

2.2 ACCESS Online Score Scale Maintenance: Reading

Series 400 to Series 303

Prepared by:
Shu Jing Yen
October 2015

Issue

The purpose of this document is to provide an overview of the planned approach to the equating analysis that links the reporting scale between the 2015-16 ACCESS for ELLs 2.0 Series 400 Online (hereafter ACCESS 2.0 Online) assessment of academic English language proficiency development expressed in the Reading domain (hereafter “Reading Test”) to the ACCESS for ELLs (hereafter ACCESS) Reading Test. Since the same method proposed for maintaining the ACCESS Listening Test score scale is also used for the ACCESS Reading Test, please refer to the Listening read-ahead for detailed methodology (Topic 2.1).

Starting with ACCESS 2.0 Online, the Reading Test is administered via computer. Students both receive stimuli and enter their responses on the computer. Table 1 presents the differences between the ACCESS 2.0 Online Reading field test and the ACCESS Reading operational test. This table shows that the construct being measured has not changed because the test blueprint and content coverage have been maintained. The only major differences for the Reading Test on ACCESS 2.0 Online are the mode of administration and the grade-level clustering in Grades 1-5. Furthermore, the ACCESS 2.0 Online Reading field test consists of items from previous operational ACCESS tests that have been adapted for computer delivery.

Background

The ACCESS 2.0 Online Reading field test was conducted in the 2013-2014 (Series 302) and 2014-2015 (Series 303) school years to collect data needed to conduct psychometric analysis. The field test was designed with the aim of collecting a sample of student performances on both the ACCESS 2.0 Online Reading field test and the ACCESS operational test such that a common-person design could be used to establish the link between the tests. The field test folders were administered to samples of students within two weeks after they had taken the operational ACCESS Reading Test. Folders were spiraled across field test forms with the goal of administering each folder to at least 350 students. ACCESS Reading Test item parameters had been pre-calibrated, and they were used as anchors to establish the link between the two tests.

The same three-step process used to maintain the ACCESS Listening scale is used for maintaining the ACCESS Reading scale. The analysis includes: (I) an outlier analysis, (II) concurrent calibration, and (III) a verification study. Steps I and II have been completed. However, an unexpected issue arose after the initial calibration was completed, necessitating further refinement of the initial calibration work. A summary of the issue and the procedure used to address this issue are summarized below.

Table 1
Differences between the ACCESS 2.0 Series 400 Online field test and ACCESS 302/303

	ACCESS Online field test	ACCESS
Grade Structure	1	1-2
	2-3	3-5
	4-5	
	6-8	6-8
	9-12	9-12
Items	Continuing (i.e., re-used) items adapted for computer delivery	
Stimulus	Computer	Paper
Response	Computer	Paper

Re-configuration of Grades 1-5 Clusters in ACCESS 2.0 Online

Since language and cognition develop rapidly and broaden in depth and breadth as younger children mature, with the advent of ACCESS 2.0 Online, Grade 1 forms its own cluster in order to better measure the language development of younger students. In addition, Grades 2-3 and Grades 4-5 comprise two different clusters. Because ACCESS 2.0 Online Reading was designed to be as comparable to ACCESS Reading as possible, reading field test folders consisted of reused ACCESS Grade 1-2 Reading folders that were adapted to the online format. Similarly, Grade 2-3 ACCESS 2.0 Online Reading field test folders consisted of reused Grade 1-2 or Grade 3-5 ACCESS Reading folders that were adapted to the online format. This grade cluster reconfiguration was not expected to be problematic due to the overlap in the population of students that these folders target. However, the reconfiguration of grade clusters appeared to have created some unexpected challenges in maintaining the characteristics of the Grades 1-5 portion of the ACCESS Reading scale.

