



CRITERIA for PROCURING and EVALUATING HIGH-QUALITY ASSESSMENTS

States have demonstrated their leadership and commitment to ensuring the success of all students by adopting college- and career-readiness standards. To realize the potential of these standards, states require assessments that match the depth, breadth, and rigor of the standards; accurately measure student progress toward college and career readiness; and provide valid data to inform teaching and learning.

Assessment of College and Career Readiness. States have taken different approaches to establishing college- and career-readiness standards and to putting in place high-quality aligned assessments. Many states have adopted the Common Core State Standards (CCSS); some have modified the CCSS to meet their state’s context and needs; and others have developed standards independent of the CCSS. To provide assessments that are aligned to these standards, many states are working together through assessment consortia, while others are taking alternative paths for transition. This document is grounded in best practices for assessment development and in the research that defines college and career readiness for English Language Arts(ELA)/literacy and mathematics. Thus, regardless of each state’s approach, this document is intended to be a useful resource for any state procuring and/or evaluating assessments aligned to their college- and career-readiness standards.

Assessment Criteria for States to Consider. This document provides criteria for states to consider as they develop procurements and evaluate options for high-quality state summative assessments aligned to college- and career-readiness standards. The criteria build on the states’ [high-quality summative assessment principles](#) (CCSSO, 2013) which articulate their commitment to high-quality assessments aligned to college and career readiness. To assist states in operationalizing their commitment, this document pays particular attention to not only the criteria states could ask vendors to meet, but also to the evidence states could ask vendors to provide to demonstrate criteria have been – or will be – met. States will, of course, adapt these criteria to reflect their context, standards, and procurement regulations.

Contents of this Document. This document begins with an overview of the assessment criteria and continues with a chart containing detailed criteria and sample evidence. These criteria do not cover every area that a state would have to address in a procurement or evaluation process. Instead, they focus on the critical characteristics that should be met by high-quality assessments aligned to college- and career-readiness standards. A more comprehensive source for the development and validation of assessments is the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 1999). The assessment criteria and evidence discussed herein were developed by referencing the *Standards for Educational and Psychological Testing* and several other key sources listed in the bibliography. Additional state-specific criteria at the end of the document highlight a few of the most important additional issues that states may wish to consider in a procurement or evaluation process.

Notes about Evidence and Terminology. This document is intended to support states in selecting assessments that meet a high bar for quality. Thus, the document suggests the evidence that states will need to review in order to make informed judgments on vendors’ claims about the quality of their proposed assessments. Of course, vendors may propose assessments that are yet to be developed, assessments in development, and/or existing assessments. In designing procurement or evaluation procedures, states may therefore find it helpful to design the process for awarding “points” so as neither to reward existing (but poor quality) tests just because they have data available, nor to reward well-intentioned conceptual designs that are not executable. To support this goal, vendors should be asked to provide the most rigorous level of evidence they have available, consistent with the stage of assessment development they are in. The types of evidence that vendors should be expected to provide at different stages of development are described below:

- For assessments to be newly created, the most rigorous level of evidence will include the vendor’s descriptions of their established and proven processes; data from similar assessments; proposed test blueprints and other specifications (e.g., test design documents, test specifications, item specifications, scoring specifications); exemplar test items, passages, and forms; proposed studies, reports, and technical documentation to be created during assessment development and operation; and the processes for responding to such data. In addition, the vendor’s prior experience, expertise, and letters of recommendation should be included.
- For assessments that are currently in development, the most rigorous level of evidence will depend on the stage of assessment development. Evidence should include test blueprints and other specifications (e.g., test design documents, test specifications, item specifications, scoring specifications), and exemplar test items, passages, and forms. In addition, evidence should include as much of the data described below regarding pre-existing assessments as is available. Where such evidence is not available, vendors should provide descriptions of their established and proven processes; data from similar assessments, proposed studies, reports, and technical documentation to be created during assessment development and operation; and the process for responding to such data. In addition, the vendor’s prior experience, expertise, and letters of recommendation should be included.
- For pre-existing assessments, the most rigorous level of evidence will include comprehensive validity evidence; test blueprints and other specifications (e.g., test design documents, test specifications, item specifications, scoring specifications); annual technical reports; results of studies on scaling, equating, and reporting; and exemplar test items, passages, and forms.

Additionally, regardless of the stage of test development, states may find it helpful to put in place best practice quality assurance and other processes so that states can monitor quality throughout development and administration, and periodically evaluate evidence to ensure criteria are being met.

Finally, a note about terminology. In this document, the term “assessments” generally refers to the entire suite of summative assessments a state would procure – that is, tests of ELA/literacy and mathematics in each grade assessed. In sections specifically about ELA/literacy or mathematics, however, the term refers to the set of summative assessments in that content area. The terms “assessment” and “test” are often used interchangeably when discussing a single grade level/content area. Throughout the document, the term “tasks” refers to extended-response, open-ended test items; “test items” refers to the stimuli used to elicit a response through, for example, multiple-choice or constructed-response items as well as tasks; and “forms” are systematic collections of test items and tasks that comprise the testing experience for a particular student in a grade/content area.

