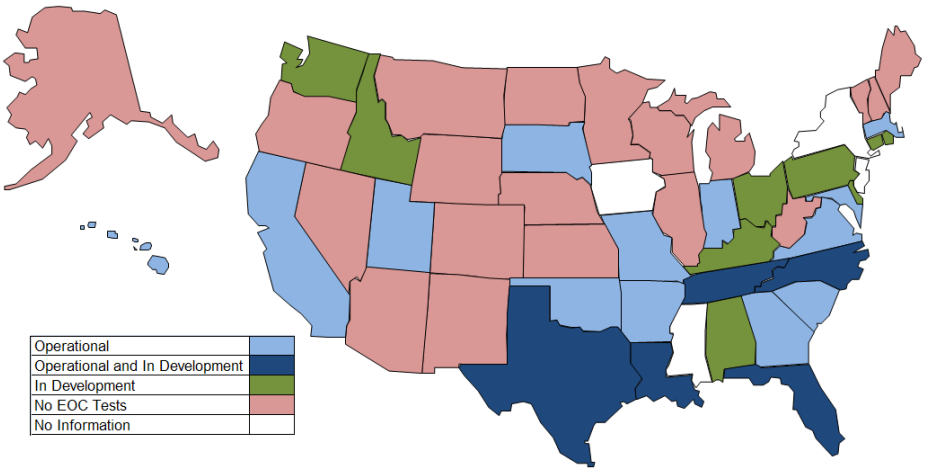
Prepared for the Technical Issues in Large Scale Assessment (TILSA)
State Collaborative on Assessment and Student Standards (SCASS)
of the Council of Chief State School Officers (CCSSO)

Chris Domaleski
Center for Assessment

Introduction

In recent years the number of states that have adopted or plan to implement end of course (EOC) tests as part of their high school assessment program has grown rapidly. As recently as 2002, only two states reported using EOC tests as part of the state assessment system[1]. Today, that number has increased to 19 with another 9 developing EOC tests for future implementation. Additionally, 5 states currently implementing EOC tests are developing new assessments.[2] This is depicted in figure 1. Clearly, state education leaders view EOC tests as a promising direction for high school assessment.

Figure 1. State Implementation of End of Course Tests



For the purposes of this document, EOC tests refer to state required, standardized exams administered at or near the completion of a term of instruction. The appeal of this approach is likely related to several factors. Perhaps foremost is the view that an assessment explicitly tied to a specific course and administered very near completion of the term will improve the connection between standards and instruction. Such an approach may also permit the development of a focused assessment

[1] Center on Education Policy. (2008). *State High School Exit Exams: A Move Toward End-of-Course Exams*. Retrieved from: http://eric.ed.gov/PDFS/ED504468.pdf.
[2] Based on a survey conducted by CCSSO in August 2010 in which 47 states and the District of Columbia responded.

that provides a more reliable and valid measure of student achievement with respect to the key knowledge and skills associated with each course.

While EOC tests certainly offer great promise, they are not without challenges. Many of the proposed uses of EOC tests open new and often complex issues related to design and implementation. The purpose of this brief is to support education leaders and policy makers in making appropriate technical and operational decisions to maximize the benefit of EOC tests and address the challenges.

Theory of Action

In considering the most fundamental questions about an EOC-based assessment program, including whether a state should develop EOC tests and for what purpose, it is important to begin with a clear idea of the end. One should consider the specific educational problems that are most important to solve and how EOC-based assessments might uniquely address these issues. This will help clarify important decisions about design and implementation.

A useful approach for accomplishing this is to construct a credible theory of action (TOA). In brief, a TOA explicates how the elements of a system work together to accomplish one or more desired outcomes.[3] Just as a good carpenter develops plans before beginning a building project, the TOA acts as a blueprint to show how the elements are intended to come together to reach the desired result.

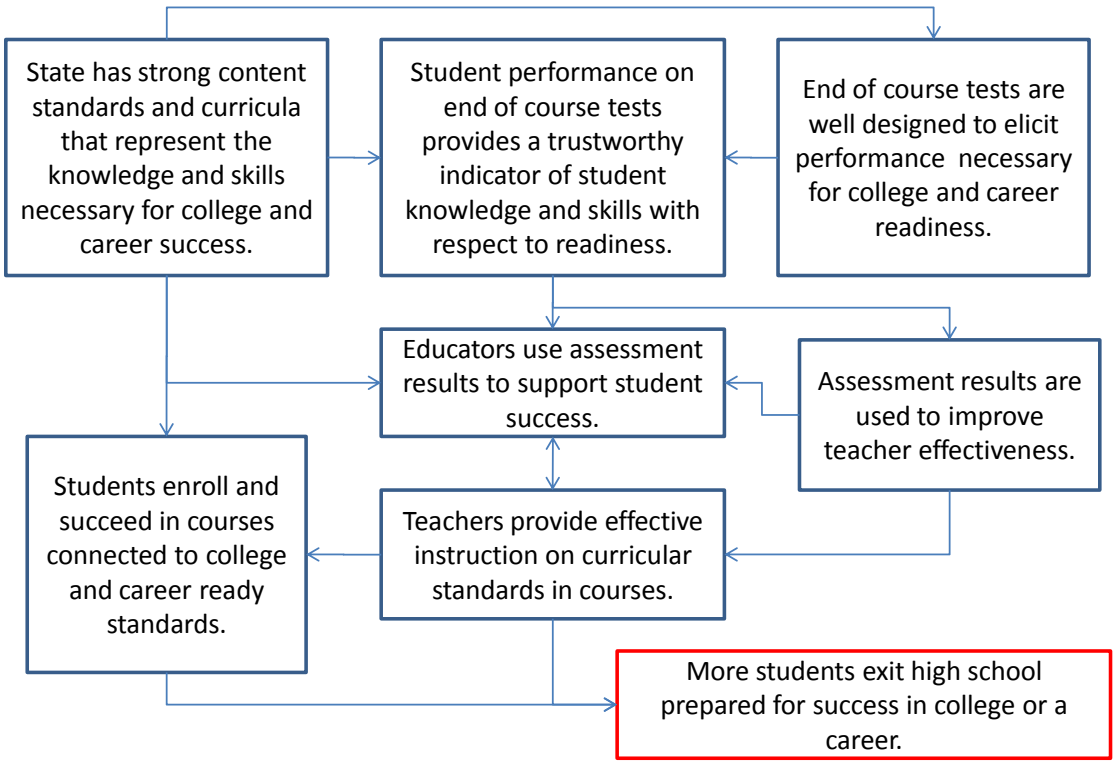
A **Theory of Action** (TOA) helps explain how the elements of a system work together to accomplish one or more desired outcomes. Developing a strong TOA can help guide key policy decisions.

Often, it is useful to depict the TOA in a diagram, taking care to identify how the intended outcomes are related to and supported by additional mediating conditions and outcomes. Not only will the TOA process help provide direction for design, but it can also serve as the basis to evaluate the extent to which the goals were achieved. A good TOA essentially works as a framework to construct and evaluate a validity argument for the assessment system.

A simplified example of a TOA is depicted in Figure 1. This TOA assumes that policy makers have prioritized the outcome of college and career readiness. Note that the theory is built on interrelated claims about the curriculum, instruction, and assessment. In particular, the model outlines specific functions of the EOC test to support college and career readiness. For example, test content helps signal what is important for teachers to teach and students to learn. Additionally, results are used to improve student success and teacher effectiveness. A more comprehensive theory would detail many more supporting claims about the necessary components to achieve each of the elements depicted. Ultimately, a clear understanding about how the assessment is hypothesized to support larger policy goals will aid in decisions about design and implementation of the program.

