Top Ten: Transitioning English Language Arts Assessments

White Paper

Delise Becker
Michael Bay-Borelli
Lee Brinkerhoff
Kellie Crain
Laurie Davis
Charles Fuhrken
Tiffany Hartmann
Jay Larkin
Kimberly O’Malley
Suzanne Trevvett

May 2011
About Pearson
Pearson, the global leader in education and education technology, provides innovative print and digital education materials for preK through college, student information systems and learning management systems, teacher licensure testing, teacher professional development, career certification programs, and testing and assessment products that set the standard for the industry. Pearson’s other primary businesses include the Financial Times Group and the Penguin Group. For more information about the Assessment & Information group of Pearson, visit http://www.pearsonassessments.com/.

About Pearson’s White Papers
Pearson’s white paper series shares the perspectives of our assessment experts on selected topics of interest to educators, researchers, policy makers and other stakeholders involved in assessment and instruction. Pearson’s publications in .pdf format may be obtained at: http://www.pearsonassessments.com/research.
Abstract

Most states will transition their English language arts (ELA) assessments in the next several years to align with new, more rigorous ELA content standards. This document is designed to identify the top ten considerations that states and consortia will need to address as they plan the transition of their ELA assessments. Because specific state context, policies, and test development processes will influence the transition plans, this document focuses on the questions to be asked, rather than attempting to provide answers to the questions or considerations posed.

*Keywords:* English language arts, transition, considerations
Top Ten: Transitioning English Language Arts Assessments

Most states will transition their English language arts (ELA) assessments in the next several years to align with new, more rigorous ELA content standards. For more than 40 states, the transition is driven by the adoption of the Common Core State Standards (CCSS) and the need to align the state ELA assessment with the new CCSS standards. Of those states that have not adopted the CCSS, many, such as Texas, Virginia, and Alaska, are revising their content standards in response to the national educational trend of increasing the rigor of what students need to know so that they are prepared for college and career. Thus, for a variety of reasons, nearly all states will transition their ELA assessments to align with changes to content standards.

This document is designed for assessment experts who work in or with state education departments. The purpose is to identify the top ten considerations that states and consortia will need to address as they plan the transition of their summative, or large-scale, ELA assessments. Because specific state context, policies, and test development processes will influence the transition plans, this document focuses on the questions to be asked, rather than attempting to provide answers to the questions or considerations posed. That is, these top ten considerations should serve as conversation starters or as a structure for the conversations that states, consortia, and assessment providers will need to have in planning the transition. Though some of the considerations may apply to other content areas, this document attends to those issues that are particularly salient to ELA.

1. Identifying the Standards to be Assessed
As the transition of ELA assessments to new curriculum standards (e.g., CCSS) is planned, will all or only some of the new ELA standards be assessed? More specific questions to ask:

- Which ELA domains will be assessed?
- If speaking and listening are assessed, will academic discussion skills be assessed?
- To what extent will research skills be assessed?
- Will skills related to multimedia be assessed?

And if these skills are to be assessed on one type of assessment (e.g., formative), will they also be assessed on other types (e.g., summative)?

Whereas all of the skills in the new ELA standards should be taught in classrooms, large-scale summative assessments cannot always measure student achievement relative to every curriculum standard. For example, the CCSS ELA standards ask students to “conduct sustained research projects” (Council of Chief State School Officers [CCSSO] & National Governors Association [NGA], June 2010, p. 17) and to “write routinely over extended time frames” (CCSSO & NGA, June 2010, p. 18). Although classroom-based assessments could feasibly address such standards, these standards would be challenging to address in a one- or two-session, end-of-year summative assessment.

In transitioning ELA assessments, a critical first step will be determining which ELA skills will be eligible (for which type of assessment) in the new testing programs. This determination is driven in part by whether the tests under consideration are classroom-based formative assessments or large-scale summative tests, and to some extent, by the testing formats
and item types to be used. Other considerations such as test length, testing time, cognitive demand, technology constraints, etc., may also factor into this determination.

