

MEMORANDUM

To: Gloria Zyskowski

From: ETS Texas STAAR® Psychometric Services Team

cc: Frederick McHale
George Powell

Date: May 6, 2016

Re: A Report on Testing Time Collected During March 2016 Test Administration

Texas House Bill 743 (HB743) added §39.023 (a-12) to Education Code specifies the following requirements for STAAR grades 3–8 assessments:

1. If administered to students in grades three through five (i.e., lower grades), 85 percent of students will be able to complete the assessment instrument within 120 minutes (2 hours); and
2. If administered to students in grades six through eight (i.e., higher grades), 85 percent of students will be able to complete the assessment instrument within 180 minutes (3 hours).

The Texas Education Agency (TEA) plans to implement these requirements in the 2016–2017 school year. In this memorandum we summarize test data from the STAAR March 2016 administration (Grade 5 and 8 Reading and Mathematics; Grade 4 and 7 Writing) that might help the TEA in evaluating whether the requirements specified in HB743 are being currently met for the Grade 3-8 tests. In the last two sections, we present some alternatives in case the above requirements are not met for the data we analyzed.

Results from Paper and Pencil Data

For students taking paper tests, test coordinators were instructed to grid in the number of minutes (in 15 minute increments) each student took to finish the test. If for this data, we observe most students (e.g., 85 %) complete the test within 120 minutes for grades 3-5 and 180 minutes for grades 6-8, then that would support the requirements specified by HB743. However as observed in the data, only a small percentage (about 17% for Grade 5 Reading for example) completed the test in 2 hours for the lower grades. Findings were more positive for the higher grade tests where approximately 70-80 % percent of students completed the tests in 3 hours. The percentages for all March tests are shown in Table 1. These percentages do not support the requirement (i.e., 85% of students completing the test in 2 hours) for the lower grades. However, the percentages for the higher grades are much closer to the target of 85% of students completing the test in 3 hours, thereby lending support to meeting the HB743 requirement.

Note that the utility of this data depends on how accurately the test coordinators recorded the time taken by each student to finish the test. For example, if test coordinators were not gridding in the time accurately or were simply gridding something after collecting all the papers, then the data would not

provide much useful information. Since this is a possibility, we filtered out data where 90% of students in a campus were marked by the test coordinator to have finished the test in a particular time slot (e.g., 2 hours, 3 hours, longer than 3 hours, etc.). By doing so, we filtered out 14,637 of a total of 347,088 cases. Figure 1, shows an example of a 90% condition which was filtered out.

Results from Online Data

For students taking online tests (STAAR Accommodated and Linguistically Accommodated), testing time was collected as the actual number of minutes that student spent on answering all items before they submit their tests. If students spent additional time reviewing the items, their item review time was not available at the time of this report. Since testing time was recorded electronically when the student completed testing, this data is viewed as more accurate than the manually gridded time recorded by the test administrator in paper and pencil testing. As seen in Table 1, the percentages of students who completed the online tests in 2 hours were higher compared to the paper and pencil condition for the lower grades. For example, 50% of students completed the Grade 5 Reading test in 2 hours compared to only about 17% in the paper and pencil condition. As in the paper and pencil condition, findings were more positive for the higher grade tests where approximately 81% to 96% of students completed the tests in 3 hours. As stated earlier, these percentages do not support the requirement specified in HB743 (i.e., 85% of students completing the test in 2 hours) for the lower grades. However, the percentages for the higher grades are much higher and lend support to HB743 requirements.

The results from the paper and pencil and online data suggested that HB743 requirement for the higher grades might be supported. However, the results did not support the requirement for the lower grades. More data will be collected during May administration of all Grade 3-8 tests. However, the results may not be drastically different than what were observed in the March administration. Unless TEA plans to repeal HB743, we present in the next sections some alternatives for TEA consideration that might help with meeting the requirement for the lower grades.