[3] See Perie, M. (2007). *Key elements for educational accountability models*. Washington, DC: Council of Chief State School Officers.

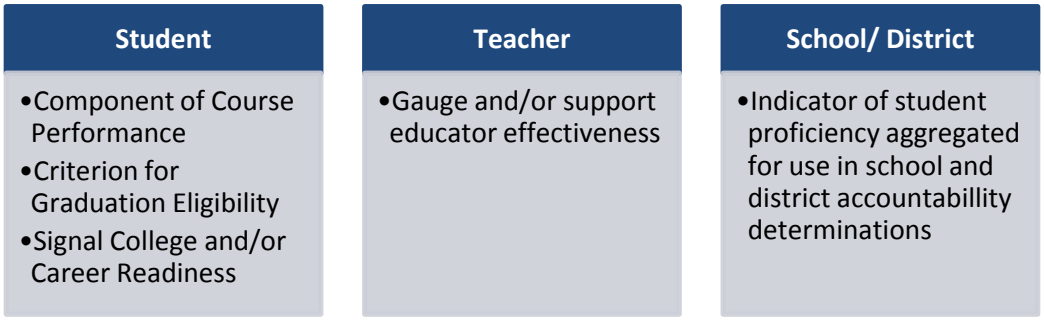
Figure 2. Sample Theory of Action for End of Course Test Program



Purpose and Uses

State policy makers increasingly rely on EOC tests to support a variety of purposes and uses. Prominent among these are accountability initiatives at the student, teacher, and/or school level. Each of these uses connects to a variety of critical issues related to design and implementation – in fact, these components serve as a useful organizational vehicle for the bulk of this document. Accordingly, the following section explores some of the common and/or emerging practices in each category along with the associated implementation considerations and challenges. A summary of uses addressed in this brief is provided in Figure 3.

Figure 3. Common Accountability Uses for State EOC Tests.



Student Accountability

Component of Course Performance

Some states use results from EOC tests as a factor in determining a student's course grade. In many cases, such policies are a response to concerns that course curricula and/or grading practices are highly variable throughout the state, including instances where the standards are not sufficiently covered or performance expectations are not appropriately rigorous. Incorporating the EOC test score into course performance bolsters confidence that earned credit represents student performance that meets state expectations. Additionally, this practice may increase student motivation for performance.

There are at least two alternatives for factoring results into course performance. First, attainment of a target score can be used as a condition to award course credit. For example, performance on the EOC test must be at the proficient level or higher in order for the student to pass or receive credit for the course. Such a policy may be independent of the teacher determined grade. That is, a student who receives an 'A' in the course but does not earn the requisite score on the assessment is not eligible for credit. Obviously, this is a rigid policy approach and ensures that the EOC score serves as a 'gateway' to course credit. Currently, 5 states have a policy that requires a passing EOC test score to earn course credit for at least some courses.

An alternative, which is used more commonly, is to combine the score on the EOC test with the student's course-based performance to arrive at a final grade or outcome. Such an approach typically establishes the EOC test as the final exam for the course. In this application, the student's evaluation on all course components except the EOC test is determined by the teacher. The state EOC test is then given either in place of or in addition to a final exam constructed by teacher. There are 9 states that report implementing this policy for some or all courses with EOC tests.

Determining Test Weight

There are a number of considerations associated with using EOC test scores as a component of course performance. Policy makers must decide how much 'weight' to assign the test and the amount of flexibility, if any, afforded when applying that weight. For example, a state may require that EOC test scores always count for at least 30%. A less stringent approach may establish a range of weights that the LEA, school, or teacher must apply (e.g. 15% to 30%). Even more flexibility is afforded when state

policy establishes the weight or range as a guideline rather than a requirement. Table 1 shows the test weights selected by the 10 states that report using EOC test scores as a component of course performance.

Table 1. Weights Used to Determine EOC Test Contribution to Course Grade

State	Course Grade Weight
Alabama	20%
Florida	30%
Georgia	15%
Louisiana	15% to 30%
North Carolina	25%
Missouri	10% to 20%
Pennsylvania	33%
South Carolina	20%
Tennessee	20%
Texas	15%

A state may also decide to combine the two alternatives. That is, it is possible to have a policy that requires students to achieve a passing score on the EOC in order to receive credit and use the score to influence course grades.

Naturally, to the degree that the state intends to influence course performance through EOC test scores, the weight should be increased and the flexibility to apply the weight should be removed. The chief advantage of this approach is that it creates a straightforward policy that carries a substantial and uniform impact across the state. In circumstances where there is considerable evidence that course expectations are not in line with state content or performance standards, this may be an attractive policy option. On the other hand, the advantage of choosing a lesser weight and/or more flexibility is that districts, schools, and/or teachers retain more control and ability to exercise professional judgment where appropriate. To this point, it should be recognized that using EOC test scores to influence course grades is typically associated with adding rigor to performance standards, but could, in fact, produce the opposite effect. For example, in a literature/ composition course that requires students to complete a substantial writing assignment, adding a heavily weighted selected response assessment could have the real or perceived effect of weakening course expectations.

Computing a Grade

Whenever EOC test scores are used as a component of course grades, the issue of ***how*** to facilitate the calculation must also be resolved. Most teacher assignments are scored by either assigning letter grades or scores which typically range from 0 to 100. On the other hand, performance on large scale standardized tests is typically reported as performance levels (e.g. basic, proficient, advanced) and a scale score which usually falls outside of the 0-100 range. A requirement to use test scores as a factor

in course grades without clearly specified procedures for how to accomplish this could negate the intended effect if the process is interpreted and applied differently.

One strategy to facilitate computations is to create an additional scale for the specific purpose of incorporating test performance in course grades. One state that uses this approach is Georgia, which produces a separate Grade Conversion Score (GCS) in addition to the primary reported scale score. This scale is produced by a linear transformation of the primary scale score values within each of three performance levels to three correspondent GCS ranges on a 0-100 scale. Scores below 70 are those that do not meet expectations; scores from 70 to 89 are those that meet expectations; and scores of 90 or greater are associated with performance that exceeds expectations. The advantage of this approach is that it produces a scale that is familiar to educators. Also, the ranges correspond to a desirable policy perspective (i.e. scores that meet expectations are associated with a score range that typically yields a 'C' or better). The drawback of this approach is that the GCS compresses the performance range which makes it a coarser measure of performance. However, the state addresses this by retaining the primary scale score to more fully describe the range of performance.

Timely Results

Still another issue associated with using EOC test scores as a component of course performance is the need to provide results in a timely fashion. When test performance impacts course grades or credit, a lengthy scoring process can not only delay production of report cards and transcripts, but can lead to consequences such as holding-up determinations of graduation eligibility. There are two primary elements that influence availability of scores: the timing of test windows and the efficiency of scoring. Test windows refer to the schedule of dates that are available for schools to administer the EOC test. Generally, policy makers experience tension between the advantages of security and efficiency associated with a tight test window that is consistent for all schools and the need for flexibility that multiple and/or longer windows afford. This flexibility can be particularly important given that few states have a single academic calendar and there are multiple schedules used by high schools that may differ by course (e.g. traditional, block, half/full credit terms, credit recovery periods etc.) Moreover, when there is an extended time period between administering the test and receiving scores, schools and/or school districts are compelled to test earlier in the term. This, of course, impacts the opportunity to learn and may call into question the validity of results.

A survey of state practices reveals that that nearly every state that implements an EOC test has established a test window for each term of instruction (e.g. fall, spring, summer). For most states, this window ranges from about 2 to 6 weeks. A few states have tighter windows – for example one state administers all state tests in two days. Conversely, a few states have flexible administrations over longer periods of time. The longest state window appears to be just over 3 months each term.

