BRIEF #4: ALABAMA ASSESSMENT TASK FORCE

TIMELINE FOR TEST DEVELOPMENT AND ADMINISTRATION

Juan D'Brot and Scott Marion, Center for Assessment

January 12, 2018

One of the most important issues when specifying a Request for Bids (RFB) for an assessment program is determining the time line for assessment design, field testing, and operational administration. The timeline of an operational administration dictates the timing and pace of development. There are many activities that are sequential in nature (i.e., the first must be completed before the next can be started) when developing an assessment, all of which are reliant on the specified purposes and uses of the assessment. The following brief describes two key concepts in assessment design and concludes by posing questions that will need to be addressed by the Task Force.

The Role and Timing of Assessments in Relation to Standards and Instruction

When focusing on the role and uses of *summative* assessments for accountability, the Task Force recognized the backwards-looking nature of the information gleaned from the test and its potential uses (e.g., evaluate achievement, monitor progress over time, support accountability). Given these uses, it is important to understand how these types of assessments follow standards and instruction but can still be used to inform practice. As noted by the Task Force, the statewide summative assessment should be aligned to state standards to help us understand whether students are making progress against grade-level expectations. After-the-fact assessment results can be used to inform broader adjustments to curriculum that may lead to revisions in instruction on a wider scale. This reiterates the notion that large-scale assessment should be dependent on state standards and thus great efforts are taken to determine the facets of the standards that are most appropriate to assess. This process is described in more detail in the next section.

The Assessment Development Process

The assessment development process must begin with a clarification of the uses and purposes of the assessment. In the case of Alabama's state summative assessment, the assessments must provide evidence of student proficiency of grade-level standards, inform progress toward college- and career-readiness (CCR), and support student and school accountability. A detailed description of the major goals established in light of the Task Force's suggested uses is provided in the December meeting summary and will be included in the final report.

In order to appropriately determine the assessment development time line, we should consider the general steps that are necessary to develop an assessment. Those steps include, but are not necessarily limited to the following[1]--depending on the uses of the assessment:

1. Develop assessment specifications, which are based upon: the state's academic standards, detailed specifications about the learning objectives that support the standards, and the rules dictating requirements for test content, format, and accessibility for all students;
2. Develop and review assessment materials, which include item development guides, scoring rubrics, graphic design requirements, a verification of content and standard alignment, and score report requirements;
3. Conduct pilot tests, usability studies (to ensure ease of use by students and educators), tryout studies (to confirm consistent and accurate scoring if relevant), and bias and sensitivity reviews (to ensure content is validly and fairly represented for all students);
4. Conduct field tests to determine how well items are performing, that items effectively represent the content being assessed, and that items can be accessed fairly and appropriately by all students;
5. Produce final assessment materials, which include final test versions, reports for educators and students, and supporting information/data that helps contextualize test results to those consuming reports from the test such as administrative manuals and interpretative guides;
6. Administer, score, and report student performance using the final version of the tests; and
7. Engage in ongoing evaluation of the assessment system to ensure the assessment is meeting the goals of the system and to determine if any refinements or revisions to improve its quality and effectiveness are needed.

While these can be considered a general set of steps for assessment development, there may be additional or fewer steps depending on the intended uses of the assessment results. Although this brief focuses only on summative assessment, there are additional components of an assessment system that may provide a more comprehensive view of student performance and school quality (e.g., locally-selected assessments, assessments common across districts, or classroom developed assessments and formative practices). Those additional components may require all, a subset, or additional steps than those listed above.

Specifying a Timeline

We intend to provide a relatively robust view of the steps required to develop an assessment in the previous section. However, the three main chunks that we will use to describe the timeline are (1) item development, (2) field testing, and (3) operational administration.

For the purposes of this brief, **Item development** includes a review of the standards, item specifications, development guides, and ensuring alignment of the items. Item development is a

[1] Adapted from DRC|CTB (2016). Designing assessment systems: A primer on the test development process.

critical aspect of test development because test claims are dependent on the assessment specifications, item quality, and the items' abilities to represent the content. Incidentally, item development is also where timelines attempt to be reduced, which can lead to troublesome shortcuts.

Field testing includes steps three, four, and five, all of which are intended to test the test. In other words, are the items measuring what they purport to measure and can we argue that the test will validly reflect the claims we hope it can make? Field tests can either be stand-alone as their own administration or embedded into other operational administrations, which increase testing time.

Operational administration includes the remainder of the steps, which result in reportable scores and in the case of summative assessments, accountability data. This is typically when students and educators realize they are interacting with the assessment.

In order to determine when operational assessments will be administered, we must determine when the new assessment should be reported. Two sample timelines are presented below with questions presented in the next section.

	Pre-Assessment Preparation		Assessment Development				Live Assessment
Activity	RFP Release	Vendor Award	Item Dev	Embedded Field Test	Form Dev	Form Dev; Training	Operational Adminstration
Semester	Spring 2018	Summer 2018	Fall 2018	Spring 2019	Summer 2019	Fall 2019	Spring 2020
School Year	SY 2017-2018		SY 2018-2019			SY 2019-2020	

Figure 1. Low-risk assessment development timeline

	Pre-Assessment Preparation		Assessment Development	Live Assessment
Activity	RFB Release	Vendor Award	Item Dev; Stand-alone Field Test; Training	Training; Form Dev; Operational Administration
Semester	Spring 2018	Summer 2018	Fall 2018	Spring 2019
School Year	SY 2017-2018		SY 2018-2019	

Figure 2. High-risk assessment development timeline

Based on the figures above, the high-risk timeline attempts to compress two years' worth of activities into a single year, which will result in higher costs and risks.

Questions to Answer

Based on the information presented above, the Task Force should be prepared to address the following questions:

1. Is it worth adopting more risk for a faster deployment of an operational assessment? Note, with risk comes the possibility of poorly deployed administration, which can result in vendor change.
2. An administration of Spring 2020 allows for longer lead up time for both development and for training. Is the Task Force willing to recommend a lower-risk timeline in support of a more appropriate level of support?