Shorten Tests Further

In spring 2016, all grade 3-8 tests were shortened by excluding field test items. Additionally, Grade 4 and 7 Writing tests were shortened from two-day to one-day testing. One way of trying to meet HB743 requirements for the lower grade tests is to further shorten the tests so that more students might be able to complete the tests within 2 hours. However, we need to make sure that shortening the tests does not cause any drastic reduction in test reliability and that the shortened tests can still meet the need of assessing content standards. In this section we discuss shortening the base tests only in the context of overall test reliability. Content specialists would need to determine if current content blueprints can still be met with the proposed test lengths and if the current reporting categories can still be supported. We used the Spearman Brown formula (Brown, 1910; Spearman, 1910) to propose test lengths while still maintaining a test reliability of around 0.85 (and 0.80 or higher for writing; Grade 4 Writing test need to be longer). Note that the proposed test lengths provided in Table 2 are slightly longer than what would be required to maintain a test reliability of 0.85 or higher (except for Writing). We added a few extra items to provide our content specialists more flexibility for covering all readiness standards and a few supporting standards each year. We have also added a column proposing the number of field test items per test form. This number was based on the number of items that need to be

field tested to maintain a healthy item pool. We used 10 times of the base test length to recommend the number of field test items per test form. For example, if there would be 34 items for Grade 3 Reading, we would need to field test 340 items (34 X 10). Grade 3 testing population can support at least 60 field test forms, we can field test 6 items per form to get 360 field test items, which is slightly over our target of 340 items.

Reduce Allowed Testing Time without Further Shortening the Tests

Could reducing testing time lead to more students completing the test in 2 hours? We don't know the answer. The current data were collected under the condition where students might not have motivation to complete the test sooner. It is reasonable to assume that they might have paced themselves slowly and used up all the allowed time. Therefore, the current estimates on testing time needed for students to complete the test might be somewhat conservative and further shortening the tests (as suggested in the first alternative) might only result in students pacing the tests even slower. The current data which shows that only a small percentage of students are completing the test in 2 hours might be a consequence of having more time (i.e., 4 hours) for completing the test. We looked at testing time offered for K-12 tests in other states. The number of minutes allowed per item was calculated by dividing the testing time allowed by the number of items in the respective tests. As seen in Appendix A, STAAR offers substantially more testing time per item compared to the other states. The only other state that offers similar or slightly more testing time per item as Texas is New Hampshire. These two states offer approximately 4-6 minutes per item compared to several other states which offer much lower testing time per item, with some states offering as low as 1-1.5 minutes per item. If other states are not offering lengthy testing times, it might be reasonable for Texas to reduce testing time as well. To provide additional evidence that students probably don't need 4-6 minutes per item, we examined the online data to find out on how many minutes on average students took to respond to an item. The online data showed that students, on average, took about 3 minutes to answer a multiple-choice or griddable item and as low as 2 minutes for tests such as Grade 7 Writing. These data show that although we offer more time per item, students don't necessarily need all of that time (see Table 3).

References

- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3, 296–322.
- Spearman, C. C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271–295.

| | Percentages of Students Completing Testing within | | | Average Break Time (in Minutes) |
|----------------------------|---|----------------------|---------|------------------------------------|
| | 2 hours | 2 hours & 15 minutes | 3 hours | |
| Grade 4 Writing | | | | |
| STAAR Paper | 26 | 38 | 72 | 9 |
| STAAR A Online | 75 | 82 | 97 | |
| Grade 5 Reading | | | | |
| STAAR Paper | 17 | 28 | 65 | 8 |
| STAAR A Online | 50 | 61 | 83 | |
| Grade 5 Mathematics | | | | |
| STAAR Paper | 27 | 38 | 70 | 8 |
| STAAR A Online | 45 | 56 | 82 | |
| STAAR L Online | 47 | 59 | 84 | |
| Grade 7 Writing | | | | |
| STAAR Paper | 36 | 50 | 82 | 7 |
| STAAR A Online | 75 | 83 | 96 | |
| Grade 8 Reading | | | | |
| STAAR Paper | 25 | 37 | 73 | 7 |
| STAAR A Online | 48 | 59 | 83 | |
| Grade 8 Mathematics | | | | |
| STAAR Paper | 42 | 54 | 82 | 7 |
| STAAR A Online | 47 | 59 | 85 | |
| STAAR L Online | 36 | 48 | 81 | |

