**** ETS Sacramento Office
 1600 K Street, Suite 4A
 Sacramento, CA 95814
 Phone: 916-403-2402
 Fax: 916-403-2462

Measuring the Power of Learning.™ www.ets.org

# Memorandum

 **Date:** February 21, 2018

 **To:** John Boivin

 California Department of Education (CDE)

 **From:** Dan Stanley

 Educational Testing Service (ETS)

 **Cc:** CDE: Michelle Center, Mao Vang, Eric Zilbert, Kelly Bacher

 ETS: Tom Foster, Mary Anne Arcilla, Fred McHale, Terran Brown, Stephanie Herrera, Helen McMahon, Patti Baron, Yi Du

 **Subject:** ETS’s Response to the Human Resources Research Organization’s (HumRRO’s) Independent Evaluation Recommendations from the California Alternate Assessments (CAAs) for English Language Arts/‌Literacy (ELA) and Mathematics Alignment Study

HumRRO’s Independent Evaluation offers the following four recommendations from the CAAs for ELA and Mathematics Alignment Study. ETS’s responses to each of the recommendations are included in this memorandum.

Questions about this response should be addressed to Dan Stanley by telephone at 916-403-2428 or by e-mail at dstanley@ets.org.

## HumRRO’s Recommendation 1

**Review the mathematics items in the grade three and eleven assessments for potential barriers to disability groups.** Panelists indicated that many grade three and eleven items exhibited barriers, particularly to students with visual and/or hearing impairments. It is important to note that assessment items were developed to include an alternate presentation allowing students to access the content of the item without modifying the construct of the item. However, panelists’ comments for some items and the assessment overall stated that the alternate text did not provide enough information to answer the item without being able to view graphic stimuli or response options. We recommend CDE content experts working with the testing contractor review alternate text for mathematics items in grade three and eleven on the versions reviewed in this alignment study as well as the versions developed for the 2017-2018 school year to ensure the alternate text is sufficient, particularly for students with visual and/or hearing impairments, to access the content of the item without changing the construct.

## ETS’s Response to HumRRO’s Recommendation 1

All items developed for CAA assessments utilize alternative text (alt text) to describe images for students with visual impairments. The alt text is an integral part of an item from the beginning of the development process, and an item is considered incomplete if it does not have alt text.

The alt text is reviewed and edited alongside the item during the ETS internal Content and Editorial review rounds, and California educators provide feedback on alt text during their review of items. Additionally, the ETS Alternative Test Formats group provides a final review of the alt text and provides feedback to the content team to ensure the alt text is as accessible as possible.

The 2015–16 administration was the initial administration of the CAAs for ELA and mathematics, and was exclusively composed of new items without prior statistics. As recommended by HumRRO, ETS reviewed items and alt text from the 2015–16 transitional year and found several items in grades three and eleven with alt text that could be revised. Specifically, ETS found that

* some items had alt text that should have provided more information for students,
* some alt text contained language that was not appropriate for the grade level, and
* several grade eleven items containing tables and graphs had alt text with long descriptions, which may have been confusing for students to follow.

ETS has made improvements in alt text development since the 2015–16 transitional year, drawing on item statistical data and feedback from California educators and the CDE. Improvements to items developed for the 2016–17 and 2017–18 administrations include

* the incorporation of simpler language in both the item text as well as the alt text,
* modified internal review checklists to ensure the alt text provides enough information for students with visual impairments to respond to the item, and
* reduction in the complexity of graphs and tables by limiting the number of data points allowed.

These measures have been captured in item specification documents that guide current and future item development. In addition, ETS will review alt text of all items in grades three through eight and grade eleven selected for placement on operational forms to ensure conformance with the guidelines as described previously. Revisions to alt text will be applied as needed. Finally, beginning with the 2018–19 administration, ETS will ensure that all items appearing on operational test forms have alt text that conforms with the guidelines as stated previously.

Although a review of item alt text is already a part of the Item Review Committee (IRC) review conducted by participating California educators, ETS believes that this review can be given added emphasis during IRC meetings and will collaborate with the CDE to structure this emphasis in the next round of meetings. ETS strives for a balance between the need to sufficiently describe graphics in the alt text while keeping the language complexity low, and will use the alignment study feedback for continued improvement. And, although these analyses are specific to mathematics, any findings will be shared and applied across content areas.