If stakeholders are to feel ownership in the new ELA testing program, the process used to determine the eligibility of curriculum standards should be inclusive. Efforts should be made to involve policy makers, content specialists, and educators in deciding which skills will be assessed on which type of assessments. When making eligibility decisions, the criteria used should be applied consistently across grades.

2. Selecting Viable Assessment Instruments for a Testing Program

*With improvements in technology and an increased desire to assess student performance throughout the school year, rather than in a single summative assessment, how do states determine which assessment instruments to use?*

Every testing program must decide the assessment instrument(s) that makes the most sense in terms of measuring and reporting student progress. Summative, or end-of-year or end-of-course assessments, are a traditional approach, but today, assessment systems composed of a combination of assessment types (such as through-course, diagnostic, and formative) may meet current needs of state departments and in ELA classrooms.

The purpose of summative assessments is to summarize student learning at a particular point in time, usually at the end of an academic year or course. Summative assessments are valuable in that they provide an overall portrait of how well students learned course content; however, because summative assessments occur at the end of an academic year or course, the data do not inform instructional decisions for those particular test takers but rather can be used to inform how teachers shape their delivery of ELA content in the future. Traditionally, summative assessments in ELA that depend largely on multiple-choice items, sometimes partnered with constructed-response items, have been a valuable instrument on a large-scale basis because they are cost-effective, can be scored exclusively or mostly by machine, and can provide reliable and objective information quickly.

Currently on the national front, through-course assessments are being proposed because of their potential to expand the summative testing window and to provide more information throughout the year to students, teachers, and administrators than a single “snapshot” summative assessment affords. Whereas many summative test administrations are limited to a single class period, through-course assessments may extend over multiple days or class periods. With the additional time for both administration and scoring these assessments do not need to be limited to multiple choice and/or constructed response; they can be performance-task oriented (such as conducting research on a nonfiction topic and producing a multimedia product). Another matter to consider is the fact that skills in ELA develop recursively, so much thought will need to be placed on the standards to be measured, and the ways in which information is reported in order for data to be helpful in informing remediation and future instruction. Finally, keeping costs low might become a priority since students would produce multiple products that would need to be scored. Therefore, artificial intelligence or teacher scoring might factor into a reasonable solution.

Besides through-course assessments, other types of assessments (e.g., diagnostic, benchmark, formative) may also provide important gauges of student learning, perhaps at a particular time during the academic year. Diagnostic testing can be used to obtain data about students’ current knowledge and skills in ELA. For a diagnostic test to be effective, it has to measure a discrete set of skills. In many cases a set of standards will need to be broken into a smaller set of elemental or foundational skills to provide any meaningful data.
for teachers to use, especially for ELA in which skills develop recursively (as mentioned above). Benchmark tests are usually administered at designated intervals within the year to assess a subset of standards assessed on the summative test. State departments may be interested in benchmark assessments because the data obtained are often used to predict how students will perform on the summative test. An agreed upon curricular scope and sequence is required before the benchmark tests can be implemented. Formative assessment has the potential to provide teachers with immediate feedback about student learning. Formative assessment—sometimes called classroom assessment—often involves, for example, teachers providing students opportunities to express their learning and teachers’ observations of students performing tasks; for instance, teachers might ask students to discuss the effectiveness of literary language in a poem in order to hear students articulate their knowledge as well as to take advantage of opportunities to push students’ thinking. Formative assessment is ongoing and necessary in order to inform decisions about students’ mastery of skills or the need to reteach or remediate. An important purpose of these assessments is to positively impact decisions about the instructional design, score, and delivery of curricular content.

Finally, the advent of computer-based testing allows for faster delivery of scores (in some cases instantaneous) and the administration of new item types. The new item types allow for more domains (e.g., speaking and listening) to be assessed than is commonly found in traditional print-based assessments as well as allow more authentic and engaging tasks to be presented to test takers, such as using a drag-and-drop feature to order the plot events of a story or organize factual details of an informational article by pinning them appropriately on a fishbone diagram.