Overview of Assessment Criteria

A. Meet Overall Assessment Goals and Ensure Technical Quality

- A.1 Indicating progress toward college and career readiness
- A.2 Ensuring that assessments are valid for required and intended purposes
- A.3 Ensuring that assessments are reliable
- A.4 Ensuring that assessments are designed and implemented to yield valid and consistent test score interpretations within and across years
- A.5 Providing accessibility to *all* students, including English learners and students with disabilities
- A.6 Ensuring transparency of test design and expectations
- A.7 Meeting all requirements for data privacy and ownership

B. Align to Standards – English Language Arts/Literacy

- B.1 Assessing student reading and writing achievement in both ELA and literacy
- B.2 Focusing on complexity of texts
- B.3 Requiring students to read closely and use evidence from texts
- B.4 Requiring a range of cognitive demand
- B.5 Assessing writing
- B.6 Emphasizing vocabulary and language skills
- B.7 Assessing research and inquiry
- B.8 Assessing speaking and listening
- B.9 Ensuring high-quality items and a variety of item types

C. Align to Standards – Mathematics

- C.1 Focusing strongly on the content most needed for success in later mathematics
- C.2 Assessing a balance of concepts, procedures, and applications
- C.3 Connecting practice to content
- C.4 Requiring a range of cognitive demand
- C.5 Ensuring high-quality items and a variety of item types

D. Yield Valuable Reports on Student Progress and Performance

- D.1 Focusing on student achievement and progress to readiness
- D.2 Providing timely data that inform instruction

E. Adhere to Best Practices in Test Administration

- E.1 Maintaining necessary standardization and ensuring test security

F. State Specific Criteria (as desired)

Sample criteria might include

- Requiring involvement of the state’s K-12 educators and institutions of higher education
- Procuring a system of aligned assessments, including diagnostic and interim assessments
- Ensuring interoperability of computer-administered items

Assessment Criteria and Evidence

A. Meet Overall Assessment Goals and Ensure Technical Quality*

| Criteria | Evidence |
|--|---|
| <p>A.1 Indicating progress toward college and career readiness: Scores and performance levels on assessments are mapped to determinations of college and career readiness at the high school level and for other grades to being on track to college and career readiness by the time of high school graduation.</p> | <ul style="list-style-type: none"> • A description is provided of the process for developing performance level descriptors and setting performance standards (i.e., “cut scores”), including <ul style="list-style-type: none"> ○ Appropriate involvement of higher education and career/technical experts in determining the score at which there is a high probability that a student is college and career ready; ○ External evidence used to inform the setting of performance standards and a rationale for why certain forms of evidence are included and others are not (e.g., student performance on current state assessments, NAEP, TIMSS, PISA, ASVAB, ACT, SAT, results from Smarter Balanced and PARCC, relevant data on post-secondary performance, remediation, and workforce readiness); ○ Evidence and a rationale that the method(s) for including external benchmarks are valid for the intended purposes; and ○ Standard setting studies, the resulting performance level descriptors and performance standards, and the specific data on which they are based (when available). • A description is provided of the intended studies that will be conducted to evaluate the validity of performance standards over time. |
| <p>A.2 Ensuring that assessments are valid for required and intended purposes: Assessments produce data, including student achievement data and student growth data required under Title I of the Elementary and Secondary Education Act (ESEA) and ESEA Flexibility, that can be used to validly inform the following:</p> <ul style="list-style-type: none"> • School effectiveness and improvement; • Individual principal and teacher effectiveness for purposes of evaluation and identification of professional development and support needs; • Individual student gains and performance; and • Other purposes defined by the state. | <ul style="list-style-type: none"> • A well-articulated validity evaluation based on an interpretive argument (e.g., Kane, 2006) is provided that includes, at a minimum <ul style="list-style-type: none"> ○ Evidence of the validity of using results from the assessments for the three primary purposes, as well as any additional purposes required by the state (specify sources of data). ○ Evidence that scoring and reporting structures are consistent with structures of the state’s standards (specify sources of data). ○ Evidence that total test and relevant sub-scores are related to external variables as expected (e.g., other measures of the construct). To the extent possible, include evidence that the items are “instructionally sensitive,” that is, that item performance is more related to the quality of instruction than to out-of-school factors such as demographic variables. ○ Evidence that the assessments lead to the intended outcomes (i.e., meet the intended purposes) and minimize unintended negative consequences. Consequential evidence |

* The term “technical quality” here refers to the qualities necessary to ensure that scoring and generalization inferences based on test scores are valid both within and across years. This document prioritizes certain aspects of technical quality, but as noted in the introduction, readers should also refer to other sources, primarily *The Standards for Educational and Psychological Testing*.