Difficulty in Maintaining the Grades 1-5 Portion of the ACCESS Reading Scale

To maintain the ACCESS Reading score scale, ACCESS 2.0 Online Reading field test items were calibrated on the ACCESS Reading scale using ACCESS operational item parameters

as an anchor. ACCESS Reading item parameters were anchored to their initial values obtained from the annual equating study. Anchor item parameters were released or estimated only if the displacement statistics suggested that adjustments were needed in order to fit the observed response patterns of the current sample. For consistency, the criteria that were used during the annual equating for releasing anchor item parameters were also used in the initial calibration; that is, if the displacement statistics were greater than the absolute value of 0.3, item parameters were re-estimated.

It was observed that the majority of the Grade 1-2A and Grade 3-5A anchor parameters needed to be released or re-estimated, mostly because these items were shown to be more difficult for the field test sample group. We investigated this issue and concluded that this could be due to the differences in the student samples between the ACCESS 2.0 Online field test and the ACCESS equating sample for Grades 1 to 5. The Grade 1A ACCESS 2.0 Online field test sample consisted of only Grade 1 students, while the Grade 1-2A ACCESS equating sample consisted of both Grades 1 and 2 students. Since the Grade 1A ACCESS 2.0 Online field test sample was less able than the Grades 1-2A ACCESS equating sample, the same items would appear to be more difficult for the field test sample. Consequently, the anchor items parameters needed to be re-estimated in order to fit the observed response patterns of the field test sample. Similarly, the Grade 4-5A ACCESS 2.0 Online field test sample consisted of only Grades 4-5 students while the Grade 3-5A ACCESS equating sample consisted of Grades 3 to 5 students.

One consequence of releasing a large portion of ACCESS 1-2A and 3-5A anchor items is that the characteristics of the Grades 1-5 portion of the ACCESS Reading score scale is not maintained. In addition, Grades 1-5A ACCESS 2.0 Online field test items, on average, are more difficult than those typically seen in Grades 1-5A ACCESS items. To mitigate this problem, the following methodology was used to conduct a secondary calibration.

Refinement of the ACCESS 2.0 Online Reading Field Test Calibration

The goal of the secondary calibration was to make adjustments to the initial calibration such that the characteristics of the Grades 1-5 portion of the ACCESS Reading scale were maintained.

Step 1. Use ACCESS 2.0 Online field test folders that are common between the Grade 4-5 and Grade 6-8 clusters as anchors to maintain the vertical scale in the secondary calibration process. The item parameters for these five field test folders were fixed to the values from the initial calibration throughout the entire process.

Step 2. Conduct a concurrent calibration across the Grades 1-5 ACCESS 2.0 Online field test and ACCESS operational items.

Step 3. Evaluate the displacement statistics for ACCESS item parameters and release them using the following criterion:

- Release anchor items with displacement statistics greater than the absolute value of 0.5. This is a less stringent criterion than is typically used in the annual equating. Relaxing the displacement statistics criterion is necessary to preserve the characteristics of the Grades 1-5 portion of the current ACCESS scale as much as possible.
- Release anchor items in pairs where one item of each pair had the highest positive and the other item had the highest negative displacement statistics among all anchor items with displacement statistics greater than the absolute value of 0.5 for a given iteration of the calibration.
- Only release Grade 1-2A (Grade 3-5A) items if the item pair being released is from both from Grade 1-2A (Grade 3-5A.) This rule was used to avoid releasing all Grade 1-2A or Grade 3-5A ACCESS anchor items that had significant change in item difficulty in a single direction (for example, releasing all Grades 1-2A anchor items that were too difficult for the field test sample) during the calibration process.

Calibration was repeated each time one anchor item pair was released and their parameters were re-estimated. In each iteration of the secondary calibration process, the average item difficulty values across all Grade 1A ACCESS 2.0 Online field test folders were computed and compared with those from ACCESS Grade 1-2A folders to ensure that the average difficulty of the latter were maintained. A similar procedure was implemented to maintain the overall difficulty level of the ACCESS Grade 3-5A folders.

Questions for the TAC

- Question: What additional analyses might the TAC recommend to ensure that the linking is conducted appropriately and that the Reading score scale is maintained?