3. Determining Appropriate Testing Modality

With the continually increasing availability of new technology and the demand for assessments that will produce more immediate results that can be used to enhance instruction, educators will need to determine which modality, or delivery system, will best support the goals of an assessment program. Some questions to ask:

- How can technology be used to authentically assess a wider range of skills in the (ELA) curriculum?
- In what modalities will the ELA assessments be offered: paper, online, handheld devices, all?
- If ELA tests are to be offered in multiple modalities, how will comparability of score interpretation (paper/online) be addressed?
- If assessments are delivered in multiple modalities, how will varying levels of student familiarity and experience with those modalities be addressed?

The continual advancement of technology has created a variety of options for delivering all types of ELA assessments. Computer-based assessments provide an opportunity to assess a wide range of ELA skills using the online tools and technology that are often used to facilitate learning in the classroom. In addition, these new modalities can provide more immediate performance reports, provide opportunities for innovative approaches such as performance tasks requiring collaboration and speaking and listening skills, and may improve accessibility to the content for all students. However, before adopting one or more delivery systems for an assessment program, consideration must be given to the effect different modalities will have on the equity for all test takers and the comparability of student scores and performance between different modalities.
Currently, all students will not have the same level of experience with and access to computer-based technology. In addition, not all schools will have the same level of access to computers or handheld devices for use in administering the tests. Strategies for dealing with these inequities, such as providing equal training with test modalities and flexible, on-demand assessment schedules, have been developed but need to be planned carefully before adopting a computer-based modality.

Using multiple modalities to support an assessment program is appealing, but consideration must also be given to the comparability of test scores. Research has shown that processes in reading and writing can differ between print-based and computer-based assessments. For example, reading in an online format typically requires scrolling through multiple screens because less text can be presented online than on a printed page (Dillon, 1992). In addition, the process of writing might differ when done in different modalities. For example, when writing a composition on paper students tend to make greater use of rough drafts and preplanning before they start writing. However, when writing a composition in a computer-based environment, students tend to do less prewriting and compensate by using readily available revision tools (Van Waes & Schellens, 2003). These fundamental differences in the processes students use to engage in writing and reading tasks may make it difficult to establish comparability in a conventional way across testing modes.

4. Choosing Item and Stimuli Types

With increasing demand for authentic assessments to enhance instruction and an emphasis on providing instruction and experience with 21st century skills, which item approaches will best measure student mastery and support the goals of the assessment program?

With advancements in computer-based technology, assessments of a wider range of skills and deeper cognitive complexity are possible. Innovative approaches offered by computer-based technology allow for item types and items that are more representative of instruction, are more cognitively sophisticated, and which can be scored more quickly. Using technology such as drag and drop, hot spot, visual and oral media, and voice recording allow for items that can better assess a wider range of skills such as speaking and listening, research, and collaboration that have been difficult to assess in more conventional paper and pencil presentations. Additionally, these item types support a curriculum aligned to 21st century skills that students will need to be successful in higher education and the workplace. Innovative approaches also include examining traditional print fixed forms and determining if computer-adaptive possibilities offer solutions to instructional and assessment goals. Conventional (traditional) assessment approaches have used item types such as multiple choice and constructed/extended response to measure selective English language arts (ELA) skills appropriate to a print-based presentation. Though items supporting a conventional approach can also be delivered in online format, the items do not leverage the opportunities offered by computer-based technology. However, these items types and assessment approaches are familiar to students and provide valid and reliable measurement of student achievement. Educators will need to examine the wide range of item types available and determine which item types, or which mixture of item types and approaches, best support the goals of a new curriculum.