| Criteria | Evidence |
|--|--|
| | <p>should flow from a well-articulated theory of action about how the assessments are intended to work and be integrated with the larger accountability system.</p> <ul style="list-style-type: none"> ○ The set of content standards against which the assessments are designed is provided. If these standards are the state’s standards, evidence is provided that the content of the assessments reflects the standards, including the cognitive demand of the standards. If they are not the state’s standards, evidence is provided of the extent of alignment with the state’s standards. ○ Evidence is provided to ensure the content validity of test forms and the usefulness of score reports (e.g., test blueprints demonstrate the learning progressions reflected in the standards, and experts in the content and progression toward readiness are significantly involved in the development process). |
| <p>A.3 Ensuring that assessments are reliable: Assessments minimize error that may distort interpretations of results, estimate the magnitude of error, and inform users of its magnitude.</p> | <ul style="list-style-type: none"> ● Evidence is provided of the reliability of assessment scores, based on the state’s student population and reported subpopulations (specify sources of data). ● Evidence is provided that the scores are reliable for the intended purposes for essentially all students, as indicated by the standard error of measurement across the score continuum (i.e., conditional standard error). ● Evidence is provided of the precision of the assessments at cut scores, and consistency of student level classification (specify sources of data). ● Evidence is provided of generalizability for all relevant sources, such as variability of groups, internal consistency of item responses, variability among schools, consistency from form to form of the test, and inter-rater consistency in scoring (specify sources of data). |
| <p>A.4 Ensuring that assessments are designed and implemented to yield valid and consistent test score interpretations within and across years:</p> <ul style="list-style-type: none"> ● Assessment forms yield consistent score meanings over time, forms within year, student groups, and delivery mechanisms (e.g., paper, computer, including multiple computer platforms). ● Score scales used facilitate accurate and meaningful inferences about test performance. | <ul style="list-style-type: none"> ● A description is provided of the process used to ensure comparability of assessments and assessment results across groups and time. ● Evidence is provided of valid and reliable linking procedures to ensure that the scores derived from the assessments are comparable within year across various test “forms” and across time. ● Evidence is provided that the linking design and results are valid for test scores across the achievement continuum. ● Evidence is provided that the procedures used to transform raw scores to scale scores is coherent with the test design and the intended claims, including the types of Item Response Theory (IRT) calibration and scaling methods (if used) and other methods for facilitating meaningful score interpretations over tests and time. ● Evidence is provided that the assessments are designed and scaled to ensure the primary |

| Criteria | Evidence |
|---|---|
| | <p>interpretations of the assessment can be fulfilled. For example, if the assessments are used as data sources for growth or value-added models for accountability purposes, evidence should be provided that the scaling and design features would support such uses, such as ensuring appropriate amounts of measurement information throughout the scale, as appropriate.</p> <ul style="list-style-type: none"> • Evidence is provided, where a vertical or other score scale is used, that the scaling design and procedures lead to valid and reliable score interpretations over the full length of the scale proposed; and evidence is provided that the scale is able to maintain these properties over time (or a description of the proposed procedures is provided). |
| <p>A.5 Providing accessibility to <i>all</i> students, including English learners and students with disabilities:</p> | |
| <ul style="list-style-type: none"> • Following the principles of universal design: The assessments are developed in accordance with the principles of universal design and sound testing practice, so that the testing interface, whether paper- or technology-based, does not impede student performance. | <ul style="list-style-type: none"> • A description is provided of the item development process used to reduce construct irrelevance (e.g., eliminating unnecessary clutter in graphics, reducing construct-irrelevant reading load as much as possible), including <ul style="list-style-type: none"> ○ The <i>test item</i> development process to remove potential challenges due to factors such as disability, ethnicity, culture, geographic location, socioeconomic condition, or gender; and ○ <i>Test form</i> development specifications that ensure that assessments are clear and comprehensible for all students. • Evidence is provided, including exemplar tests (paper and pencil forms or screen shots) illustrating principles of universal design. |
| <ul style="list-style-type: none"> • Offering appropriate accommodations and modifications: Allowable accommodations and modifications that maintain the constructs being assessed are offered where feasible and appropriate, and consider the access needs (e.g., cognitive, processing, sensory, physical, language) of the vast majority of students. | <ul style="list-style-type: none"> • A description is provided of the accessibility features that will be available, consistent with state policy (e.g., magnification, audio representation of graphic elements, linguistic simplification, text-to-speech, speech-to-text, Braille). • A description is provided of access to translations and definitions, consistent with state policy. • A description is provided of the construct validity of the available accessibility features with a plan that ensures that the scores of students who have accommodations or modifications that do not maintain the construct being assessed are not combined with those of the bulk of students when computing or reporting scores. |
| <ul style="list-style-type: none"> • Assessments produce valid and reliable scores for English learners. | <ul style="list-style-type: none"> • Evidence is provided that test items and accessibility features permit English learners to demonstrate their knowledge and abilities and do not contain features that unnecessarily prevent them from accessing the content of the item. Evidence should address: presentation, response, setting, and timing and scheduling (specify sources of data). |
| <ul style="list-style-type: none"> • Assessments produce valid and reliable scores for students with disabilities. | <ul style="list-style-type: none"> • Evidence is provided that test items and accessibility features permit students with disabilities to demonstrate their knowledge and abilities and do not contain features that |