## HumRRO’s Recommendation 2

**Review the DOK match between items and CCCs for grade eleven ELA.** Panelists rated the grade eleven ELA items at lower DOK levels in comparison to the DOK level of the corresponding CCC. We recommend CDE work with the testing contractor to review the grade eleven items on the versions reviewed in this alignment study as well as the versions developed for the 2017-2018 school year to ensure items are written to a range of DOK levels to reflect the content depth indicated by the CCCs. We suggest CDE obtain the number of items at each DOK level in the bank from the testing contractor to determine whether item development goals need to be amended to include more item development at certain DOK levels.

## ETS’s Response to HumRRO’s Recommendation 2

ETS reviewed grade eleven items from the 2015–16 transitional administration used for the alignment study, as well as items from the 2017–18 administration. Depth of Knowledge (DOK) levels for these items can be provided to the CDE upon request. Additional considerations include the following:

1. The report indicated that alignment panelists “provided DOK ratings for each CAA test item. This allowed for direct comparisons of item cognitive complexity to that of the matched Core Content Connectors (CCCs) and allowed determination of consistency between item and CCC DOKs.” DOK levels for the CAAs are assigned to the test items for all grades. ETS ensures there is a range of DOK levels on each test by methodically developing items with a range of DOK levels during item development.
2. ETS will evaluate the items for grades three through eight and grade eleven in the bank in light of the DOK analysis from the HumRRO report. The items were designed to be aligned to the CCCs, and the design considered the EUs as well, in order to include the appropriate DOK levels and grade appropriateness of the items. However, ETS found no evidence that the EUs were included as part of the alignment study. ETS believes that if the EUs had been included as part of the analysis by the alignment panel, the judgments by the panelists would have included evidence of the lower range of DOK levels. ETS will review the item bank and review the forms to ensure a range of DOK levels. The goals will be to identify any gaps and develop the appropriate items to remove the gaps.

## DOK Comparison Chart

The following is a comparison chart of the HumRRO and the CAA DOK scales.

| HumRRO Tasks | HumRRO DOK | ETS CAA DOK |
| --- | --- | --- |
| Attention | DOK 1 | N/A \* |
| Memorize/Recall | DOK 2 | DOK 1 |
| Performance | DOK 3 | DOK 1–DOK 3 |
| Comprehension | DOK 4 | DOK 1–DOK 3 |
| Application | DOK 5 | DOK 3 |
| Analysis, Synthesis, Evaluation | DOK 6 | DOK 3 |

\* *Attention* is not indicated in Norman Webb’s Depth of Knowledge levels.

Performance and Comprehension are shown as DOK 1–DOK 3 in the DOK scale used on the CAA and are based on the premise that performance and comprehension can be demonstrated at different DOK levels.

For example:

* Performance and comprehension at DOK 1 can be shown by calculating, listing, or labeling.
* Performance and comprehension at DOK 2 can be demonstrated by categorizing or comparing.
* Performance and comprehension at DOK 3 can be shown by differentiating or drawing conclusions.

## Confirming DOK Levels

DOKs are reviewed internally at ETS through a standard process that includes at least three content reviews. DOK checks are tracked using Item Review Checklists. DOKs for the 2015–16 administration were not reviewed as part of the IRC Meetings with California educators for that year, nor have they been reviewed at any annual educator meeting except for the pilot year. ETS suggests a further process refinement by adding an item-by-item DOK review with California educators as part of future IRC meetings.

Figure 1 shows an excerpt of a CAA ELA comment sheet that was used by California educators in Item Review Meetings for the 2015–16 administration. Its first column denotes the grade. Its second column denotes the number in the sequence in which the item will be presented to the student. Its third column denotes the “accnum,” which is the identification number used for a specific item. The fourth column denotes tier, which denotes the complexity of an item. The fifth column denotes the Core Content Connector to which the item aligns. The sixth and final column denotes the short description of the item. It lacks a *DOK* column, indicating that depth of knowledge was not evaluated by educators in 2015–16.



Figure . A portion of the ETS CAA ELA comment sheet without a DOK column

ETS suggests a further process refinement by adding an item-by-item DOK review with California educators as part of future IRC meetings.