5. Selecting Texts

When assessing English language arts standards, what are the text considerations (e.g., types of texts; literature vs. informational emphasis; permissioned vs. commissioned selections; readability) that should be examined so that the tasks presented on an assessment more closely mirror authentic classroom practice?
The selection of texts for an English language arts assessment is a crucial issue because the texts themselves serve as the basis for many of the test tasks students will encounter; for example, to measure students’ understandings of important reading skills, test takers are asked first to read a text and then respond to items or tasks about the text (note that the term text in this paper represents texts in an online or paper format).

The types of texts that students will see on an assessment are usually indicated in the curriculum standards. For example, standards may be organized around categories of texts such as fictional texts, informational texts, and media. Embedded standards may further provide information about types of texts, such as literature includes realistic fiction, historical fiction, and poetry, and informational texts include expository and content-area reading, which might entail maps, text boxes, and other data displays. Assessments should feature these types of texts in order to mirror the everyday reading that students are completing in their literacy classrooms.

The emphasis of certain types of texts over others is an ongoing concern and debate on the national front. Traditionally, students have read more literary than informational texts both in classrooms and on assessments, especially at the lower grade levels. But about a decade ago, more informational texts began appearing in classrooms in an effort to prepare students as readers, writers, and researchers across the content areas (Harvey, 1998). College and career readiness standards emphasize the reading of nonfiction in both literacy and content-area classrooms because nonfiction is the kind of reading that happens in college and the workplace (Closing the Expectations Gap, 2011). Furthermore, standards may require students to synthesize ideas across multiple sources, so appropriate pairings of passages may also be an important consideration. In addition, the text must support the need for items developed to a variety of difficulty levels as well as cognitive complexity constructs. Therefore, the amount of or emphasis on literary and informational texts (as well as any other types of texts indicated in curriculum standards) must be taken into account when developing the test design and blueprint of an assessment.

The texts for use on an English language arts assessment typically come from two sources: permissioned texts (previously published texts for which a fee is paid to use them on an assessment) and commissioned texts (selections written by freelance writers for the sole purpose of appearing on an assessment). Permissioned texts are appealing for use on an assessment because they provide authentic reading experiences for students. However, the parameters of their use can pose challenges; for instance, publishers/authors may not allow changes to content (such as excising a reference that might confuse students) and may place restrictions on the use of the text (such as not allowing the text to be released to the public and/or charging high fees). Commissioned texts are appealing because test publishers essentially buy all rights to the texts and therefore can augment or modify the text as needed and can release the texts to the public without restrictions; nonetheless, commissioned texts may sometimes be perceived as not being as well written as permissioned texts. Furthermore, some assessed standards may be specific about the requirement of the text source; for instance, a “well-known speech” must come from a permissioned rather than commissioned source, whereas realistic fiction could be either permissioned or commissioned. Therefore, the source of texts is another important consideration when determining eligible texts to appear on an assessment.

The “readability” of texts (meaning the ease with which a text can be read and comprehended) is an equally important consideration when selecting texts appropriate for an assessment. During an assessment, students read independently and do not have the
benefit of teacher scaffolding or assistance in order to make sense of texts. Readability software (e.g., Harris-Jacobson, Fry, Spache) measures text complexity mostly by taking into account syntactic features (e.g., the complexity of sentence structure). Because of this limitation, the expertise of classroom teachers is often needed to consider more dynamic factors that can increase or decrease the difficulty of texts such as the amount of picture support, the organization of the text, students’ background knowledge about and interest in the topic, and so forth. Readability formulae paired with teachers’ knowledge of students’ independent reading abilities are often used to place texts at appropriate grade levels for assessments.

6. Choosing Assessment Materials for Public Consumption

In order to prepare for a new assessment, a state, district, or organization may publish a variety of materials. Materials such as sample items, practice tests, and specification documents provide valuable functions in transitioning the field from one set of standards or assessments to another. What must be considered when publishing these materials?