The second element to consider is the efficiency of scoring. It generally holds that assessment results can either be of high quality, provided quickly, or produced cheaply. In the best case, perhaps two of these conditions can be met — but almost certainly not all three. Therefore, policy makers must determine which to prioritize and how to address the consequences of this decision. For example, if swift results are the prime consideration, it is likely that at least some of the quality checks that typically

accompany large-scale, high-stakes assessments will need to be relaxed. Moreover, it may require the dedication of additional resources – in terms of personnel, funds, or both – to accommodate rapid reporting. In most cases, states will attempt to balance these competing considerations to obtain a satisfactory, if not optimal, solution for each. Table A3 in the appendix of this document shows the turn-around time for providing student level results to the school or district once they are received for scoring.

An additional factor that influences reporting is the number and type of constructed response items on the assessment. To the extent that the EOC exam contains items that require review by one or more trained raters, the time and cost of scoring will typically increase. To address this, some states have explored local scoring and/or distributed scoring for all or part of the assessment. Local scoring typically involves having teachers or other trained raters evaluate student responses either at the school or district location. Distributed scoring refers to the practice of scoring student work remotely, typically via electronic submission. Another promising approach may involve the use of artificial intelligence (AI) applications as the primary or secondary scoring method for some student responses, such as writing samples. It should be acknowledged that these strategies are not widely used in large-scale, high-stakes state testing programs. Therefore, policy makers are urged to carefully research the opportunities and challenges associated with each. For example, there are multiple approaches to AI scoring and states will want to fully understand the advantages and limitations of each – to include vetting the alternatives with an appropriate technical advisory group – before deciding whether to adopt a method, which one to select, and/or how best to implement it.

Graduation Eligibility

Another student level use for EOC tests is as a criterion for graduation eligibility – a practice currently in place in 13 states. Such policies are usually inspired by the desire to assure that a high school diploma is a meaningful indicator of requisite student achievement. This may also help establish consistency in the curriculum across the state and increase student motivation to meet expectations.

Combining Multiple Measures

Historically, states with similar policy objectives relied on a single cumulative assessment, perhaps made-up of content area subtests, to serve as an 'exit exam.' With EOC tests, it is important to consider how multiple measures obtained variably throughout the student's high school experience can be combined to render a decision of 'good enough' performance to qualify for a diploma. That is, what decision rules will be used to determine if the student has met the graduation standard?

There are at least four approaches to combining multiple indicators to yield a single outcome: *compensatory*, *conjunctive*, *disjunctive,* and *profile* methods. Compensatory means that higher performance in one measure may offset or compensate for lower performance on another measure. Conjunctive means that

How can multiple indicators be combined into a single outcome?

- **Compensatory:** higher performance on one measure may offset lower performance on another measure.
- **Conjunctive:** performance must be acceptable for every measure.
- **Disjunctive:** performance must be acceptable on at least one measure
- **Profile:** identify conditional rules to identify the patterns or profiles that are assigned certain values.

acceptable performance must be achieved for every measure. Disjunctive means that performance must be acceptable on at least one measure. A profile refers to a defined pattern of performance that is judged to be satisfactory, unsatisfactory, or equivalent.

A compensatory approach recognizes that some degree of variability in performance across indicators may be expected. Such an approach has a higher degree of reliability because the overall decision is based on multiple indicators evaluated more holistically. Moreover, reliability improves because random error in multiple measures tends to cancel. Conjunctive decisions are less reliable because errors accumulate across multiple judgments meaning a student might fail due to the least reliable measure. However, such an approach may be desirable when it is important to assure that a student does not fall below established standards on any one criterion. A disjunctive method is desirable when any one assessment is viewed as adequate assurance the student has met graduation standards. This is uncommon across content areas, but may be an appealing alternative within content areas, especially when assessments are judged to classify attainment of graduation standards equally well. Finally, profiles are useful especially when there are certain patterns that can be described that reflect valued performance that are not easily captured, usually because the combinations of criteria are judged to be not equivalent. For example, assume a state has algebra I, algebra II, and geometry EOC tests. It may be determined that algebra II is most comprehensive and rigorous and that passing this test can compensate for failing to meet the standard on the other two. However, if a student does not pass algebra II the only other acceptable profile is to pass both algebra I and geometry.

One example of a fully compensatory model is implemented in Maryland, where students are required to take and pass Maryland High School Assessments (HSA) following completion of coursework in four content areas in order to be eligible for graduation. Through what is termed the 'combined score option' Maryland students can meet the graduation requirement by obtaining an overall score for all four tests that is equal to the sum of the passing score for each test. Specifically, the courses and passing scores are: algebra 412, biology 400, government 394, and English 396. Students meet the graduation requirement by earning a combined score of 1602 or higher. This allows for stronger performance in one area to offset weaker performance in another.

A state approach that combines conjunctive and disjunctive elements is used in Louisiana. Beginning in 2010-11 all incoming freshmen in Louisiana must pass three EOC tests in the following categories to earn a standard diploma:

- English II or English III
- algebra I or geometry
- biology or American history

The policy is conjunctive in that students must meet the requirement in **each** of the three categories, which assures that all students with a standard diploma have met performance expectations in English, math, and either science or history. However, the policy is disjunctive in that students can meet the standard on either **one** of the two assessments within each category. By so doing, the state has established a uniform exit standard, but offers some flexibility in meeting that standard.

Transfer Students and Reciprocity

When graduation requirements are based on multiple measures at various points in high school, it is also important to consider how to handle transfer students. In particular, states should consider how students who are new to state public schools (e.g. those from out-of-state, private schools, or home schools) will meet accountability requirements. Such students may have taken and received credit for courses in another school for which the state has an associated EOC that is used to determine credit, determine graduation eligibility, or both.

One approach is to specify that transfer students must meet the state's accountability requirements with respect to EOC tests. This involves requiring transfer students to retake courses with associated EOC tests, take and pass the state EOC tests, or both. This offers maximum assurance that no student can be a candidate for graduation without meeting state standards. However, this practice will be most onerous for students who seek to transfer in a large number of credits. In fact, a student transferring in his or her senior year may be required to take several courses and/or exams, which is burdensome to the student and non-trivial for the school.

What are alternatives for handling transfer students?

- **Maintain requirements:** state accountability policies apply to transfer students
- **Full reciprocity:** accept transfer credit from sending institution
- **Conditional reciprocity:** accept transfer credit if other requirements are met

A more flexible approach is to offer reciprocity – with or without conditions. Conditional reciprocity may involve accepting transfer course credit if the student has taken and passed another qualifying assessment, such as the sending state's EOC test or an Advanced Placement exam. Full reciprocity without conditions refers to accepting credits without imposing additional requirements. Reciprocity allows local school districts flexibility to more efficiently manage transfer student admissions. However, depending on the conditions, it sets up the possibility that some students who have not met the state standards will be eligible for graduation. More troubling, some students may intentionally seek to circumvent state requirements by earning credit outside the high school that does not measure up to state standards.

States vary with respect to practices for transfer students and EOC tests. Many states either have not explicitly addressed this in state policy or empower the local school district to make determinations about awarding credit for courses with EOC tests.

Massachusetts provides an example of state that requires incoming transfer students to take required tests, but allows an appeal for some circumstances. State policy stipulates that students who transferred to a Massachusetts high school must participate in all tests available to them. However, if the student transferred late in his or her senior year and did not have the opportunity to participate in state testing, a 'transcript appeal' may be filed to evaluate the student's eligibility for graduation.

