Smarter Balanced Summative Assessments Simulation Results

2016–2017 Test Administrations

English Language Arts/Literacy Grades 3–8, 11

Mathematics Grades 3–8, 11

October 2016

American Institutes for Research

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

1 INTRODUCTION

The Smarter Balanced summative assessment consists of two parts: a computerized adaptive test (CAT), and a performance task (PT) component. Each student is allowed a single opportunity to take the summative assessment. For the computer-adaptive test, prior to the operational testing window, simulation studies are conducted to evaluate and ensure the implementation and quality of the adaptive item-selection algorithm and the scoring algorithm. The simulation tool enables us to manipulate key blueprint configuration settings to match the blueprint and minimize measurement error. The adaptive tests are administered in one segment for English language arts/literacy (ELA/L) grades 3–8 and 11 and mathematics grades 3–5, and in two segments for mathematics grades 6–8 and 11- a calculator and a no calculator segment- each of which is configured separately. The performance tasks are also taken on a computer, but are not computer adaptive. Each student receives a set of items in a fixed-form. After the blueprint configurations for the adaptive tests are finalized, simulations are run on the combined tests, including both CAT and PT, to check the scoring algorithm.

This document describes the results of simulated adaptive test administrations (CAT component) used to configure and evaluate the adequacy of the item selection algorithm used to administer the CAT component of the 2016–17 Smarter Balanced summative assessments. The purpose of the simulations is to configure the adaptive algorithm to optimize item selection in order to meet blueprint specifications, while also targeting test information to student ability. When the adaptive algorithm is optimized, the student ability is estimated more precisely than would otherwise be possible in a fixed-form environment, especially for high- and low-performing students. Consequently, the test administrations (forms) generated by the adaptive algorithm will not be statistically parallel. Nevertheless, scores from the assessment should be comparable, and each test form should measure the same content, albeit with a different set of test items.

The Smarter Balanced summative test blueprints describe the content of the ELA/L and mathematics summative assessments for all grades tested, and how that content will be assessed. The summative test blueprints reflect the depth and breadth of the performance expectations of the Common Core State Standards. The test blueprints include critical information about the number of items and depth of knowledge (DoK) for items associated with each assessment target.

For the Smarter Balanced item pool, all items are developed in English. To accommodate students who use Braille and students who need tests in Spanish, a portion of the English item pool was transcribed in Braille or translated into Spanish. The ELA/L pool is available in English and Braille. The mathematics pool is available in English, Braille, and Spanish. The English 2016–17 CAT operational item pool was augmented by 2,572 items for ELA/L and 2,393 items for mathematics, compared to the 2015–16 CAT item pool, while the item pools in Braille and Spanish are similar to the 2015–16 item pools, with a few replaced or added items.

Simulations were run on the item pool in English to set key blueprint and configuration settings to match the blueprint and minimize measurement error. The settings for the English pool were then applied to the item pool in Braille and Spanish.

This report summarizes simulation results of the adaptive administrations in grades 3–8, and 11 for ELA/L in English and Braille, and mathematics in English, Braille, and Spanish in seven sections: (1) introduction; (2) summary of adaptive algorithm; (3) summary of adaptive operational item pool; (4) statistical summary indices, describing the specific indices used to examine score recovery precision; (5) summary of statistical analyses, including summary of simulation results on blueprint satisfaction, bias and precision of the

proficiency estimates that were obtained in CAT, item exposure rates, and off-grade item usage; (6) summary of field-test items; and (7) summary.

2 SUMMARY OF ADAPTIVE ALGORITHM

For the Smarter Balanced Summative assessments, item selection rules ensure that each student receives an assessment representing an adequate sample of the domain with appropriate difficulty. The algorithm maximizes the information for each student and allows for certain constraints to be set, ensuring that items selected represent the required content distribution. The Test Delivery System ensures that students are not exposed to the same items or passages in subsequent assessments if they attempt multiple opportunities for the same content area.

The adaptive algorithm selects items to administer on each student's assessment to match the test specifications (test blueprints), and to minimize the measurement error by administering an assessment with items targeted to a student's ability. Items selected for each student depend on the student's performance on previously selected items. The accuracy of the student responses to items determines the next item or passage that the student will see. Thus, each student is presented with a set of items that most accurately aligns with his or her proficiency level, based on grade-level content. Higher performance is followed by more difficult items, and lower performance is followed by less difficult items until test length constraints are met.