| Criteria | Evidence |
|--|---|
| | unnecessarily prevent them from accessing the content of the item. Evidence should address: presentation, response, setting, and timing and scheduling (specify sources of data). |
| <p>A.6 Ensuring transparency of test design and expectations: Assessment design documents (e.g., item and test specifications) and sample test questions are made publicly available so that all stakeholders understand the purposes, expectations, and uses of the college- and career-ready assessments.</p> | <ul style="list-style-type: none"> • Evidence is provided, including test blueprints, showing the range of state standards covered, reporting categories, and percentage of assessment items and score points by reporting category. • Evidence is provided, including a release plan, showing the extent to which a representative sample of items will be released on a regular basis (e.g., annually) across every grade level and content area. • Sample items with annotations and answer rationales are provided. • Scoring rubrics for constructed-response items with sample responses are provided for each level of the rubric. • Item development specifications are provided. • Additional information is provided to the state to demonstrate the overall quality of the assessment design, including <ul style="list-style-type: none"> ○ Estimated testing time by grade level and content area; ○ Number of forms available by grade level and content area; ○ Plan for what percentage of items will be refreshed and how frequently; ○ Specifications for the various levels of cognitive demand and how each is to be represented by grade level and content area; and ○ For ELA/Literacy, data from text complexity analyses. |
| <p>A.7 Meeting all requirements for data privacy and ownership: All assessments must meet federal and state requirements for student privacy, and all data is owned exclusively by the state.</p> | <ul style="list-style-type: none"> • An assurance is provided of student privacy protection, reflecting compliance with all applicable federal and state laws and requirements. • An assurance is provided of state ownership of all data, reflecting knowledge of state laws and requirements. • An assurance is provided that the state will receive all underlying data, in a timely and useable fashion, so it can do further analysis as desired, including, for example, achievement, verification, forensic, and security analyses. • A description is provided for how data will be managed securely, including, for example, as data is transferred between vendors and the state. |

B. Align to Standards – English Language Arts/Literacy

| Criteria | Evidence |
|---|---|
| <p>B.1 Assessing student reading and writing achievement in both ELA and literacy: The assessments are English language arts and literacy tests that are based on an aligned balance of high-quality literary and informational texts.</p> | <ul style="list-style-type: none"> • Test blueprints and other specifications as well as exemplar literary and informational passages are provided for each grade level, demonstrating the expectations below are met. • Texts are balanced across literary and informational text types and across genres, with more informational than literary texts used as the assessments move up in the grade bands, as the state’s standards require. <i>For example, for common core aligned assessments, goals include</i> <ul style="list-style-type: none"> ○ <i>In grades 3-8, approximately half of the texts are literature and half are informational;</i> ○ <i>In high school, because comprehension of complex informational texts is crucial for readiness, texts are approximately one-third literature and two-thirds informational; and</i> ○ <i>In all grades, informational texts are primarily expository rather than narrative in structure, and in grades 6-12, informational texts are approximately one-third each literary nonfiction, history/social science, and science/technical.</i> • Texts and other stimuli (e.g., audio, visual, graphic) are previously published or of publishable quality. They are content-rich, exhibit exceptional craft and thought, and/or provide useful information. • History/social studies and science/technical texts, specifically, reflect the quality of writing that is produced by authorities in the particular academic discipline. |
| <p>B.2 Focusing on complexity of texts: The assessments require appropriate levels of text complexity; they raise the bar for text complexity each year so students are ready for the demands of college- and career-level reading no later than the end of high school. Multiple forms of authentic, previously published texts are assessed, including written, audio, visual, and graphic, as technology and assessment constraints permit.</p> | <ul style="list-style-type: none"> • Text complexity measurements, exemplar literary and informational passages for each grade level, and other evidence (e.g., data, tools, procedures) are provided to demonstrate the expectations below are met. • At each grade, reading texts have sufficient complexity, and the average complexity of texts increases grade-by-grade, meeting college- and career-ready levels by the end of high school. • A rationale and evidence are provided for how text complexity is quantitatively and qualitatively measured and used to place each text at the appropriate grade level. <i>For example, for common core aligned assessments, goals include</i> <ul style="list-style-type: none"> ○ <i>Texts are placed in a grade band using at least one research-based quantitative measure;</i> ○ <i>Texts are placed at a grade level using a qualitative analysis measure, reflecting the expert judgment of educators; and</i> ○ <i>Most of the texts are placed within the grade band indicated by the quantitative</i> |