Virginia is an example of a state that offers reciprocity. If the student has passed an EOC test in another state, then the passing score on that state's test may be used to award what Virginia terms the

'verified credit' needed for graduation. On the other hand, if the student has passed the class outside the state public school system but has not taken an EOC test, the student may be administered the state EOC test to earn verified credit.

College and Career Readiness

State education leaders are increasingly viewing the primary focus of high school as preparing students for college and/or careers. In fact, the United States Department of Education's blueprint for reauthorization of the Elementary and Secondary Education Act (ESEA) specifically identified an emphasis on assessments that are trustworthy measures of progress toward or attainment of college and career readiness[4]. While a full treatment of this topic is beyond the scope of this paper, it is useful to identify the primary considerations to support use of EOC tests to signal readiness.

How can EOC test scores signal readiness for college or career?

- Establish clear definition of readiness supported by state content standards.
- Ensure these content standards are well-represented on EOC tests.
- Establish meaningful achievement standards linked to readiness.
- Evaluate performance and outcomes.

Developing meaningful measures of college and career readiness (CCR) has implications that go well beyond establishing a performance level on the EOC tests. First, the state must establish a clear and coherent definition of what it means to be CCR. In the best case, this definition is informed by expertise and research that reaches outside the K-12 system. Then, expectations of what CCR students should know and be able to do should inform the development of content standards and, subsequently, the test items and forms that assess these skills. If CCR performance calls for students to demonstrate higher order thinking skills, a test that requires only low level tasks, such as identification and recall, will not suffice. Moreover, states may have to look beyond tests that contain only selected response items to include item types that better capture more complex skills.

Prior to establishing standards, CCR expectations should be expressed in concise policy definitions that indicate "good enough" performance at each performance level. Such statements should clarify the target and allow policy makers to indicate what the desired achievement looks like (e.g. ready for post-secondary coursework). Next, the specific knowledge and skills correspondent with each level should be expressed in detailed performance level descriptors that will be used to guide the standard setting process.

There is certainly more than one approach to setting standards and no single method is "right." The selected design should take into consideration the features of the assessment, type and availability of performance data prior to standard setting, proposed use of results and many other factors.

[4] United States Department of Education (2010). *A Blueprint for Reform: The Reauthorization of the Elementary and Secondary Education Act.* Retrieved from: http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf

Common and appropriate methods may be item based (e.g. Angoff or Bookmark), performance based (e.g. Contrasting Groups, Body of Work), or include elements of both approaches.[5]

Given the central importance of the performance standards in signaling readiness, policy makers may also consider the following elements in the standard setting process.

- Incorporate broad based input early and often. Experts and stakeholders from a variety of groups should have an opportunity to participate in the process and contribute to the decision making.
- Consider use of external data to inform and evaluate the proposed achievement standards. For example, projected impact can be generated and compared to other benchmarks of readiness such as ACT or SAT scores.
- Include a plan for continued monitoring and review of the standards after they are established. Such a plan should include ongoing validity studies and allow for policy makers to act on findings. For example, a study of the relationship between EOC test performance and success in college would provide critical information about the achievement standards.

Currently, of the 28 states that have or are developing an EOC test, 6 states have established that the purpose of the tests is to signal readiness for college and/or career. Many other states may be considering, but have not yet determined, whether or how to link EOC tests to readiness. However, less is known regarding how states intend to use these indicators. This may involve producing summary measures to track readiness (e.g. percent CCR by school) and providing incentives to meet readiness targets. In other cases, states may develop articulation agreements with institutions of higher education that apply to students meeting CCR standards. Given the prominence of CCR in the blueprint for ESEA reauthorization and the priority given to readiness in the federally funded common core assessment initiatives, this is an area that will likely develop rapidly.

Teacher Accountability

In some instances states are considering using test scores as a component to determine teacher effectiveness. EOC tests are considered particularly appealing for this purpose, given the ability to associate test scores with a particular teacher over a term of instruction. It should be recognized that such use is far from straightforward and requires carefully building a system and process to address numerous challenges, many of which elude broad consensus over how or even whether they can be fully resolved. Notwithstanding, the essential elements that should be in place and the challenges to consider are presented in this section.

It is generally acknowledged that any use of test data to inform teacher effectiveness should control for prior performance.[6] Therefore, the assessment system must produce a measure that reflects the progress or growth of the student during the period of time the teacher provided instruction.

---

[5] See: Zieky, M.J., Perie, M., and Livingston, S.A. (2008). *Cutscores: A manual for setting standards of performance on educational and occupational tests*. Educational Testing Service.

[6] Some models may include controls for factors beyond prior performance, such as student characteristics.

Broadly, there are two primary elements that must be in place to accomplish this goal: 1) availability of one or more prior scores and 2) application of a suitable analytic method.

In order to obtain a suitable prior score, it is possible to either use a previous test, such as a banked EOC test or end of grade test, or administer an additional assessment during the course. If a previous test is to be used, one must consider the sequence, coherence, and timing. Sequence refers to the order that the student encounters the EOC tests. Commonly, students are permitted to take courses at different grades and in a different order (e.g. one student takes algebra in grade 8 and geometry in grade 9; another student takes geometry in grade 9 and algebra in grade 10). It is not reasonable to assume that one prior score is as good as another or that growth between the two can be interpreted similarly. What's more, courses may not be coherently related, even when taken in sequence in the same content area. That is, two EOC tests (e.g. biology and chemistry) may not share the same construct such that one is meaningfully related to the other. Finally, it is important to consider the impact of timing. For example, it would be very challenging to evaluate any one teacher based on performance changes during the three year gap between a grade 8 test and a grade 11 test.

What are the key considerations related to using EOC test scores to inform evaluations of teacher effectiveness?

- **Account for prior performance:** requires a suitable prior measure and analytic technique to calculate growth
- **Test characteristics:** assessment should represent what teachers should be teaching and students should be learning
- **Addresses attribution**: data system must connect scores to the instructor; in many cases, only a small set of teachers/courses are included
- **Research and evaluation**: engage in systematic program of research and evaluation to quantify sources of error and address validity claims

An alternative is to administer additional assessments at the beginning of and/or at multiple times during the term of instruction – that is, a "pre-post" design within the course. This relies on the assumption that the pre-test is well suited for its intended use. It should be highly correlated with the outcome assessment and, to the extent it represents the construct of interest, claims that gains are associated with instruction are better supported. This method may better control for extraneous influences, such as the effect of student gains or losses between terms. However, if students are assessed more frequently there is less opportunity for instruction to impact performance between assessment events. This reduces variance in performance which, in turn, decreases reliability.

The second consideration for producing a meaningful growth score is the implementation of an appropriate analytic method. There are a variety of approaches to consider and a full treatment of this topic is beyond the scope of this paper. Common alternatives include gain scores based on developmental or vertical scales or using regression based approaches such as value-added models (VAM)[7] or Student Growth Percentiles (SGP)[8]. Each approach offers specific advantages and limitations

[7] See Braun, Chudowsky, & Koening, 2010 and McCaffrey et al. (2003).

[8] See Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Educational Assessment: Issues and Practices, 48* (4),pp. 42-51.

and policy makers are encouraged to carefully evaluate the options with an independent technical advisory group.

Finally, it is critical to address the characteristics of the EOC test. In order to incentivize the desired instructional practices, the assessment must represent that which is most important for teachers to teach and students to learn. That is, it should have sufficient breadth to cover the full range of content and sufficient depth to address these standards beyond a superficial level. Beyond the content represented on the assessment, the range of performance measures produced must be sufficient. If the assessment is to produce useful information about students' progress to inform educator effectiveness, it must have 'high ceiling' and a 'low floor.' If the range is not sufficiently broad, the assessment will not reliably detect gains between multiple assessments for students of high or low ability.