The initial ability is used to select the first item or the first item group. The system selects the first item from *k* items providing the most information, given prior student achievement, e.g., student achievement in 2016 (Cohen, C., & Albright, L., 2014a). The parameter *k* is a configurable parameter that can be used to mitigate item exposure or more closely match a student's performance depending on its value. Larger values of *k* provide more exposure control at the expense of optional selection.

For the initial ability to select the first item or the first item group, the algorithm can start the assessment in the following ways:

- If prior score is available, the algorithm starts the assessment with the score in the previous grade, grade 3 score for grade 4 student.
- If no prior score is available, the algorithm can:
    - Start an assessment with an item of average difficulty near the average ability of students in the previous administration, assuming the same initial ability for all students; or
    - Select the first item randomly from the pool.

In the 2016–17 CAT administration, in consultation with Smarter Balanced, students' first item will be selected randomly from the pool to minimize the unused item rate while providing the most information given prior student achievement. Subsequent items are selected based on student responses. After the first item or the first item group is administered, the algorithm identifies the best item to administer using the criteria presented in next sections.

2.1 Match to the Blueprint

The algorithm first selects items to maximize fit to the test blueprint. Blueprints specify a range of items to be administered in each claim for each assessment, with a collection of *constraint sets*. A constraint set is

a set of exhaustive, mutually exclusive classifications of items. For example, if a claim consists of four targets and each item measures one—and only one—of the claims, the claim classifications constitute a constraint set.

During item selection, the algorithm "rewards" claims that have not yet reached the minimum number of items. For example, if the measurement claim requires that an assessment contain between eight and nine items, measurement is the constrained feature. At any point in time, the minimum constraint on some features may have already been satisfied, while others may not have been. Other features may be approaching the maximum defined by the constraint. The value measure must reward items that have not yet met minimum constraints and penalize items that would exceed the maximum constraints. The algorithm stops administering items when the specified assessment length is met.

2.2 Match to Student Ability

In addition to rewarding items that match the blueprint, the adaptive algorithm also places greater value on items that maximize test information near the student's estimated ability, ensuring the most precise estimate of student ability possible, given the constraints of the item pool and satisfaction of the blueprint match requirement. After each response is submitted, the algorithm recalculates a student ability. As more answers are provided, the estimate becomes more precise, and the difficulty of the items selected for administration more closely aligns to the student's ability level. Higher performance (answering items correctly) is followed by more difficult items, and lower performance (answering items incorrectly) is followed by less difficult items. When the assessment is completed, the algorithm scores the overall assessment and each claim.

The algorithm allows previously answered items to be changed; however, it does not allow items to be skipped. Item selection requires iteratively updating the estimate of the overall and claim ability estimates after each item is answered. When a previously answered item is changed, the proficiency estimate is adjusted to account for the changed responses when the next new item is selected. While the update of the ability estimates is performed at each iteration, the overall and claim scores are recalculated using all data at the end of the assessment for the final score.

2.3 Off-Grade Item Selection

For students who are performing very well or very poorly on the test, if an item pool does not include a wide enough range of item difficulties for every test blueprint constraint, the item banks may run out of items that measure the student's proficiency sufficiently. This could potentially result in imprecise measurement for students in the tails of the proficiency distribution. Smarter Balanced selected off-grade items—one or two grades above and one to three grades below—and realigned the off-grade items to the on-grade blueprints.

Constraints enforced in administering off-grade items are as follows:

- Administer off-grade after a student responds to two-thirds of the operational items.
- The system should make it extremely unlikely that students could achieve a "proficient" determination based on below-grade content or could be denied a "proficient" determination based on above-grade content.

- The system should not allow off-grade items while a student maintains a non-trivial possibility of achieving proficiency (or dropping below it) based on on-grade items.

The on-grade item pool expands adding the off-grade items when a student reaches two-thirds of the test length, depending on a student's performance. At or after two-thirds of the test length, when a student's performance falls below the standard (not proficient) with a probability (p) < 0.0000001, the below-grade items are added to the on-grade item pool. Likewise, if a student's performance is above the standard (proficient) with a probability (p) < 0.0000001, the above-grade items are added to the on-grade item pool. More detailed statistical criteria for expanding the item pool can be found in the off-grade item selection approach document (Cohen, C., & Albright, L., 2014b).