Table 1. *Cumulative Distribution of Student Testing Time Gridded by Test Administrators and through Online Testing*

| Test | | Number of Base Items (2016 Test Length) | Number of Base Items (Alternative) | Field Test Items per Form (60 Forms) | Alternative Test Length |
|----------------|---|--|---------------------------------------|---|----------------------------|
| Reading | 3 | 40 | 34 | 6 | 40 |
| | 4 | 44 | 36 | 6 | 42 |
| | 5 | 46 | 38 | 6 | 44 |
| | 6 | 48 | 40 | 7 | 47 |
| | 7 | 50 | 42 | 7 | 49 |
| | 8 | 52 | 44 | 7 | 51 |
| Mathematics | 3 | 46 | 32 | 5 | 37 |
| | 4 | 48 | 34 | 6 | 40 |
| | 5 | 50 | 36 | 6 | 42 |
| | 6 | 52 | 38 | 6 | 44 |
| | 7 | 54 | 40 | 7 | 47 |
| | 8 | 56 | 42 | 7 | 49 |
| Writing | 4 | 18 MC + Essay | 30 MC + Essay | 5 | 36 |
| | 7 | 30 MC + Essay | 30 MC + Essay | 5 | 36 |
| Science | 5 | 44 | 36 | 6 | 42 |
| | 8 | 54 | 42 | 7 | 49 |
| Social Studies | 8 | 52 | 44 | 7 | 51 |

Table 2. Suggested Test Lengths based on Maintaining Test Reliability in the Mid to High 80s

| Test | | Number of Items (2016) | Allowed Minutes Per Item | Actual Minutes used Per Item (A Form) | Actual Minutes used Per Item (L Form) |
|-------------|---|---------------------------|-----------------------------|--|--|
| Reading | 5 | 46 | 5.2 | 2.8 | N/A |
| | 8 | 52 | 4.6 | 2.5 | N/A |
| Mathematics | 5 | 50 | 4.8 | 2.6 | 2.6 |
| | 8 | 56 | 4.3 | 2.3 | 2.5 |
| Writing | 4 | 18 MC + Essay | N/A | 2.6 | N/A |
| | 7 | 30 MC + Essay | N/A | 2.0 | N/A |