Even with a well-designed assessment system that produces a trustworthy measure of student progress, a number of challenges must be addressed in order to move to the next step of associating those results with teacher effectiveness. These challenges include the following:[9]

- *Limited involvement of grades and subjects*. Where assessment information is not available, results cannot be produced. For many high schools, it is conceivable that a relatively small proportion of teachers will be included. Therefore, states that prioritize evaluation of teacher effectiveness, may wish to develop tests for all courses within the scope of educator evaluation.
- *Assigning accountability.* It is critical to determine which teacher should be held accountable for a student's performance when students receive instruction over tested material from multiple teachers.
- *Extraneous factors.* It is not enough simply to link scores to educators, it is critical to address extraneous factors that threaten the interpretation that it was the teacher's behavior that led to the observed gains. These factors might include those that advantage performance, such as availability of home enrichment or factors that mitigate performance, such as a student who infrequently attends class or experiences a family crisis during the instructional term.

Education leaders and policy makers are encouraged to engage in systematic data collection and research to address these challenges. Such research should explore the extent to which the system functions as intended, acknowledges and quantifies sources of error, and promotes desired outcomes. In the end, to the extent that policy makers intend to use results for high stakes applications (e.g. merit pay, grounds for termination, etc.) the burden to demonstrate that the system is fair and accurate is increased. Moreover, as the stakes elevate, it is important to guard against unintended consequences.

---

[9] For a more complete list and description see: Domaleski, C. & Hill, R. (2010) Considerations for Using Assessment Data to Inform Determinations of Teacher Effectiveness. Retrieved from: http://www.nciea.org/papers-UsingAssessmentData4-29-10.pdf

School and District Accountability

Federal Accountability

No Child Left Behind (NCLB) requires states to measure student performance in reading/ language arts and mathematics annually in grades 3-8 and at least once in high school. In addition, science must be assessed at least once in grades 3-5, 6-9 and 10-12. Results from reading/ language arts and mathematics are used to determine if schools/ districts are meeting established performance targets to evaluate adequate yearly progress (AYP). As the number of states using EOC tests in high school have grown, there has been a corresponding interest in using the results from these assessments to satisfy federal requirements. To date, 16 states report using EOC test results in their NCLB accountability system, while many other states may be exploring this option.

Including Results in Accountability

One issue associated with using EOC tests in high school accountability systems is determining how results will be included. With a test that is administered at a single point in time for all students, such as an end of grade test or comprehensive high school exit exam, inclusion is relatively straightforward. Results can be incorporated from a single test event (e.g. spring 2010) and participation rate is calculated as the number of examinees divided by the number of enrolled students in the grade. However, as indicated earlier, students often encounter EOC tests at different points in their high school experience. Determining which tests to include, when to include them, and how to calculate participation rate becomes less than straightforward.

How can results be included in accountability systems?

- **Annual inclusion:** all scores are used in the year the test is administered.
- **Cohort inclusion:** scores are included at an identified point (e.g. grade 10) when all or most students in an identified cohort should have taken the assessment – prior administrations may be banked.

There are at least two general approaches that can be considered to address this issue – annual inclusion and cohort-based inclusion. Annual inclusion describes an approach in which all scores are used in the year the test is administered, regardless of grade. Typically, this would capture the first-time administrations of each EOC test administered in that year for inclusion in the accountability system. For example, if geometry results are factored into accountability determinations, all qualifying administrations of geometry in any of grades 9-12 would be used in accountability calculations for the school. [10] An advantage of this approach is that it more directly reflects the performance of a school for the academic year reported, because scores are not 'lagged' from previous years. However, a method to account for participation needs to be addressed. For instance, if it is possible for a student to graduate without taking geometry (i.e. students can earn mathematics credit with an alternative mathematics course) then the school will never be accountable for the performance of the students who do not take this course. To the extent that students are systematically excluded from determinations, the integrity of the accountability results will be in question.

[10] 'Qualifying' refers to any established requirements for inclusion, such as scores associated with students who have been enrolled for the full academic year.

A second approach is cohort-based inclusion. This method establishes a specified point to calculate determinations when all or most students in an identified cohort should have taken the assessment. This typically involves banked assessment data for students who take the test prior to the year in which accountability determinations are made. For example, if most students are expected to take geometry on or before grade 10, geometry performance for the group of students enrolled in grade 10 only is used as the basis for accountability. For any grade 10 students who took geometry earlier, their scores are retrieved from the bank and included in calculations. The chief advantage of this approach is that it creates a well defined cohort of students to use as the denominator to account for test participation. Additionally, depending on the grade selected for determinations, it could include retests, permitting schools to use a student's best performance among multiple administrations. A drawback of this approach is that accountability results are lagged while the scores are banked. Additionally, a policy for assigning scores to schools will need to be addressed when the student tests at one school but is enrolled at another when determinations are made. Finally, it is necessary to consider how to handle students who do not test before the point at which calculations are made.

A related accountability issue relevant to both approaches is how to handle results from students who tested prior to high school. The most common example is algebra, which many students take in grade 8 or earlier. Advocates for including these scores in high school determinations argue that these are typically high performing examinees and schools should get 'credit' for favorable results. Some opine that to do otherwise could incentivize unintended consequences. Others make the case that results should be included at the school where the student was instructed. Therefore, high schools should not be held accountable (favorable or not) for student performance that occurred outside the grades served. Ultimately, this is a policy decision that should be carefully considered.

There is no single best approach for addressing inclusion. Policy makers should consider how each approach would support the primary goals of accountability as well as the logistical constraints, such as course-taking patterns and student mobility, in determining the approach that best fits. Examples can be found of each approach from among the states that use EOC tests in accountability.

One state that uses the annual inclusion approach is Virginia. Results for the applicable EOC tests in mathematics and reading are incorporated into accountability determinations in the year that they are administered. Scores apply to the school at which the student tested, including tests taken prior to high school. For example, results for an 8th grade administration would be included at the middle school. Within Virginia's approach, participation is calculated based on the number of students enrolled in EOC courses. Additionally, inclusion of all students in the accountability system is supported by a policy that ties graduation requirements to courses with EOC tests.

North Carolina provides an example of a state that uses the cohort approach. In North Carolina EOC tests in algebra I, English I and writing contribute to school accountability determinations. These determinations are based on the cohort enrolled in grade 10 each year. Scores for students who test earlier are banked and included when the student's cohort reaches grade 10. This enables North Carolina to use all enrolled students in the grade as both the denominator for calculating participation and the basis for attributing student performance to schools.

Comparability

Another issue that states may need to address when using EOC test results for school accountability is comparability between or among multiple tests. That is, state policy may permit flexibility in the timing and/or choice of courses that a student may take to satisfy curricular requirements. For example, one student may meet mathematics requirements with a geometry course and another with an algebra II course. If the state determines that performance on an EOC test from either course will be used for federal accountability, it is necessary to demonstrate that tests are comparable.

Establishing comparability between or among different tests is non-trivial. This will require presenting evidence that the tests are constructed to the same high quality technical standards and yield equally valid and reliable results. Moreover, the state must demonstrate that the tests have comparable level of difficulty and provide consistent determinations of student performance.

Cross Cutting Issues

Curricular Coherence

For a state that is considering the adoption or development of EOC tests, it is first necessary to determine that expectations for student learning are sufficiently well-defined and consistent throughout the state. Although there is some variability in terminology, generally this is accomplished through the creation of a uniform curriculum and/or framework for each course that defines how state standards will be addressed in instruction. In the absence of such clarity, it is not safe to assume that the courses for which EOC tests are intended exist in all schools or that students receive similar instruction in these courses. The creation of standard course curricula is critical to ensure that students will have an opportunity to learn and that assessment results will be valid for the intended use(s).