| Criteria | Evidence |
|--|--|
| <p>B.3 Requiring students to read closely and use evidence from texts: Reading assessments consist of test questions or tasks, as appropriate, that demand that students read carefully and deeply and use specific evidence from increasingly complex texts to obtain and defend correct responses.</p> | <p><i>analysis, with exceptions usually found in high school literary texts.</i></p> <ul style="list-style-type: none"> • Test blueprints and other specifications as well as exemplar test items are provided for each grade level, demonstrating the expectations below are met. • All reading questions are text-dependent and <ul style="list-style-type: none"> ○ Arise from and require close reading and analysis of text; ○ Focus on the central ideas and important particulars of the text, rather than on superficial or peripheral concepts; and ○ Assess the depth and specific requirements delineated in the standards at each grade level (i.e., the concepts, topics, and texts specifically named in the grade-level standards). • Many reading questions require students to directly provide textual evidence in support of their responses. <p><i>For example, for common core aligned assessments, goals include</i></p> <ul style="list-style-type: none"> ○ A majority of reading score points is devoted to questions that ask students to directly provide textual evidence in support of their responses (e.g., constructed-response and/or two-part evidence-based selected-response item formats). |
| <p>B.4 Requiring a range of cognitive demand: The assessments require all students to demonstrate a range of higher-order, analytical thinking skills in reading and writing based on the depth and complexity of college- and career-ready standards, allowing robust information to be gathered for students with varied levels of achievement.</p> | <ul style="list-style-type: none"> • Test blueprints and other specifications are provided to demonstrate that the distribution of cognitive demand for each grade level and content area is sufficient to assess the depth and complexity of the state’s standards, as evidenced by use of a generic taxonomy (e.g., Webb’s Depth of Knowledge) or, preferably, classifications specific to the discipline and drawn from the requirements of the standards themselves and item response modes, such as <ul style="list-style-type: none"> ○ The complexity of the text on which an item is based; ○ The range of textual evidence an item requires (how many parts of text[s] students must locate and use to response to the item correctly); ○ The level of inference required; and ○ The mode of student response (e.g., selected-response, constructed-response). • A rationale is provided justifying the distribution of cognitive demand for each grade level and content area. • Exemplar test items for each grade level are provided, illustrating each level of cognitive demand, and accompanied by a description of the process used to determine an item’s cognitive level. |
| <p>B.5 Assessing writing: Assessments emphasize writing tasks that require students to engage in close reading and analysis of texts so that students can demonstrate college- and career-ready abilities.</p> | <ul style="list-style-type: none"> • Test blueprints and other specifications as well as exemplar test items for each grade level are provided, demonstrating the expectations below are met. • Writing tasks reflect the types of writing that will prepare students for the work required in college and the workplace, balancing expository, persuasive/argument, and narrative writing, as state standards require. At higher grade levels, the balance shifts toward more exposition and argument. |

| Criteria | Evidence |
|--|--|
| | <p><i>For example, for common core aligned assessments, goals include</i></p> <ul style="list-style-type: none"> ○ <i>Taking all forms of the test together, writing tasks are approximately one-third each exposition, argument, and narrative (some tasks may represent blended structures), with the balance shifting toward more exposition and argument at the higher grade levels.</i> <ul style="list-style-type: none"> ● Tasks (including narrative tasks) require students to confront text or other stimuli directly, to draw on textual evidence, and to support valid inferences from text or stimuli. |
| <p>B.6 Emphasizing vocabulary and language skills: The assessments require students to demonstrate proficiency in the use of language, including vocabulary and conventions.</p> | <ul style="list-style-type: none"> ● Test blueprints and other specifications as well as exemplar test items for each grade level are provided, demonstrating the expectations below are met. ● Vocabulary items reflect requirements for college and career readiness, including <ul style="list-style-type: none"> ○ Focusing on general academic (tier 2) words; ○ Asking students to use context to determine meaning; and ○ Assessing words that are important to the central ideas of the text. ● Language is assessed within writing assessments as part of the scoring rubric, or it is assessed with test items that specifically address language skills. Language assessments reflect requirements for college and career readiness by <ul style="list-style-type: none"> ○ Mirroring real-world activities (e.g., actual editing or revision, actual writing); and ○ Focusing on common student errors and those conventions most important for readiness. ● Assessments place sufficient emphasis on vocabulary and language skills (i.e., a significant percentage of the score points is devoted to these skills). |
| <p>B.7 Assessing research and inquiry: The assessments require students to demonstrate research and inquiry skills, demonstrated by the ability to find, process, synthesize, organize, and use information from sources.</p> | <ul style="list-style-type: none"> ● Test blueprints and other specifications as well as exemplar test items for each grade level are provided, demonstrating the expectations below are met. ● Test items assessing research and inquiry mirror real world activities and require students to analyze, synthesize, organize, and use information from sources. <p><i>For example, for common core aligned assessments, goals include</i></p> <ul style="list-style-type: none"> ○ <i>Research tasks require writing to sources, including analyzing, selecting, and organizing evidence from more than one source, and often from sources in diverse formats; and</i> ○ <i>When assessment constraints permit, real or simulated research tasks comprise a significant percentage of score points when all forms of the reading and writing test are considered together.</i> |
| <p>B.8 Assessing speaking and listening: Over time, and as assessment advances allow, the assessments measure the speaking and listening communication skills students need for college and career readiness.</p> | <ul style="list-style-type: none"> ● Over time, and as assessment advances allow, the speaking and listening skills required for college and career readiness are assessed. <p><i>For example, for common core aligned assessments, test items assessing speaking</i></p> <ul style="list-style-type: none"> ○ <i>Assess students’ ability to express well-supported ideas clearly and to probe others’ ideas; and</i> |

