

BRIEF #2: ALABAMA ASSESSMENT TASK FORCE

CONTENT COVERAGE WITH A FOCUS ON WRITING ON ALABAMA'S NEXT ASSESSMENT SYSTEM

Scott Marion & Juan D'Brot, Center for Assessment

January 12, 2018

Alignment to state standards was one of the most important goals articulated by the Task Force at the December 2017 meeting. Task Force members indicated that it was critical that the new state assessment accurately reflect the standards that teachers are expected to teach and students are expected to learn. However, what does alignment really mean? The standards include such things as listening, speaking, and research, but when asked, Task Force members acknowledged that there was little interest in trying to assess such learning targets on an on-demand state assessment. We use this example to make the point that all assessments require choices about what will and will not be included on any given assessment. It is beyond the scope of the Task Force to make these sorts of content decisions, but the Task Force can recommend convening content experts to produce a “test specifications” document to make such design decisions explicit. However, the Task Force can and should weigh in on how writing is assessed, which is the focus of the remainder of this brief.

Many states rely on extended constructed-response tasks as the primary way to assess writing. If we want to measure writing achievement, it makes sense to have students write. Including direct writing on state assessments has been shown to increase the amount of writing that students do in classrooms, at least in the grades where writing is assessed. Further, the newer approaches to writing tasks that call on students to frame arguments based on evidence from reading stimuli rather than the “imaginary narrative” prompts of the past can help incentivize such practices in classrooms.

However, there are some measurement challenges associated with the use of a single writing prompt. Even though student response times can range from 30-90 minutes, the score from the single writing task contributes very little test information to an overall English language arts score. Additionally, there are known challenges with the generalizability of the results from a single task. In other words, since students perform differentially on various prompts, it is hard to support valid inferences about individual student writing achievement based on a single prompt.

The solution to this problem—administering two or more writing tasks to each student—often is not practically feasible.

Given these challenges, what are some options for assessing writing in a meaningful way? First, the Task Force must decide if the priority is student- or school- level scores. Focusing on school-level information does not mean that the state is giving up on student-level scores, but it might mean that it will tolerate somewhat lower levels of student comparability in order to get more information at the school level. For example, the state could administer multiple writing tasks (at least three and probably not more than eight) at the school level with each student completing only one task. This means that students would be completing different prompts, but with writing as a component of ELA, student-level score comparability will not be much of an issue. On the other hand, the multiple prompts at the school level will produce a robust writing score at the school level that may even be able to support writing subscores (e.g., by genre).

If student-level scores are the priority, there are ways to gather more “information” than can be gathered through a single writing response. The state can choose to include a few short writing responses throughout the test in addition to the extended-response task to noticeably increase the amount of writing information generated by the test. Some states and consortia have explored measuring both reading and writing with these shorter writing tasks and while this sounds intuitively sensible, it has proved challenging because of phenomena such as “halo effects” (e.g., students getting the same score on both reading and writing).

The Task Force must consider how often it would like writing assessed. Most states assess writing as part of its state assessment in grades 3-8 and at least once in high schools. However, some states have scaled back writing to once per grade span and other states that claim to assess reading rather than ELA standards may not assess writing at all. As noted above, there is fairly strong evidence that including direct writing on the state assessment has demonstrated positive consequences for teaching and learning of writing. The RFB will have to specify whether writing will be part of an overall ELA score or it will have its own stand-alone score that may be combined with reading in a post-hoc composite. If writing is not assessed in every grade, the post-hoc composite is likely the more sensible approach.

Finally, the RFB will have to indicate whether the contractor should bid automated scoring or require human scoring only. The artificial intelligence (AI) software has improved considerably over the years and, for many types of prompts, AI scoring is as or more reliable as human scoring and it is much more cost effective depending on the number of student papers being scored.

Questions to Answer

We would like the Task Force to weigh in on the following questions:

1. Does the Task Force want to require a writing assessment as part of the RFP? If so...
2. Should the RFB ask for writing to be included in all grades where reading is assessed?
3. Does the Task Force want to prioritize school or student level information?
4. Does the Task Force want to explore ways to generate more writing points than can be yielded from a single prompt?
5. Does the Task Force want to encourage bidders to propose artificial intelligence approaches for scoring writing?
6. Does the Task Force want to recommend convening an assessment specifications work group for each of the content areas assessed?