Many states also find that addressing variability in course nomenclature and patterns will help successfully implement an EOC testing program. For example, the standards intended to be covered on an algebra EOC test may be taught in courses termed algebra I, concepts of algebra, applied algebra, etc. The state should explicitly identify which courses (e.g. by name and/or course number) that will administer an EOC test. Additionally, it is important to address how to handle situations where students may receive partial instruction of the standards across multiple courses and when these students should test.

Assessment Design

Developers have numerous options when determining the format and design of EOC tests. As stated from the outset, these decisions should be guided by the overarching purpose of the assessment program and the intended uses of the results. In most cases, the primary purpose of the EOC tests will be to provide summative information, such as a pass/fail classification, with respect to student performance on the standards associated with the course. Policy makers may also wish for these tests to serve other purposes, such as render diagnostic information about student achievement (i.e. How is the student progressing on learning goals to guide instruction?). However, an assessment cannot serve multiple purposes well. As a general principle, as the intended uses of a single test increase, the ability

of that test to satisfy each intended use decreases. Therefore, policy makers are encouraged to consider how an EOC test may be situated within a broader formative, interim, summative assessment system to support multiple goals.[11]

In contrast with a cumulative high school exam that covers multiple courses, EOC tests are generally intended to focus more narrowly on specified content standards. By so doing, designers are able to produce a measure that goes beyond a superficial level and elicits knowledge and skills associated with a deeper understanding of the curriculum. To accomplish this requires engaging in a careful process to define the construct of interest and determine the format and design that 'fits.'

While there is certainly more than one approach to guide the process of assessment design and development, the *evidence centered design* (ECD) is widely used.[12] As described by Mislevy, Steinberg, and Almond the ECD model consists of three components:

- **Student model**: define the knowledge and skills that should be tested (i.e., the construct)
- **Evidence model**: determine the performance or evidence that represents the construct
- **Task model**: identify the tasks or items that elicit evidence of performance

Using ECD principles, developers engage in a process to determine what is most important to measure and how it is best measured. This is typically codified in test specifications and test blueprints that clearly define the parameters of the assessment.

Though certainly not exhaustive, there are three general categories of item types that may appear on EOC tests: selected response, short constructed response, and extended constructed response. Selected response, or multiple choice, items are the most widely used, as they are efficient and inexpensive to develop and score. Because examinees can typically work through them relatively quickly, they provide a means to cover a large number of standards. However, constructed response items are better suited to elicit more complex or higher order knowledge and skills. Short constructed response items typically require the examinee to independently produce an answer sometimes accompanied by supporting information, such as solving mathematical expressions and showing work. Extended constructed response items require more in-depth student work products, such as producing an essay to argue a position in response to a prompt.

Although more expensive and time consuming to produce and score, constructed response items serve the important purpose of drawing-out skills such as conceptual understanding that may be a critical element of student performance. As indicated in table A3 located in the appendix to this document, there are 10 states that report using constructed response items on some or all EOC tests. Ultimately, policy makers are encouraged to consider the priorities of the assessment program and

---

[11] See Marianne Perie, Scott Marion, Brian Gong, and Judy Wurtzel (2007). *The Role of Interim Assessments in a Comprehensive Assessment System: A Policy Brief*. Achieve, Inc., The Aspen Institute, and The National Center for the Improvement of Educational Assessment, Inc.

[12] See Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (1999). Evidence Centered Assessment Design. Educational Testing Service. Available at: http://www.education.umd.edu/EDMS/mislevy/papers/ECD_overview.html

weigh the relative advantages and limitations of each approach to work through item format and broader design decisions.

Through-Course Assessment

Recently, interest has risen in what some believe is an innovative and promising approach for evaluating student performance — *through-course* testing. Through-course assessment is a special case of EOC testing and refers to an approach in which students encounter assessments at multiple points throughout the term, which are then combined into a single summative judgment. Just as a teacher might give quizzes or assignments following several units of instruction and combine these into a single course grade, proponents of through-course testing advocate that large scale summative testing does not have to be based on one event. A through-course approach essentially adds interim components to an existing EOC test. To be sure, there are some clear advantages associated with through-course testing. Not only does it allow the time between instruction and assessment to be minimized, but it also permits the assessment to more precisely focus on the content covered. Additionally, it may overcome a logistical obstacle to including constructed response items or performance tasks on assessments that measure higher-order and/or traditionally difficult to assess skills. By administering these items earlier in the course, more time is available to score them, reducing the delay between the conclusion of the course and the availability of student score reports.

Although conceptually appealing, this approach is not without complications and challenges. The most prominent issue to address is determining what skills will be assessed at what points and for what purpose. For example, if four tests are given throughout the course, will each test cover a non-overlapping set of standards? Alternatively, will each test cover the same set of standards but differ in terms of cognitive complexity or expectations for student performance? Will the tests be used for more than combining results into a summative judgment of performance, such as measuring growth throughout the course? Many other uses and designs are possible, each necessitating a distinct design approach.

Additionally, there is not well-established guidance for exactly how multiple indicators should be combined into a single score with through-course tests. For a skill that may develop over the course of an instructional term such as writing, should students be evaluated on their early work if they demonstrate higher quality work later? How should the components be weighted and what should be done about missing components, such as may occur when a student transfers-in during the term?

These and other issues are important to address from the outset when considering the suitability of adopting a through-course approach.

Scope of the Assessment Program

Another broad factor to consider in the development of an EOC assessment program is the scope of implementation. That is, for what courses should an EOC test be developed? Should all content areas (e.g. mathematics, language arts, science, social studies, others) be assessed and to what extent should multiple courses within a content area have an assessment? While these decisions will necessarily differ by state, Table A1 in the appendix of this document shows the courses currently

assessed by states that have adopted EOC tests. Ultimately, decisions about scope are policy considerations. To that end, it is useful to explore the primary factors that may shape implementation policy.

To start with, the availability of resources is, unavoidably, a prime consideration. To be sure, cost and staff capacity, have a significant impact on options that are feasible. Additionally, elements such as the format of the assessment, the frequency of administration, or the scope of ongoing development and support, may make some options more feasible than others.

Also, the structure of high school curriculum and course-taking patterns will influence the scope. In cases where students unvaryingly take a specified course, the decision may be relatively uncomplicated. However, as indicated earlier, frequently students may satisfy course completion requirements by taking one from among a choice of courses. In these instances, policy makers must decide if all choices should be tested or a subset of these courses. If the state tests a subset of courses, it is important to consider unintended consequences. For example, if physical science is assessed but not biology, will course-taking patterns change to avoid the tested content area? On the other hand, if student accountability requirements (e.g., graduation eligibility) are associated with one tested course, what will be the incentive for taking non-tested courses? One way to address this threat is to offer EOC tests in more than one, but not all courses, and require some number of credits to be earned from the tested courses. This allows some choice in course taking patterns while establishing a minimum testing requirement. However, if the tests are not comparable in rigor, this, too, may influence course decisions.

The proposed use of the assessment is yet another consideration that influences scope decisions. For example, if the system is intended to fulfill the strictures of NCLB accountability, the state must minimally test in reading/ language arts, mathematics, and science. Moreover, all students must participate in the tests. As described above, this may require multiple tests within a content area and it will necessitate alternate assessments, which is addressed more fully later in this document. If the tests are to serve as a graduation eligibility criterion or to signal college and career readiness, then policy makers must consider the essential components to support this use. For instance, is it reasonable to base a college-ready classification on the performance of an algebra assessment usually administered in the ninth grade? Or, does this proposed use necessitate expanding the scope to include higher level mathematics courses? As referenced earlier in this document, if policy makers desire to use EOC test performance for teacher evaluation, it will be necessary to assess a sufficient number of courses to include the educators intended to be evaluated. Based on the number of courses typically assessed, the overwhelming majority of teachers would be excluded.