| Criteria | Evidence |
|--|--|
| | <ul style="list-style-type: none"> ○ <i>Include items that measure students' ability to marshal evidence from research and orally present findings in a performance task.</i> <p><i>For example, for common core aligned assessments, test items assessing listening</i></p> <ul style="list-style-type: none"> ○ <i>Are based on texts and other stimuli that meet the criteria for complexity, range, and quality outlined in criteria B.1 and B.2 above; and</i> ○ <i>Permit the evaluation of active listening skills (e.g., taking notes on main ideas, elaborating on remarks of others).</i> |
| <p>B.9 Ensuring high-quality items and a variety of item types: High-quality items and a variety of types are strategically used to appropriately assess the standard(s).</p> | <ul style="list-style-type: none"> • Specifications are provided to demonstrate that the distribution of item types for each grade level and content area is sufficient to strategically assess the depth and complexity of the standards being addressed. Item types may include, for example, selected-response, two-part evidence-based selected-response, short and extended constructed-response, technology-enhanced, and performance tasks. • To support claims of quality, the following are provided: <ul style="list-style-type: none"> ○ Exemplar items for each item type used in each grade band; ○ Rationales for the use of the specific item types; ○ Specifications showing the proportion of item types on a form; ○ For constructed response and performance tasks, a scoring plan (e.g., machine-scored, hand-scored, by whom, how trained), scoring rubrics, and sample student work to confirm the validity of the scoring process; and ○ A description of the process used for ensuring the technical quality, alignment to standards, and editorial accuracy of the items. |

C. Align to Standards – Mathematics

| Criteria | Evidence |
|--|--|
| <p>C.1 Focusing strongly on the content most needed for success in later mathematics: The assessments help educators keep students on track to readiness by focusing strongly on the content most needed in each grade or course for later mathematics.</p> | <ul style="list-style-type: none"> • Test blueprints and other specifications are provided, demonstrating that the vast majority of score points in each assessment focuses on the content that is most important for students to master in that grade band in order to reach college and career readiness. For each grade band, this content consists of <ul style="list-style-type: none"> ○ Elementary grades – number and operations; ○ Middle school – ratio, proportional relationships, pre-algebra, and algebra; and ○ High school – prerequisites for careers and a wide range of postsecondary studies, particularly algebra, functions, and modeling applications. <p><i>For example, for common core aligned assessments, goals include</i></p> <ul style="list-style-type: none"> ○ <i>In elementary grades, at least three-quarters of the points in each grade align exclusively to the major work of the grade;</i> ○ <i>In middle school grades, at least two-thirds of the points in each grade align exclusively</i> |

| Criteria | Evidence |
|---|---|
| | <p style="text-align: center;"><i>to the major work of the grade; and</i></p> <ul style="list-style-type: none"> ○ <i>In high school, at least half of the points in each course align exclusively to prerequisites for careers and a wide range of postsecondary studies.</i> • The assessment design reflects the state’s standards and reflects a coherent progression of mathematics content from grade to grade and course to course. |
| <p>C.2 Assessing a balance of concepts, procedures, and applications: The assessments measure conceptual understanding, fluency and procedural skill, and application of mathematics, as set out in college- and career-ready standards.</p> | <ul style="list-style-type: none"> • Test blueprints and other specifications as well as exemplar test items for each grade level are provided, demonstrating the expectations below are met. • The distribution of score points reflects a balance of mathematical concepts, procedures/fluency, and applications, as the state’s standards require. <i>For example, for common core aligned assessments, at least one-quarter of the points come from each of the following categories:</i> <ul style="list-style-type: none"> ○ <i>Conceptual understanding problems in which students to respond to well-designed conceptual problems;</i> ○ <i>Procedural skill and fluency problems (e.g., purely procedural problems, some requiring use of efficient algorithms, and others inviting opportunistic strategies); and</i> ○ <i>Application problems (e.g., in elementary and middle grades, solving grade-appropriate word problems reflecting growing complexity across the grades; in high school, rich application problems requiring students to demonstrate college and career readiness).</i> • All students, whether high performing or low performing, are required to respond to items within the categories of conceptual understanding, procedural skill and fluency, and applications, so they have the opportunity to show what they know and can do. |
| <p>C.3 Connecting practice to content: The assessments include brief questions and also longer questions that connect the most important mathematical content of the grade or course to mathematical practices, for example, modeling and making mathematical arguments.</p> | <ul style="list-style-type: none"> • Test blueprints and other specifications as well as exemplar test items for each grade level are provided, demonstrating the expectations below are met. • Assessments for each grade and course meaningfully connect mathematical practices and processes with mathematical content (especially with the most important mathematical content at each grade), as required by the state’s standards. • Explanatory materials (citing test blueprints and other specifications) describe the connection for each grade or course between content and mathematical practices and processes. <i>For example, for common core aligned assessments, goals include</i> <ul style="list-style-type: none"> ○ <i>Every test item that assesses mathematical practices is also aligned to one or more content standards (most often within the major work of the grade); and</i> ○ <i>Through the grades, test items reflect growing sophistication of mathematical practices with appropriate expectations at each grade level.</i> |