## HumRRO’s Recommendation 3

**Review the percent of items per domain on the assessment and compare to the percent of items per domain in the CAA Blueprint.** In this alignment study, panelists were tasked with assigning a CCC that best matches the content of the assessment items. Particularly in the ELA grades, panelists’ assignment of CCC to item resulted in the percent of items per domain varying from the percent of item targets listed in the CAA Blueprint. We recommend reviewing the percent of items per domain based on the CCC assigned during item development and the percent of items per domain in the CAA Blueprint to determine if they are comparable (differ by less than 10%). If they are comparable, we recommend additional professional development for teachers on the CCCs for ELA, especially regarding how a reading item can primarily measure writing content. If the comparison is not comparable (differ by 10 percent or more), then additional target checks may need to be implemented during test construction to ensure CAA Blueprint targets are met.

## ETS’s Response to HumRRO’s Recommendation 3

The absolute percentage differences of items per domain between the forms and blueprints are summarized in Table 1 for the 2015–16 transitional administration, in Table 2 for the 2016–17 administration, and in Table 3 for the 2017–18 administration. Overall, the blueprint alignment for the 2015–16 administration is good—differences are less than 10 percent—with some exceptions in ELA 05, ELA 06, ELA 08, and ELA 11 (bolded in Table 1, Table 2, and Table 3). The 2015–16 administration was a transitional year with a limited pool of items for constructing test forms. Subsequent item development efforts have alleviated these item bank constraints.

Table 1. Absolute Percentage Differences of Items per Domain for 2015–16 Administration

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 2015–16  | R1 + Easy | R1 + Moderate | R1 + Hard | R2 + Easy | R2 + Moderate | R2 + Hard |
| ELA 3 | 0%~3% | 0%~5% | 0%~3% | 0%~3% | 0%~5% | 0%~3% |
| ELA 4 | 0%~2% | 0%~4% | 0%~4% | 0%~2% | 0%~4% | 0%~4% |
| ELA 5 | 0%~3% | 0%~3% | 0%~11% | 0%~3% | 0%~3% | 0%~7% |
| ELA 6 | 0%~19% | 0%~8% | 3%~12% | 0%~12% | 0%~3% | 2%~4% |
| ELA 7 | 0%~9% | 1%~3% | 1%~3% | 0%~9% | 1%~3% | 1%~3% |
| ELA 8 | 1%~8% | 1%~10% | 1%~8% | 1%~8% | 1%~8% | 1%~8% |
| ELA 11 | 1%~11% | 1%~15% | 1%~11% | 3%~8% | 3%~11% | 3%~8% |
| Mathematics 3 | 2%~3% | 2%~3% | 2%~3% | 0%~2% | 0%~2% | 0%~2% |
| Mathematics 4 | 0%~3% | 0%~3% | 0%~3% | 1%~4% | 1%~4% | 1%~4% |
| Mathematics 5 | 0%~4% | 0%~4% | 0%~4% | 0%~4% | 0%~4% | 0%~4% |
| Mathematics 6 | 1%~5% | 1%~5% | 1%~5% | 1%~8% | 1%~8% | 1%~8% |
| Mathematics 7 | 2%~9% | 2%~9% | 2%~9% | 0%~7% | 0%~7% | 0%~7% |
| Mathematics 8 | 0%~6% | 0%~6% | 0%~6% | 2%~3% | 2%~3% | 2%~3% |
| Mathematics 11 | 1%~9% | 1%~9% | 1%~9% | 1%~6% | 1%~6% | 1%~6% |

As shown in Table 2, the percent of items per domain based on the CCC assigned during item development and those in the CAA blueprint are comparable—differences are less than 10 percent, with only one exception of 10 percent, at grade three—for all grades in the 2016–‍17 administration. For the “Router + Hard” form of the CAA for ELA in grade three, there was a need to swap an item passage set that resulted in the overall form not adhering as closely to the blueprint as other forms.

Table 2. Absolute Percentage Differences of Items per Domain for the 2016–17 Administration

|  |  |  |  |
| --- | --- | --- | --- |
| 2016–17 | Router + Easy | Router + Moderate | Router + Hard |
| ELA 3 | 1%~6% | 1%~3% | 1%~**10%** |
| ELA 4 | 1%~3% | 1%~3% | 1%~2% |
| ELA 5 | 2%~2% | 2%~2% | 2%~2% |
| ELA 6 | 0%~6% | 0%~2% | 0%~6% |
| ELA 7 | 0%~2% | 0%~2% | 0%~2% |
| ELA 8 | 0%~2% | 0%~2% | 0%~2% |
| ELA 11 | 1%~2% | 1%~2% | 1%~2% |
| Mathematics 3 | 0%~2% | 0%~2% | 0%~2% |
| Mathematics 4 | 1%~3% | 1%~3% | 1%~3% |
| Mathematics 5 | 0%~2% | 0%~2% | 0%~2% |
| Mathematics 6 | 0%~2% | 0%~2% | 0%~2% |
| Mathematics 7 | 0%~3% | 0%~5% | 0%~3% |
| Mathematics 8 | 1%~2% | 1%~2% | 1%~2% |
| Mathematics 11 | 0%~2% | 0%~2% | 0%~2% |