Table 3. Allowed and Actual Minutes Per Item

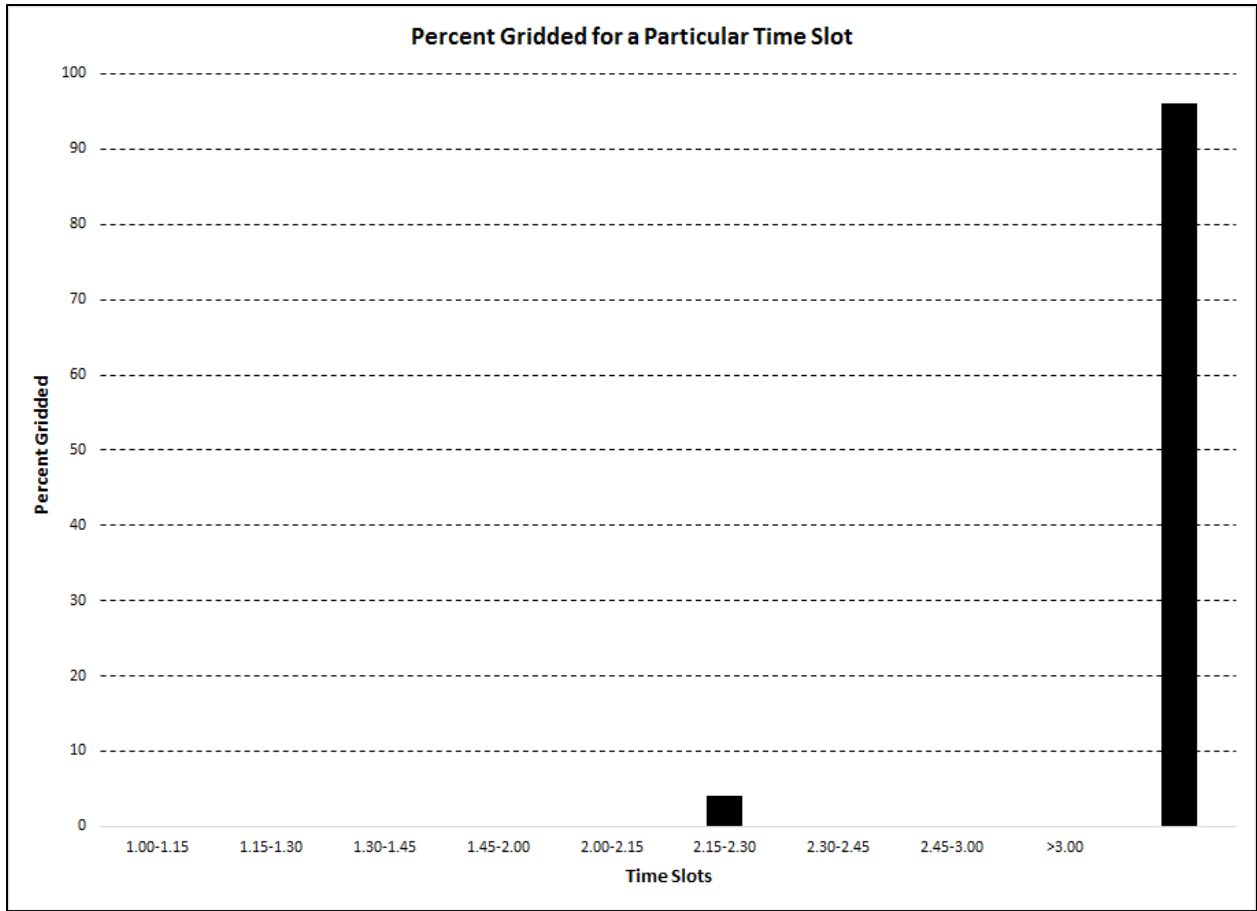


Figure 1. *Percent Gridded for a Particular Time Slot (90% Filtered Condition)*

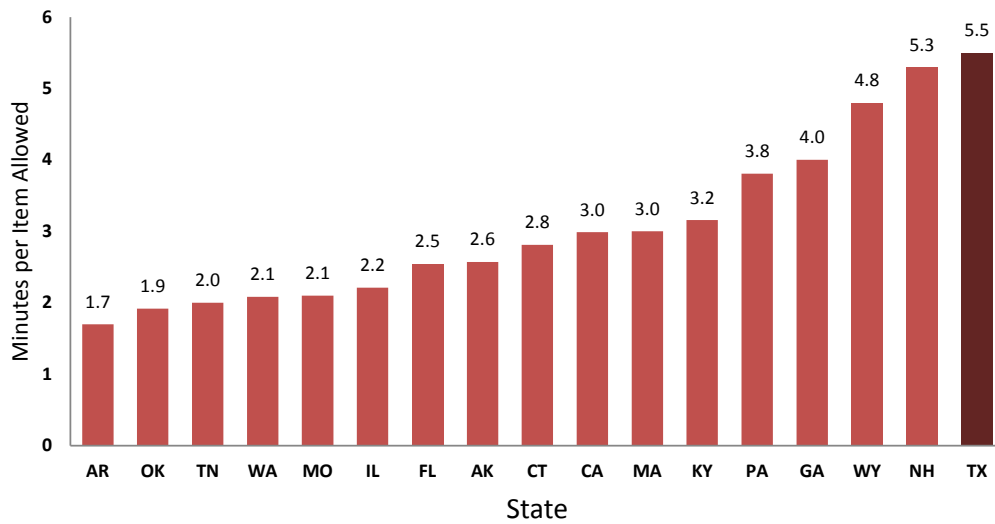
Appendix A

Comparison of Testing Time per Item Allowed in State Assessments

Grade 3 Reading (ELA)



Grade 4 Reading (ELA)



Grade 5 Reading (ELA)