2.4 Parameters Used to Simulate Student Proficiency

The testing of the adaptive item-selection algorithm begins by generating a sample of examinees with true thetas from a normal (μ,σ) distribution for each grade and subject. The parameters for the normal distribution are based on students' operational scores in the 2015–16 Smarter Balanced summative tests conducted. Each simulated examinee is administered one test opportunity for ELA/L and mathematics. The initial ability (prior ability) used to initiate the test by choosing the first few items is drawn from a uniform distribution within the range of true theta plus or minus 1. Table 1 provides the means and standard deviations used to generate a sample of student true abilities in the simulation by grade and subject.

Table 1. Population Parameters Used to Generate
True Ability Distributions for Simulated Test Administrations

Grade	ELA/L		Mathematics	
	Mean	SD	Mean	SD
3	–0.873	1.014	–0.950	1.006
4	–0.387	1.062	–0.429	1.017
5	0.051	1.080	–0.098	1.112
6	0.297	1.072	0.120	1.295
7	0.584	1.114	0.369	1.329
8	0.777	1.125	0.546	1.462
11	1.163	1.233	0.803	1.510

3 Operational Item Pool for Adaptive Tests

3.1 ELA/L Adaptive Item Pool

The ELA/L item pool in English and Braille is summarized in Tables 2 and 3, including both on-grade and off-grade items. In ELA/L, the items in claims 1 and 3 are associated with passages, while the items in claims 2 and 4 are discrete items. The Braille pool consists of a portion of the English item pool, ranging from 22% to 38% for items and from 21% to 44% for passages across grades. The number of off-grade items is summarized in Table 4. The off-grade items are selected from one or two grades above or one to three grades below the on-grade. The average difficulties of off-grade items are provided in Appendix B. Figure 1 presents the distribution of item difficulties for items in the English and Braille pools.

Table 2. Number of Items in the ELA/L Adaptive Operational Item Pool

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
English							
Total	890	873	886	826	763	808	2,463
Claim 1 Reading	325	265	303	241	234	225	859
Claim 2 Writing	250	267	276	261	251	285	688
Claim 3 Listening	184	192	163	161	170	186	560
Claim 4 Research	131	149	144	163	108	112	356
Braille							
Total	300	306	337	306	287	305	540
Claim 1 Reading	102	88	110	90	94	90	218
Claim 2 Writing	92	90	100	83	83	100	125
Claim 3 Listening	67	73	70	80	72	81	133
Claim 4 Research	39	55	57	53	38	34	64

Table 3. Number of Passages in the ELA/L Adaptive Operational Item Pool

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
English							
Total	131	121	117	108	107	110	371
Claim 1 Long Literary	14	11	13	11	10	9	46
Claim 1 Short Literary	17	14	15				
Claim 1 Long Information	11	10	11	11	7	12	39
Claim 1 Short Information	15	11	11	20	22	15	68
Claim 3 Listening	74	75	67	66	68	74	218
Braille							
Total	44	42	45	48	43	46	79
Claim 1 Long Literary	5	3	3	4	4	2	9
Claim 1 Short Literary	3	4	6				
Claim 1 Long Information	4	3	4	3	2	4	7
Claim 1 Short Information	5	4	4	8	10	8	15
Claim 3 Listening	27	28	28	33	27	32	48

Table 4. Number of Off-Grade Items in the ELA/L Adaptive Operational Item Pool

Grade	English		Braille	
	Above Grade	Below Grade	Above Grade	Below Grade
3	11	n/a	8	n/a
4	18	20	12	15
5	15	43	6	32
6	16	41	5	31
7	19	45	11	31
8	15	42	3	20
11	n/a	28	n/a	12