As shown in Table 3, the percent of items per domain based on the CCC assigned during item development and those in the CAA blueprint are also comparable—differences are less than 10 percent, with only one exception of 10 percent—for all grades in the 2017–18 administration. For the “Router + Easy” form of the CAA for ELA in grade four, a shortage of available items meeting both blueprint and other test form specifications led to a difference of 10 percent from the specified blueprint.

Table 3. Absolute Percentage Differences of Items per Domain for the 2017–18 Administration

|  |  |  |  |
| --- | --- | --- | --- |
| 2017–18 | Router + Easy | Router + Moderate | Router + Hard |
| ELA 3 | 1%~2% | 1%~6% | 1%~2% |
| ELA 4 | 3%~**10%** | 1%~6% | 1%~6% |
| ELA 5 | 2%~2% | 2%~2% | 2%~2% |
| ELA 6 | 0%~2% | 0%~2% | 0%~2% |
| ELA 7 | 0%~2% | 0%~6% | 2%~6% |
| ELA 8 | 0%~2% | 0%~2% | 0%~2% |
| ELA 11 | 1%~2% | 1%~3% | 1%~2% |
| Mathematics 3 | 0%~2% | 0%~2% | 0%~2% |
| Mathematics 4 | 0%~2% | 0%~3% | 0%~3% |
| Mathematics 5 | 0%~2% | 0%~2% | 0%~2% |
| Mathematics 6 | 0%~2% | 0%~2% | 0%~2% |
| Mathematics 7 | 0%~3% | 0%~3% | 0%~3% |
| Mathematics 8 | 0%~2% | 0%~5% | 0%~2% |
| Mathematics 11 | 0%~3% | 0%~3% | 0%~3% |

Note that the CAA test design allows for struggling students to end testing early. Early exit cases were excluded from this blueprint analysis.

More details with regard to blueprint alignment can be found in the technical reports for the 2015–16 (CDE, 2017) and 2016–17 (report pending) administrations. The technical report for the 2017–18 administration that includes blueprint alignment details will be available in 2019.

## HumRRO’s Recommendation 4

**Request that the testing contractor include version- and tier-level reliability information in the CAA Technical Report.** While not directly evaluated for this study, ETS references information related to test reliability from psychometric analyses found in the CAA Technical Report***[[1]](#footnote-1)***. This information supports alignment conclusions related to test reporting and is a more reasonable indicator of reliability than item counts. The reported reliability estimates in the technical report are strong, ranging from 0.74 to 0.89 (alpha coefficients). However, the estimates are aggregated across versions and tiers. We recommend expanding this information to ensure that reliability estimates are not substantially different by version or tier.

## ETS’s Response to HumRRO’s Recommendation 4

## Version Reliability

ETS does not recommend providing reliability by tier due to the restriction of range associated with routing in this multistage test (MST). No ETS-developed assessment that uses an MST design provides this type of reliability. An explanation of this recommendation follows.

Because one of the two available router tests is chosen randomly for each student, it is reasonable to question whether this choice materially affects the quality of measurement. As such, the reliability computations were applied to each of two test versions, each defined by one of the routing tests being paired with each of the three targeted second-stage item blocks. Differences in reliability across these two versions is due solely to the performance of the different items in the two routing tests. Marginal reliability proposed by Green et al. (1984) was calculated, which is an appropriate reliability estimate in the context of multistage testing. As expected, differences in reliability were small, as shown in Table 4.