Due to the cycle of assessment development, reading and language arts stimuli frequently are developed as early as two years in advance of the estimated field-test year. Generally, this is to provide sufficient time so that any applicable bias and sensitivity, regional, and content issues have been resolved to meet the specifications of the state. When new standards are adopted and applied to an assessment, states will want to scrutinize legacy passages already in item banks prior to being shared with the public in order to ascertain if they remain robust enough to support both the cognitive and content requirements of the new standards.

When material is shared with the education field, it is possible that it will have impact beyond the intended audience. During the early years of new standards implementation, the limited availability in item and passage banks may perhaps curtail the amount of material published, and will make it challenging to provide enough material to represent the full spectrum of what is being tested by the new standards. An additional challenge for English language arts is that many items are passage based and reflect an application of a particular standard that is most likely unique to that passage. Therefore, an item testing identification of theme might not be as illustrative or concrete as a math item that assesses the order of operations. Additionally, a specific genre of passage will rarely allow the writing of items to all standards, or provide the definitive explanation or guide to preparing for a new assessment. States may want to plan additional item development at the start of new ELA assessments, so that materials can be released early in the transition process and can help students, teachers, and the public make a smooth transition.

Another consideration is the need to evaluate that the materials initially released to the field remain current as the testing program evolves. New training materials are often published in advance to prepare the field for a new assessment that is aligned to new state ELA standards. A practice test and preliminary set of item specifications are developed and may be published on the state website for field usage. Once the data on the field-test items have been analyzed, modifications to the item specifications and item types may be applied to capture more closely the intended skill of the new standards. Updates to the item specifications and item types may cause previously published materials to become out of date, and new materials will be needed in order to accommodate the evolution of the style and item types of the most recent standard set.
7. Assessing New Writing Standards

*How will the adoption of a new set of writing standards impact decisions about skill organization, item types, the inclusion of research, scoring and reporting, and technology and assessment mediums?*

Writing tasks and the associative language skills included within the writing domain can vary based on a preferred philosophical and educational approach. When transitioning to a new set of standards, many writing elements may need to be considered. For example, what research implies about each writing element, how scoring and reporting may need to be changed, and what mediums and item types would most effectively cover the new standards.

The organization of writing skills is generally reflected by either encapsulating the universal writing process or mode-specific tasks to measure a student’s skills. Understanding the standards and the organizational methodology helps guide decisions regarding the most appropriate approach in assessing the skills. Implementation of diverse or innovative item types may necessitate some trials and analysis to measure effectiveness and validity. More research is recommended on the various models and in how corresponding results aid in both determining these decisions and providing substantial support.

For some states, the adoption of the CCSS may require adjustments in scoring and reporting of the writing standards due to the organization of the ELA domains. For example, many states include in their writing standards and benchmarks concepts related to conventions such as grammar and usage. However, in the CCSS, the conventions standards are located in the language domain which also includes vocabulary and knowledge of language. Another example relates to research, often found in the reading domain of many state standards. In the CCSS, research is primarily addressed in the writing domain. The domain differences will involve the determination of an appropriate scoring and reporting strategy that most effectively reflects the intended skill set of the new standards. Additional considerations related to these efforts are addressed in more detail in the section below.

8. Addressing Dimensionality Issues

*At what level will students’ scores be reported when an ELA assessment measures a variety of standards and domains that might fall along different dimensions? Will reading, writing, listening, and speaking be reported as a single score, as separate scores, or as a composite? Will separate subscores be needed for individual strands/standards within the assessment? Can individual items contribute to multiple subscores or only to one?*

As states begin a transition to a new ELA assessment, it will be important to simultaneously consider how the content will be developed along with the desired scoring and reporting features for the assessment. ELA curricula have traditionally included four domains: reading, writing, speaking, and listening; however, many states have limited the assessable standards to the reading and writing domains. With the advent of the CCSS, states’ efforts to assess listening and speaking content standards will likely be increased. Whereas there is a logic and simplicity to talking about these domains as if they represented a single concept, it is not always so straightforward to build assessments that can provide both a unified ELA test score as well as feedback on domain-specific skills within the ELA domain.