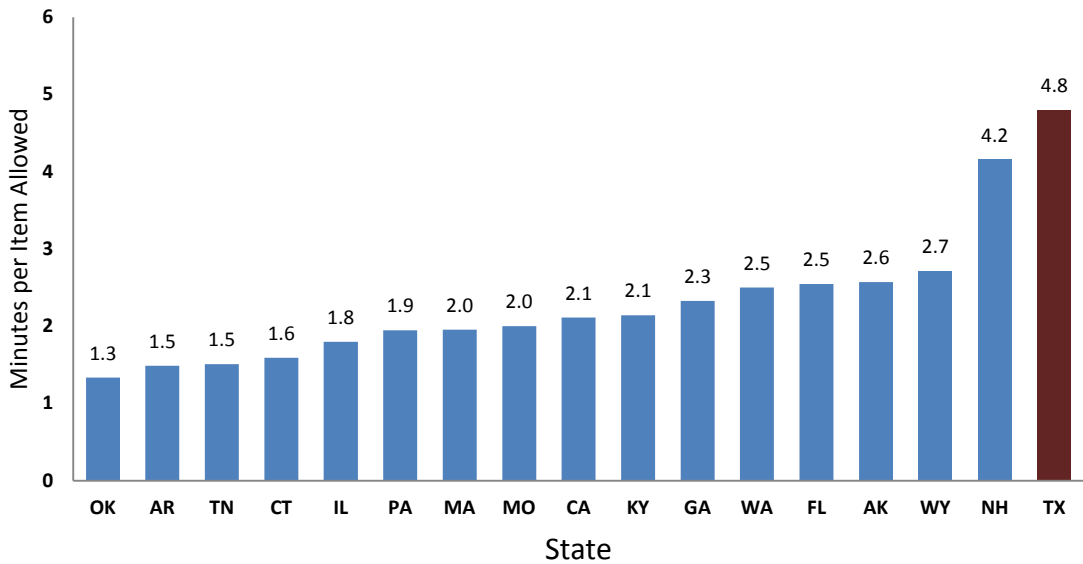
Grade 3 Mathematics



Grade 4 Mathematics



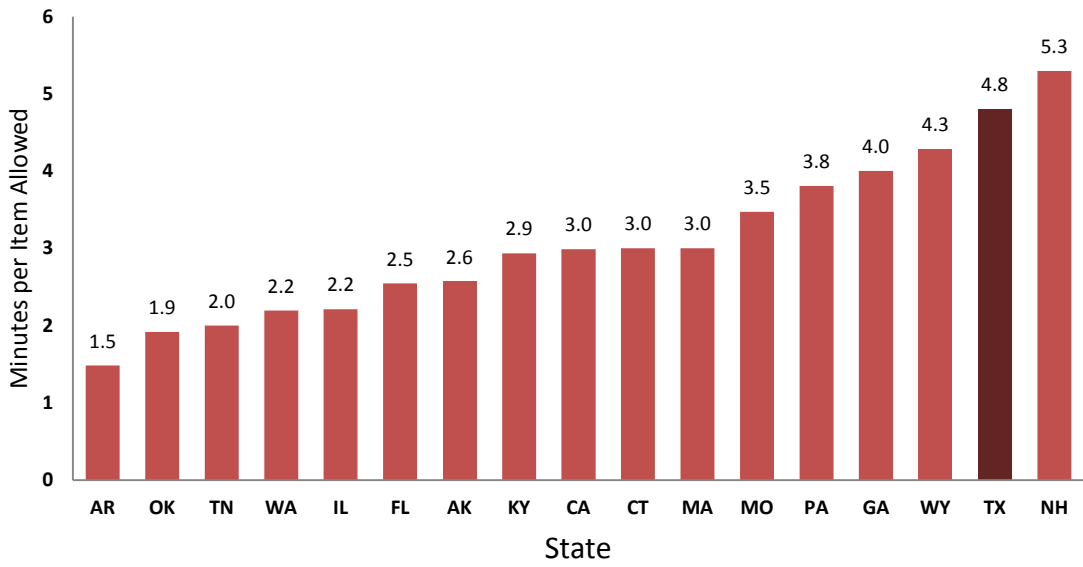
Grade 5 Mathematics



Grade 6 Reading (ELA)



Grade 7 Reading (ELA)