| Criteria | Evidence |
|---|---|
| <p>C.4 Requiring a range of cognitive demand: The assessments require all students to demonstrate a range of higher-order, analytical thinking skills in reading and writing based on the depth and complexity of college- and career-ready standards, allowing robust information to be gathered for students with varied levels of achievement. Assessments include questions, tasks, and prompts about the basic content of the grade or course as well as questions that reflect the complex challenge of college- and career-ready standards.</p> | <ul style="list-style-type: none"> • Test blueprints and other specifications are provided to demonstrate that the distribution of cognitive demand for each grade level is sufficient to assess the depth and complexity of the state’s standards, as evidenced by use a of generic taxonomy (e.g., Webb’s Depth of Knowledge) or, preferably, classifications specific to the discipline and drawn from mathematical factors, such as <ul style="list-style-type: none"> ○ Mathematical topic coverage in the task (single topic vs. two topics vs. three topics vs. four or more topics); ○ Nature of reasoning (none, simple, moderate, complex); ○ Nature of computation (none, simple numeric, complex numeric or simple symbolic, complex symbolic); ○ Nature of application (none, routine word problem, non-routine or less well-posed word problem, fuller coverage of the modeling cycle); and ○ Cognitive actions (knowing or remembering, executing, understanding, investigating, or proving). • A rationale is provided justifying the distribution of cognitive demand for each grade level and content area. • Exemplar test items for each grade level are provided, illustrating each level of cognitive demand, and accompanied by a description of the process used to determine an item’s cognitive level. |
| <p>C.5 Ensuring high-quality items and a variety of item types: High-quality items and a variety of item types are strategically used to appropriately assess the standard(s).</p> | <ul style="list-style-type: none"> • Specifications are provided to demonstrate that the distribution of item types for each grade level and content area is sufficient to strategically assess the depth and complexity of the standards being addressed. Item types may include selected-response, short and extended constructed-response, technology-enhanced, and multi-step problems. • To support claims of quality the following are provided: <ul style="list-style-type: none"> ○ The list and distribution of the types of work students will be asked to produce (e.g., facts, computation, diagrams, models, explanations); ○ Exemplar items for each item type used in each grade band; ○ Rationales for the use of the specific item types; ○ Specifications showing the proportion of item types on a form; ○ For constructed response items, a scoring plan (e.g., machine-scored, hand-scored, by whom, how trained), scoring rubrics, and sample student work to confirm the validity of the scoring process; and ○ A description of the process used for ensuring the technical quality, alignment to standards, and editorial accuracy of the items. |

D. Yield Valuable Reports on Student Progress and Performance

| Criteria | Evidence |
|---|--|
| <p>D.1 Focusing on student achievement and progress to readiness: Score reports illustrate a student’s progress on the continuum toward college and career readiness, grade by grade, and course by course. Reports stress the most important content, skills, and processes, and how the assessment focuses on them, to show whether or not students are on track to readiness.</p> | <ul style="list-style-type: none"> • A list of reports is provided, and for each report, a sample that shows, at a minimum <ul style="list-style-type: none"> ○ Scores and sub-scores that will be reported with emphasis on the most important content, skills, and processes for each grade or course; ○ Explanations of results that are instructionally valuable and easily understood by essentially all audiences; ○ Results expressed in terms of performance standards (i.e., proficiency “cut scores”), not just scale scores or percentiles; and ○ Progress on the continuum toward college and career readiness, which can be expressed by whether a student has sufficiently mastered the current grade or course content and is therefore prepared for the next level. <p>(Note: Not all reporting information need be numerical; for example, actual student work on a released item could be presented, along with the rubric for the item and a discussion of common errors.)</p> <ul style="list-style-type: none"> • The reporting structure can be supported by the assessment design, as demonstrated by evidence, including data confirming that test blueprints include a sufficient number of items for each reporting category, so that scores and sub-scores lead to the intended interpretations and minimize the possibility of misinterpretation. |
| <p>D.2 Providing timely data that inform instruction: Reports are instructionally valuable, easy to understand by all audiences, and delivered in time to provide useful, actionable data to students, parents, and teachers.</p> | <ul style="list-style-type: none"> • A timeline and other evidence are provided to show when assessment results will be available for each report. • A description is provided of the process and technology that will be used to issue reports in as timely a manner as possible. • Evidence, including results of user testing, is provided to demonstrate the utility of the reports for each intended audience. |