Figure 1. ELA/L Item Difficulty Distribution for English and Braille Pool

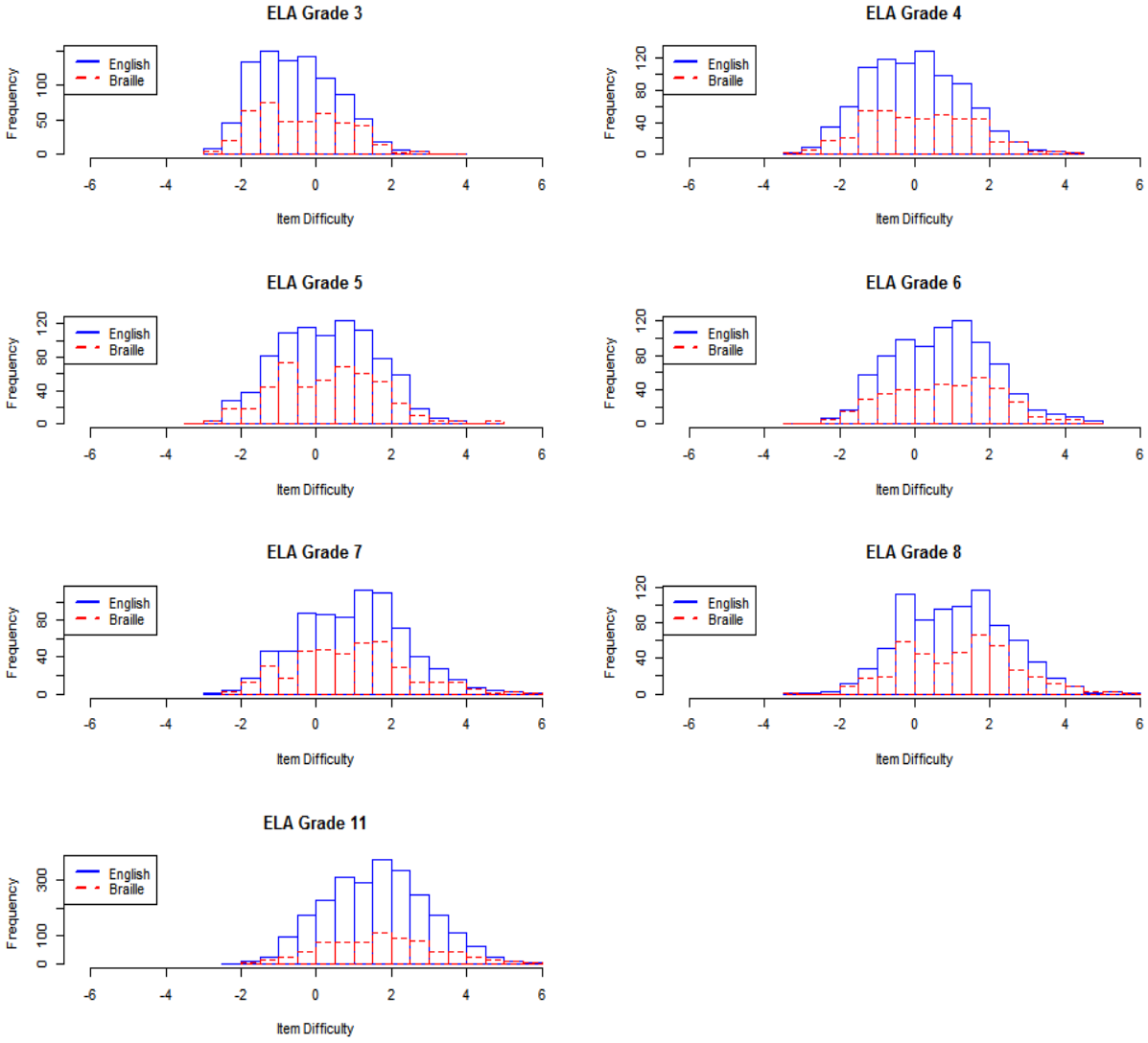


Table 5 provides the average item difficulty for the adaptive operational item pool and the average estimated ability for the simulated students. The average item difficulties are higher than the average student abilities in all grades in both English and Braille, with larger differences in grades 6 and 11. Relative to the 2015–16 pool, the difference between the average item difficulties and the average student abilities, however, decreased with the 2016–17 pool, which was augmented with easy items. The distribution of item difficulties in 2015–16 and 2016–17 adaptive operational item pool can be found in Appendix A. The distribution of item difficulties and estimated abilities is overlaid in Figure 2 for the English pool and Figure 3 for the Braille pool.