Many teachers believe that reading and writing and listening and speaking skills are interrelated. They often teach pairs of subjects together, yet the pairs of domains can
behave differently when assessed. For example, students may have relatively strong reading comprehension skills, but be unable to coherently express themselves in a written composition. Assessments that have tasks targeted either at reading comprehension skills or writing skills may provide useful feedback about skills in the respective domains but may not lend themselves well to generating a single ELA test score.

Ways to address this would be to design assessment tasks which blend reading, writing, listening, and speaking skills. For example, it might be possible to design assessment tasks that emulate a research activity. Within the assessment students might be provided with a set of reference materials that they would have to read, discuss, and synthesize to generate and present a written product. These types of assessment tasks would likely provide a strong measure of students’ ELA skills but may not lend themselves well to generating more granular feedback about the students’ proficiency within reading or writing.

These issues may be compounded with the addition of speaking and listening domains that include skills such as presentation, conversation, and active listening. Given the diversity of these skills, one might expect students to have different profiles of proficiencies in each of these areas. States might look to English language proficiency testing for ideas on how to combine information across domains through mechanisms such as composite scores rather than through generation of a single ELA scale score.

9. Planning for National and International Comparisons
If student scores on the state assessment will be used for making national and international comparisons, which of the new ELA content standards and assessments will facilitate and which will challenge those comparisons?

What process should be implemented to understand how to best benchmark ELA scores across state, national, and international assessments?

How will the state balance the inclusion of authentic and culturally rich text in the assessments and the broadening of the language and content to facilitate national and international comparisons?

With the adoption of the CCSS standards and the impact of the Race to the Top initiative, it is likely that states will no longer report the results on their state tests in isolation. Educational reformists are pushing for states to benchmark their students’ performance against other states and countries. Increased popularity of the National Assessment of Educational Progress (NAEP) results illustrates the nation’s intense interest with cross-state comparisons. NAEP and the National Center for Educational Statistics (NCES) coordinated their international assessment schedules so they report results at about the same time (National Assessment of Educational Progress [NAEP], 2010). In addition, the recent and planned studies to compare states and countries on reading and mathematics highlight the importance of international comparisons. Educational organizations are not the only ones pushing for national and international comparisons. The federal government is promoting this trend as well. In a March 20, 2009, speech to the National Science Teachers Association, Arne Duncan stated, “A nation that does not benchmark its standards against the highest international standards is crippling its children in the competition for jobs.” (United States Department of Education [USDE], 2009).

States promoting national and international comparisons in English language arts will want to compare score results (e.g., percentages in performance levels, percentiles) on their state assessments to those on national or international assessments. They might also want
Several aspects specific to the English language arts assessments will require consideration in making national and international comparisons. First, states will want to review all of the content standards and identify those that will help promote national and international comparisons. They will also want to flag those content standards that will challenge comparisons across states and nations. For example, if states include text or passages that are state or nation specific, that part of the score derived from that state or nation-specific text will impede interpretations. For example, the CCSS in ELA require the inclusion of content such as “foundational U.S. documents, seminal works of American literature, and the writings of Shakespeare” (CCSSO & NGA, 2010, p. 1). When state assessment scores come from content of this type, states will need to consider how to interpret scores in state and national comparisons.

A second aspect of English language arts content standards and assessments that directly relates to national and international score comparisons is the inclusion of language that may not be used across the comparison units. For example, the Reading CCSS for grades 6-8 in Informational Text lists under Craft and Structure that students need to determine the meaning of words and phrases as they are used in text, including “figurative, connotative, and technical meanings” (CCSSO & NGA, June 2010, p. 39). States will want to carefully consider how including this skill in the assessments of the CCSS will influence score interpretations that are intended for comparing student performance in the state with that in the nation and world. For example, idiomatic phrases specific to a particular region of the U.S. may pose challenges for students in states not within that region or for those students who are not native speakers of English. Similarly, students from a particular region of the U.S. may not be familiar with technical terms that are familiar to students living in another region. Students living in southern states, for instance, may not understand weather-related technical terms that students in northern states may know.