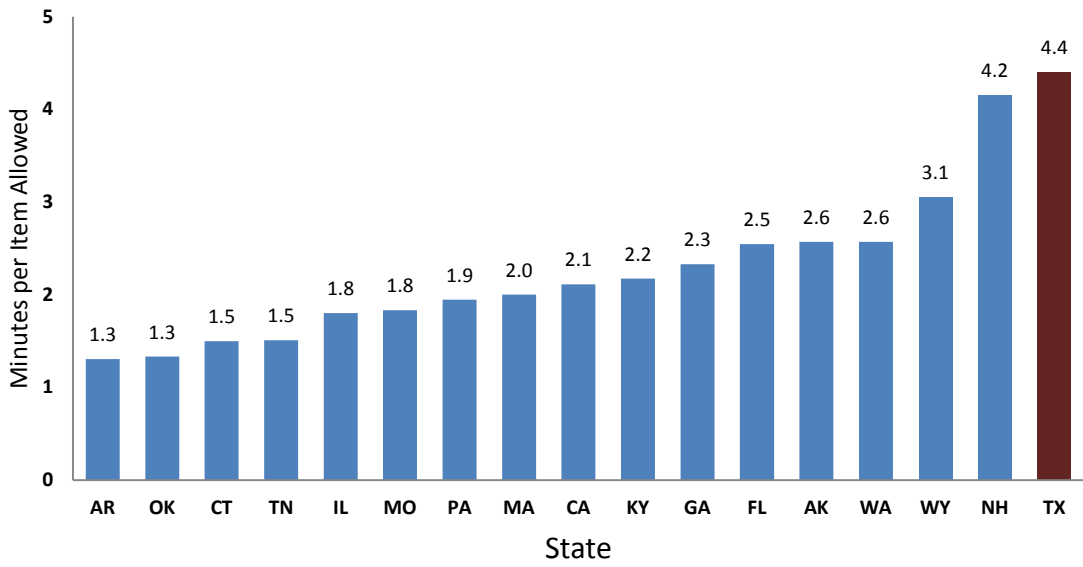
Grade 8 Reading (ELA)



Grade 6 Mathematics



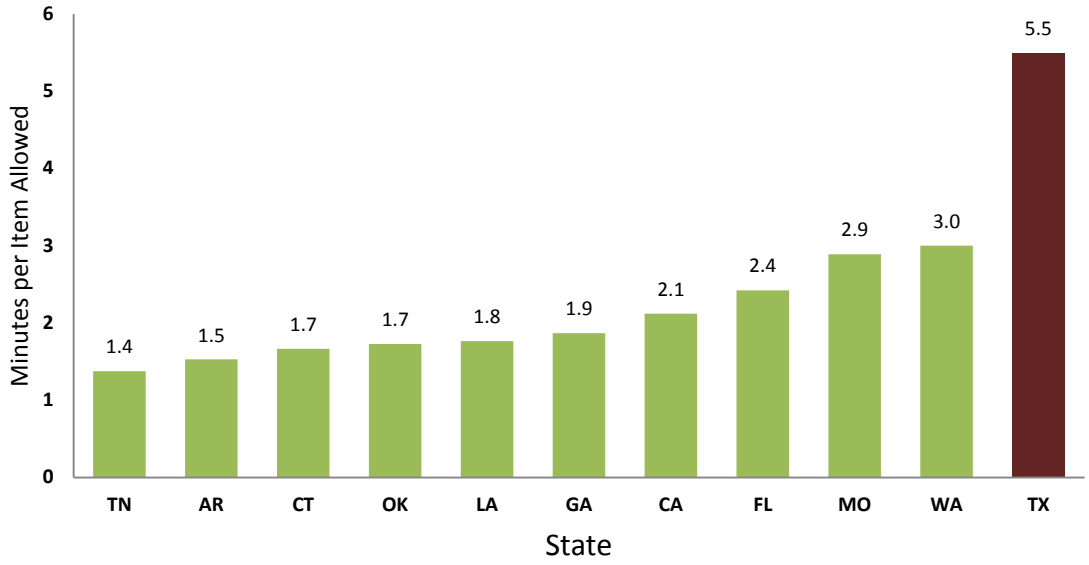
Grade 7 Mathematics



Grade 8 Mathematics



Grade 5 Science



Grade 8 Science



Grade 8 Social Studies