E. Adhere to Best Practices in Test Administration

| Criteria | Evidence |
|--|---|
| <p>E.1 Maintaining necessary standardization and ensuring test security: In order to ensure the validity, fairness, and integrity of state test results, the assessment systems maintain the security of the items and tests as well as the answer documents and related ancillary materials that result from test administrations.</p> | <ul style="list-style-type: none"> • A comprehensive security plan is provided with auditable policies and procedures for test development, administration, score reporting, data management, and detection of irregularities consistent with NCEs and CCSSO recommendations for, at a minimum <ul style="list-style-type: none"> ○ Training for all personnel – both test developers and administrators; ○ Secure management of assessments and assessment data, so that no individual gains access to unauthorized information; ○ Test administration and environment; and ○ Methods used to detect testing irregularities before, during, and after testing, and steps |

| Criteria | Evidence |
|----------|---|
| | <p style="text-align: center;">to address them.</p> <ul style="list-style-type: none"> • A description is provided of how security safeguards have been tested and validated for computer-based tests and for paper-and-pencil tests, as relevant. |

F. State Specific Criteria (as desired)

It is likely that states will supplement the above criteria with criteria specific to their needs. These might, for example, include

- **Requiring involvement of the state’s K-12 educators, institutions of higher education, and career/technical experts** in the design, development, and/or scoring of the assessments;
- **Procuring a system of aligned assessments, including diagnostic and interim assessments** designed to target and improve instruction as well as measure progress and performance; and
- **Ensuring interoperability of computer-administered items** consistent in all ways with the specifications laid out in the *Assessment Interoperability Framework* (2012) developed by the Common Education Data Standards (CEDS) project, so that tests and items owned by the state can be easily ported from one technology platform to another.

Bibliography

- American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) Joint Committee on Standards for Educational and Psychological Testing. (1999). *Standards for educational and psychological testing*. Washington DC: AERA. Retrieved March 21, 2014, from <http://www.apa.org/science/programs/testing/standards.aspx>.
- Chingos, M.M. (2013). *Standardized testing and the common core standards: You get what you pay for?* Washington, DC: Brown Center on Education Policy at Brookings. Retrieved March 21, 2014, from <http://www.brookings.edu/research/reports/2013/10/30-standardized-testing-and-the-common-core-chingos>.
- CCSSO. (2013). *States' commitment to high-quality assessments aligned to college- and career-readiness*. Washington, DC: Author. Retrieved March 21, 2014, from <http://www.ccsso.org/Documents/2013/CCSSO%20Assessment%20Quality%20Principles%2010-13%20FINAL.pdf>.
- CCSSO and the Association of Test Publishers (ATP). (2010). *Operational best practices for statewide large-scale assessment programs*. Washington, DC: Author. Retrieved March 21, 2014, from http://www.ccsso.org/resources/publications/operational_best_practices_for_statewide_large-scale_assessment_programs.html.
- Common Education Data Standards (CEDS). (2012). *Assessment interoperability framework*. Washington, DC: Author. Retrieved March 21, 2014, from <https://ceds.ed.gov/aif.aspx>.
- Darling-Hammond, L., & Adamson, F. (2013). *Developing assessments of deeper learning: The costs and benefits of using tests that help students learn*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education. Retrieved March 21, 2014, from <http://edpolicy.stanford.edu/publications/pubs/745>.
- Darling-Hammond, L., Herman, J., Pellegrino, J., et al. (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity. Retrieved March 21, 2014, from <https://edpolicy.stanford.edu/publications/pubs/847>.
- Kane, M.T. (2006). Validation. In R.L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Washington, DC: The National Council on Measurement in Education and the American Council on Education.
- Michigan Department of Education. (2013). *Report on options for assessments aligned with the common core state standards, submitted to the Michigan Legislature*. Lansing, MI: Michigan Department of Education. Retrieved March 21, 2014, from http://www.michigan.gov/documents/mde/Common_Core_Assessment_Option_Report_441322_7.pdf.
- Student Achievement Partners. (2013). *Assessment evaluation tool for CCSS alignment in ELA/literacy grades 3-12*. NYC: Author. Retrieved March 21, 2014, from <http://www.achievethecore.org/page/298/aet-ela-literacy-grades-3-12>.
- Student Achievement Partners. (2013). *Assessment evaluation tool for CCSS alignment in mathematics grades K-HS*. NYC: Author. Retrieved March 21, 2014, from <http://www.achievethecore.org/page/297/aet-mathematics-grades-k-hs>.
- U.S. Department of Education. (2010). *Overview information: Race to the top fund, notice inviting applications for new awards for fiscal year (FY) 2010*. Washington, DC: Author. Retrieved March 21, 2014, from <http://www2.ed.gov/programs/racetothetop-assessment/resources.html>.
- U.S. Department of Education. (2013). *Race to the top assessment program, technical review process, April 2013*. Washington, DC: Author. Retrieved March 21, 2014, from <http://www2.ed.gov/programs/racetothetop-assessment/technical-review-process.pdf>.

- U.S. Department of Education, Institute of Education Sciences National Center for Education Statistics. (2013). *Testing integrity symposium: Issues and recommendations for best practice*. Washington, DC: Author. Retrieved March 21, 2014, from <http://nces.ed.gov/pubs2013/2013454.pdf>.
- U.S. Department of Education, Office of Elementary and Secondary Education. (2007). *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001*. Washington, DC: Author. Retrieved March 21, 2014, from <http://www2.ed.gov/policy/elsec/guid/saaprguidance.pdf>.