The way in which states incorporate content covering cultural issues specific to the state or nation is a third aspect of the English language arts content standards and assessments that affects national and international score comparisons. States will need to balance the goals of increasing the authenticity of the content included on assessments and of increasing student engagement through relevant content with the goals of national and international comparisons.

10. Defining the Timeline
To prepare for the possibility of assessing new standards as soon as possible, what timeline will assist the prioritization of tasks such as creating a test design, aligning passage and item banks, and commissioning or permissioning field-test material?

Those involved in adopting new standards need to answer a number of questions in the timeline for planning an operational test. Factors include (but are not limited to) the following:

- How and when will the new standards be incorporated into the curriculum/instruction?
- What will the timeline be for developing test specifications, blueprints, and reporting requirements?
- When should the decision about the number of test forms be made?
- What methods should be used to acquire stimuli for use in the ELA assessment?
- What is the best way to schedule the steps of the item development process?
- How will decisions about field-test methodology (embedded vs. stand alone) impact the timeline?

After creating a test design and blueprint, states will take the next logical step of conducting a gap analysis of both passages and items. Legacy passages may be reviewed to see if they will support items written to the new standards. Additionally, field testing will need to occur in a manner that shrinks gaps in banks without creating an assessment that does not represent the breadth of the content standards accurately. Data associated with new field-test items may reflect a learning curve, especially if innovative items are used. If so, some accommodation must be allowed to account for the “newness” of the types of items and targeted standards.

The CSS will necessitate the use of complex stimuli; thus, acquisition of these stimuli will likely be the initial test development step. For the most part, passage and stimuli development for field testing followed by item development should occur two years prior to implementation of the first operational test. The field testing and data review would take place the next year, or one year prior to operational testing. In cases where it is preferred that stimuli be developed one year before item development, passage and stimuli acquisition would need to start three years prior to the first operational administration. Table 1 illustrates an example timeline of activities. If multiple states are working together and item banks or other resources are shared, this timeline may be condensed.

Table 1
A Sample Timeline for Field Testing

<table>
<thead>
<tr>
<th>YEAR</th>
<th>ACTIVITIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>• Passage and stimuli acquisition</td>
</tr>
<tr>
<td></td>
<td>• Item development</td>
</tr>
<tr>
<td>2</td>
<td>• Field test</td>
</tr>
<tr>
<td></td>
<td>• Data review</td>
</tr>
<tr>
<td>3</td>
<td>• Operational test</td>
</tr>
</tbody>
</table>

Note: For projects that develop stimuli one year before beginning item development, the above timeline would need to be adjusted to reflect four years, with passage and stimuli acquisition the first year.

Material permissioned or commissioned to assess new standards must represent the appropriate genres as well as the appropriate depth and rigor. When looking ahead, it is worth considering whether item writers will need additional time to write items due to unfamiliarity with new standards. If so, development time must be cushioned in a way that accounts for a learning curve for all parties involved.

To build a healthy passage and item bank, a significant amount of material must be field tested. In addition, the material field tested requires exposure to a group of students large enough to provide reliable and usable data.

Special consideration must be given to the formatting of field-test items developed to new content standards. Whereas perceptive students can sometimes detect field-test items in even the most traditional assessments, students’ ability to detect field-test items becomes more likely during the transition to new assessments and standards. For embedded field tests, new items might be too distinctive to blend in with the rest of the test items, both in what they are assessing and whether innovative items are presented. Field-testing methods for embedding different or new item types should be carefully considered in the transition of
ELA assessments. States might determine that stand-alone field-testing methods are better. If stand-alone field tests are considered, however, thought must be given to the students’ motivation and effort during the administration.
References