Other components of the state assessment system have a bearing on decisions about scope. If the state retains a cumulative high school exam that can either replace or support the accountability functions of the EOC, the development scope may be reduced. In other instances, states look to commercially available assessments, such as AP or IB exams to serve some of the functions otherwise associated with EOC tests.

Transition

It is also important to consider transition issues whenever new development is contemplated, particularly when tests are used for high stakes purposes. At a minimum, this requires adequate notification and opportunity to prepare for transition. Because a primary function of EOC tests is to ensure some degree of standardization in course curricula, teachers and administrators must clearly understand the requirements well in advance and receive the necessary training and support to meet expectations. Second, students must have opportunity to learn prior to testing. This concern is augmented when courses/ tests are designed to build on knowledge and skills that should be acquired over multiple courses. For example, if the curriculum and assessment are revamped for both algebra I and algebra II, will students going directly into algebra II have an adequate opportunity to meet performance expectations? For this reason, it is common for high school curriculum and assessment changes to be 'phased-in' with a selected cohort, such that students do not encounter a combination of new and old expectations.

Standardization/ Flexibility

With any large-scale testing program, there is a tension between standardization and flexibility. To the extent the assessment is administered under comparable conditions and a well-specified scoring process is consistently applied, the results from various tests can be interpreted similarly. However, at times it is desirable to relax some of the rigidity to allow for special circumstances. As discussed earlier, this is particularly important for EOC tests where variable course-taking patterns and the need for quick turn-around of results calls for flexibility. For example, it may seem sensible to allow a school to administer the EOC test on different days within the same school to allow students on a block schedule to test on the day their class meets. However, unless multiple forms are available, this could lead to concerns regarding test security.

Table 2 presents some issues that test developers and policy makers often consider when developing the operational procedures for assessment programs. Many of these are pertinent to virtually any large-scale program, but all are particularly important for EOC tests.

Table 2. Assessment Focus on Standardization Versus Flexibility.

Issue	Focus on Standardization	Focus on Flexibility
Test windows	All schools test on the same day or within a short window, which promotes consistency and security.	There are no test windows (i.e. 'on-demand' administration) or lengthy test administration windows, which allows students to test immediately following instruction even when calendars or course taking patterns are variable.
Administration Mode	A single mode (e.g., paper or computer) may provide best case for comparability of results.	Multiple modes allow students and schools options for administering the test in the most efficient manner.
Administration Conditions/ Resources	Students have access to the same resources during testing, such as calculators with the same features, such that results can be interpreted similarly.	Allowing some variability in test resources, such as allowing students to use different calculators, may better match the assessment with the student's instructional experience.
Scoring	Centralized scoring of constructed response items by a group of raters receiving the same training and working in the same conditions maximizes security and consistency of results.	Local scoring will likely accelerate the time it takes to score items and may be a valuable professional development activity for educators.

Retests

Virtually any assessment that has consequences for students, such as influencing course outcome or graduation eligibility, should provide opportunities to retest. This is certainly the case for many applications of EOC testing. However, because these assessments are explicitly connected to a course, the issue of when and under what circumstances retesting should be either allowed or required is less than straightforward. Because there are two factors (the test and course) each with two possible outcomes (pass or fail), there are four conditions to consider when determining retest policy. These are presented in Table 3.

Table 3. Retest Alternatives.

Course Outcome	Test Outcome	Retest Requirements
Student passes course	Student passes test	Most likely a retest is not offered in this circumstance. However, policy should address requests to retest from students who wish to earn a higher score.
Student passes course	Student fails test	Important to offer retests if results are used for stakes unrelated to the course, such as graduation eligibility.
Student fails course	Student passes test	If results are a component of course grades, students will either test again when retaking the course or a procedure to 'bank' the score to apply to the future course attempt should be in place.
Student fails course	Student fails test	Most likely students will retake the course and the test in this circumstance. Exceptions may occur if the course is optional and the student does not reattempt and there are no other student stakes.

Policy should also address when and how students should retest. In some instances, it may be necessary to retake the course in order to retake the assessment. This is likely to be the case if course credit was not earned and the course is specifically required for graduation. However, for students who do not need or do not choose to retake the course, the requirements and/or conditions for retesting should be defined. These students would not be encountering the assessment at the completion of a course as designed; therefore some program of remediation may be appropriate to support student success on the subsequent attempt.

Assessing Students with Disabilities

The Individuals with Disabilities Education Act (IDEA) as revised in 1997 and 2004, specifically requires the participation of students with disabilities (SWD) in statewide assessments. These assessments must be appropriate for the population and aligned with state standards. This is further addressed in No Child Left Behind (NCLB) regulations and guidance, in which five alternatives for participation of SWD are outlined:

- Participation in the general grade level assessment
- Participation in the general grade level assessment with accommodations
- Participation in an assessment based on alternate achievement standards (AA-AAS) – for students with the most significant cognitive disabilities (1% cap on proficient scores if used for NCLB accountability)
- Participation in an assessment based on modified academic achievement standards (AA-MAS) – for the small group of SWD who are unlikely to achieve grade-level proficiency within the year (2% cap on proficient scores if used for NCLB accountability)
- Participation in an assessment based on grade level academic achievement standards (AA-GLAS) – an alternate assessment that covers the same grade level content and has the same performance expectations as the general assessment

It is important to consider how the intended purpose and uses of the assessment system interact with the participation options available. For example, if the state uses EOC tests as a criterion for graduation eligibility and offers an alternate assessment for students with disabilities, it must consider how performance on the alternate assessment is evaluated with respect to the graduation standard. Federal policy prohibits states from precluding students who take an alternate assessment from attempting to complete the requirements of a high school diploma. This does not compel the state to regard the scores as comparable, rather it necessitates that states clearly outline the qualifications for a diploma and allow all students the opportunity to meet those requirements.

Whatever participation alternatives are offered, the state must ensure that the assessment system is accessible and yields valid and reliable results for all examinees. EOC tests present both opportunities and challenges in this area. For example, the flexible nature of EOC tests may better allow for opportunity to learn, as students can prepare at a suitable pace and encounter the course content when they are ready (e.g., following one or more preparatory courses). However, states should recognize that a system of tests connected to distinct courses, may necessitate substantial reworking of inclusion and support strategies to ensure it 'fits' for all courses and tests. Stated another way, it is unlikely that one solution (e.g., a uniform approach to accommodations) will meet the needs of all EOC tests.

Evaluation

Finally, it is important to engage in an ongoing evaluation process to determine the degree to which the EOC assessment supports the state's goals. Such a plan should include, but go beyond, established criteria in the *Standards for Educational and Psychological Testing*[13] and, if applicable, the *NCLB Standards and Assessments Peer Review Guidance.*[14] In the best case, the plan should be developed from the outset and include a systematic process to evaluate the explicit claims in the theory of action. Often such plans will be developed in consultation with the state's technical advisory committee, which can help guide the state in collecting and evaluating the appropriate evidence.

Strong evaluations go beyond addressing the traditional psychometric properties of the assessment. They should be tailored to address the central purpose and uses of the system. For example, if the assessment is used to signal college and career readiness, student performance may be compared to other measures of readiness, such as the SAT or ACT. Additional evidence such as college-going rates and performance in credit bearing college courses further illuminate the extent to which claims of readiness may be supported.