Table 4. Marginal Version Reliability

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test | Version | N | Mean of CSEM2 | Reliability | SEM of θ |
| ELA 3 | 1 | 1,956 | 0.18 | 0.82 | 0.54 |
| ELA 3 | 2 | 2,057 | 0.19 | 0.81 | 0.56 |
| ELA 4 | 1 | 2,131 | 0.16 | 0.84 | 0.42 |
| ELA 4 | 2 | 2,180 | 0.15 | 0.85 | 0.39 |
| ELA 5 | 1 | 1,971 | 0.13 | 0.87 | 0.33 |
| ELA 5 | 2 | 2,248 | 0.15 | 0.85 | 0.36 |
| ELA 6 | 1 | 2,127 | 0.14 | 0.86 | 0.36 |
| ELA 6 | 2 | 2,116 | 0.16 | 0.84 | 0.41 |
| ELA 7 | 1 | 2,129 | 0.15 | 0.85 | 0.38 |
| ELA 7 | 2 | 2,036 | 0.16 | 0.84 | 0.40 |
| ELA 8 | 1 | 2,009 | 0.12 | 0.88 | 0.32 |
| ELA 8 | 2 | 1,858 | 0.13 | 0.87 | 0.34 |
| ELA 11 | 1 | 1,751 | 0.14 | 0.86 | 0.35 |
| ELA 11 | 2 | 1,897 | 0.14 | 0.86 | 0.36 |
| Mathematics 3 | 1 | 1,882 | 0.19 | 0.81 | 0.39 |
| Mathematics 3 | 2 | 1,997 | 0.17 | 0.83 | 0.32 |
| Mathematics 4 | 1 | 2,103 | 0.16 | 0.84 | 0.34 |
| Mathematics 4 | 2 | 2,186 | 0.17 | 0.83 | 0.38 |
| Mathematics 5 | 1 | 1,895 | 0.14 | 0.86 | 0.29 |
| Mathematics 5 | 2 | 2,172 | 0.14 | 0.86 | 0.31 |
| Mathematics 6 | 1 | 2,068 | 0.19 | 0.81 | 0.36 |
| Mathematics 6 | 2 | 2,052 | 0.16 | 0.84 | 0.37 |
| Mathematics 7 | 1 | 2,078 | 0.19 | 0.81 | 0.41 |
| Mathematics 7 | 2 | 1,998 | 0.12 | 0.88 | 0.30 |
| Mathematics 8 | 1 | 1,962 | 0.19 | 0.81 | 0.40 |
| Mathematics 8 | 2 | 1,829 | 0.18 | 0.82 | 0.37 |
| Mathematics 11 | 1 | 1,730 | 0.21 | 0.79 | 0.45 |
| Mathematics 11 | 2 | 1,843 | 0.22 | 0.78 | 0.44 |

## Tier-/Form-Reliability

In the context of multistage testing, tier-/form- reliability is not desirable as each form contains restricted range of difficulty of items and is administered to only a narrow sample of the performance range. Instead, marginal reliabilities averaged across student levels and all possible paths/forms were estimated and reported in Table 5. A more detailed rationale can be provided upon request.

Table 5. Marginal Test Reliability

|  |  |  |  |
| --- | --- | --- | --- |
| Test | Number of Items | Number of Examinees | Marginal Reliability |
| ELA 3 | 50 | 4,013 | 0.82 |
| ELA 4 | 50 | 4,311 | 0.85 |
| ELA 5 | 50 | 4,219 | 0.86 |
| ELA 6 | 50 | 4,243 | 0.85 |
| ELA 7 | 50 | 4,165 | 0.85 |
| ELA 8 | 50 | 3,867 | 0.87 |
| ELA 11 | 50 | 3,648 | 0.86 |
| Mathematics 3 | 50 | 3,879 | 0.82 |
| Mathematics 4 | 50 | 4,289 | 0.83 |
| Mathematics 5 | 50 | 4,067 | 0.86 |
| Mathematics 6 | 50 | 4,120 | 0.82 |
| Mathematics 7 | 50 | 4,076 | 0.84 |
| Mathematics 8 | 50 | 3,791 | 0.81 |
| Mathematics 11 | 50 | 3,573 | 0.78 |

## References:

California Department of Education. (2017). *California Alternate Assessment technical report, 2015–16 administration*. Retrieved from [https://www.cde.ca.gov/ta/tg/ca/‌documents/caa2016techrpt.pdf](https://www.cde.ca.gov/ta/tg/ca/documents/caa2016techrpt.pdf)

Green, B.F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement,* 21(4). 347–360.

Samejima, F. (1977). A use of the information function in tailored testing. *Applied Psychological Measurement,* 1. 233–247.

Posted by the California Department of Education | June 2018.

1. CDE, 2017 [↑](#footnote-ref-1)