Moreover, it is critical to go beyond the assessment and consider the *supporting* claims and conditions that promote the theory of action. For example, if the theory claims that educators will use

[13] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

[14] United States Department of Education (2007). *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001.* Retrieved from: http://www2.ed.gov/policy/elsec/guid/saaprguidance.pdf

assessment results to inform instruction and support student success toward college and career ready outcomes, it may be useful to evaluate some or all of the following supporting conditions:

- Educators receive training on interpretation and use of assessment results.
- Student performance information is used to tailor curriculum and instruction.
- Students who score below assessment targets receive appropriate remediation and support.
- Increased attention to rigorous academic standards creates increased engagement and motivation from students, educators, and parents to support student achievement.

Again, the nature of the evaluation will differ depending on goals and actions outlined in the theory of action. However, a comprehensive evaluation plan should include ongoing collection of a variety of qualitative and quantitative evidence. Finally, the state should regularly monitor the evidence and act on findings to support continual improvement.

Conclusion

As the number of EOC tests has increased along with an expansion of their role in accountability, development and implementation decisions are more important than ever. This process starts with articulating how the assessment fits into a credible theory of action that describes how all elements of the system will work together to promote the desired educational outcomes. Once clarified, policy makers are encouraged to carefully study the full range of policy, technical, and practical considerations associated with the intended purposes and uses of the assessment, many of which have been discussed in this document. Finally, an ongoing monitoring and evaluation plan should accompany implementation to support system improvement. By so doing, state leaders are best positioned to leverage the promise of EOC tests and mitigate unintended consequences.

Appendix – Results of Survey of State EOC Test Implementation

Table A1. Number of EOC Tests Operational and In Development by State.

	Mathematics		English, Literature/ Comp		Science		Social Studies		Total Operational	Total In Development
State	Operational	In Development	Operational	In Development	Operational	In Development	Operational	In Development		
AL	-	1	-	2	-	1	-	1	0	5
AR	3	-	-	-	1	-	-	-	4	0
CA	8	-	-	-	8	-	1	-	17	0
CT	-	1	-	-	-	-	-	-	0	1
DC	-	-	-	-	1	-	-	-	1	0
DE	-	2	-	1	-	1	-	1	0	5
FL	1	1	-	-	-	1	-	2	1	4
GA	2	-	2	-	2	-	2	-	8	0
HI	-	-	-	-	1	-	-	-	1	0
ID	-	1	-	-	-	3	-	-	0	4
IN	2	-	1	-	1	-	-	-	4	0
KY	-	1	-	1	-	1	-	1	0	4
LA	2	-	1	1	1	-	-	1	4	2
MA	-	-	-	-	4	-	-	-	4	0
MD	1	-	1	-	1	-	1	-	4	0
MO	3	-	2	-	1	-	2	-	8	0
NC	2	3	1	-	2	-	3	-	8	3
OH*	-	-	-	-	-	-	-	-	0	0
OK	3	-	2	-	1	-	1	-	7	0
PA	-	3	-	2	-	2	-	3	0	10
RI	-	-	-	-	-	4	-	4	0	8
SC	1	-	1	-	2	-	1	-	5	0
SD	3	-	-	-	3	-	4	-	10	0
TN	2	-	2	1	1	-	1	-	6	1
TX	3	-	1	2	3	-	2	1	9	3
UT	3	-	5	-	4	-	-	-	12	0
VA	3	-	2	-	3	-	2	-	10	0
WA	-	4	-	-	-	1	-	-	0	5

*** Legislative requirement to develop, but courses have not been determined.**

Table A2. Summary of Purposes and Uses for State End of Course Testing Programs

State	Required for Course Credit	Component of Course Grade	Course Grade Weight	Contributes to Graduation Eligibility	Indicates College/ Career Readiness	Used in NCLB Accountability
AL	No	Yes - all	20%	No	Yes - all	No
AR	Yes - some	No	-	Yes - some	Yes - some	-
CA	UND	No	-	No	No	Yes - some
CT	No	UND	-	UND	No	No
DC	No	No	-	No	No	Yes - all
DE	Yes - all	UND	-	UND	No	Yes - all
FL	Yes - all	Yes - some	30%	Yes - some	No	Yes - some
GA	No	Yes - all	15%	No	No	No
HI	No	No	-	No	No	No
ID	UND	No	-	UND	No	UND
IN	No	No	-	Yes - some	No	Yes - some
KY	UND	UND	-	UND	UND	UND
LA	No	Yes - all	15% to 30%	Yes - some	UND	Yes - some
MA	No	No	-	Yes - all	No	Yes - all
MD	No	No	-	Yes - all	No	Yes - some
MO	No	Yes	10%-20%	No	No	Yes - some
NC	No	Yes - all	25%	Yes - some	No	Yes - some
OH	UND	UND	-	UND	UND	UND
OK	No	No	-	Yes - some	Yes – some	Yes - some
PA	No	Yes - some	33%	No	UND	UND
RI	UND	UND	-	UND	UND	No
SC	No	Yes - all	20%	No	No	Yes - some
SD	Yes - all	No	-	Yes - all	No	No
TN	No	Yes - all	20%	Yes - some	Yes - some	Yes - some
TX	Yes - all	Yes - all	15%	Yes - all	Yes - some	UND
UT	No	No	-	No	No	Yes - some
VA	No	No	-	Yes - some	Yes - some	Yes - some
WA	No	No	-	Yes - all	No	Yes - some

"Yes - all" indicates that every state EOC test is used for the applicable purpose; "Yes - some" indicates that a subset of the state EOC tests are used; "UND" indicates state policy was undetermined as of the data collection

Table A3. Summary of State End of Course Test Program Characteristics

State	Established Test Windows	Includes Constructed Response	Turnaround Time for Reports	Local Scoring Permitted
AL	Yes	No	2-4 weeks	No
AR	-	-	-	-
CA	Yes	No	4-6 months	No
CT	UND	Yes - some	3-7 days	Yes - some
DC	Yes	Yes - all	2-4 months	No
DE	UND	UND	1 day or less	No
FL	Yes	No	1-2 weeks	No
GA	Yes	No	1-3 days	Yes – some[6]
HI	UND	No	1 day or less	No
ID	UND	UND	UND	UND
IN	Yes	Yes - all	3-7 days[1]	No
KY	Yes	UND	UND	UND
LA	Yes	Yes - all	1-3 days	No
MA	Yes	Yes - all	2-4 months	No
MD	Yes	No	1-2 weeks	No
MO	Yes	Yes - some	3-7 days	No
NC	Yes	No	3-7 days[2]	Yes - all
OH	UND	UND	UND	UND
OK	Yes	Yes - some	1-2 weeks	No
PA	UND	Yes - all	1-2 weeks	No
RI	Yes	UND	UND	UND
SC	Yes	No	1-3 days	No
SD	No	No	1 day or less	Yes - all
TN	Yes	No	3-7 days[3]	No
TX	Yes	Yes - some	1-2 months[4]	No
UT	Yes	No	1-3 days	No
VA	Yes	Yes - some	1 day or less[5]	No
WA	Yes	No	2-4 months	No

"Yes - all" indicates condition applies for each state EOC test; "Yes - some" indicates conditions applies for a subset of the state EOC tests; "UND" indicates state policy was undetermined as of the data collection

Notes

1: online 24hours, paper 7 days
2: because of scoring at LEA, turnaround times vary
3: time may vary depending on preliminary item analyses
4: response refers to English EOC test
5: does not include writing which takes approximately 8 weeks
6: not all LEAs participate in local scoring