Table 5. Average Difficulty of the ELA/L Adaptive Operational Item Pool and Average Observed Student Ability Estimates for Simulated Test Administrations

Grade	English				Braille			
	Items		Ability		Items		Ability	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
3	-0.516	1.087	-0.870	1.025	-0.481	1.222	-0.878	1.157
4	-0.006	1.276	-0.367	1.149	0.000	1.456	-0.395	1.082
5	0.298	1.292	0.032	1.156	0.145	1.402	0.024	1.140
6	0.753	1.342	0.332	1.126	0.666	1.485	0.263	1.174
7	0.965	1.386	0.611	1.223	0.810	1.519	0.601	1.210
8	1.070	1.373	0.810	1.240	1.065	1.483	0.807	1.203
11	1.628	1.346	1.133	1.364	1.667	1.482	1.098	1.297

Figure 2. ELA/L Student Ability–Item Difficulty Distribution (English)

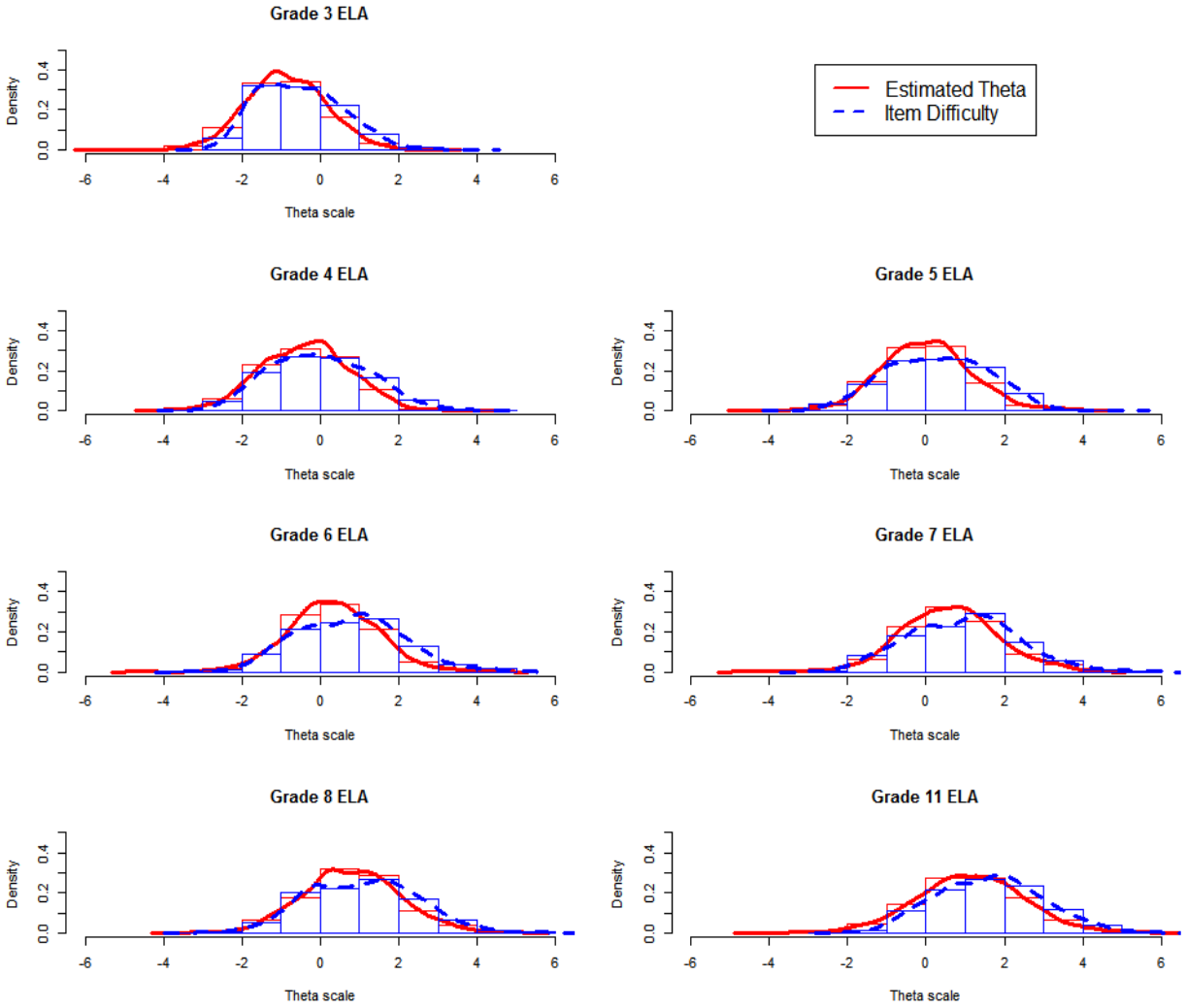
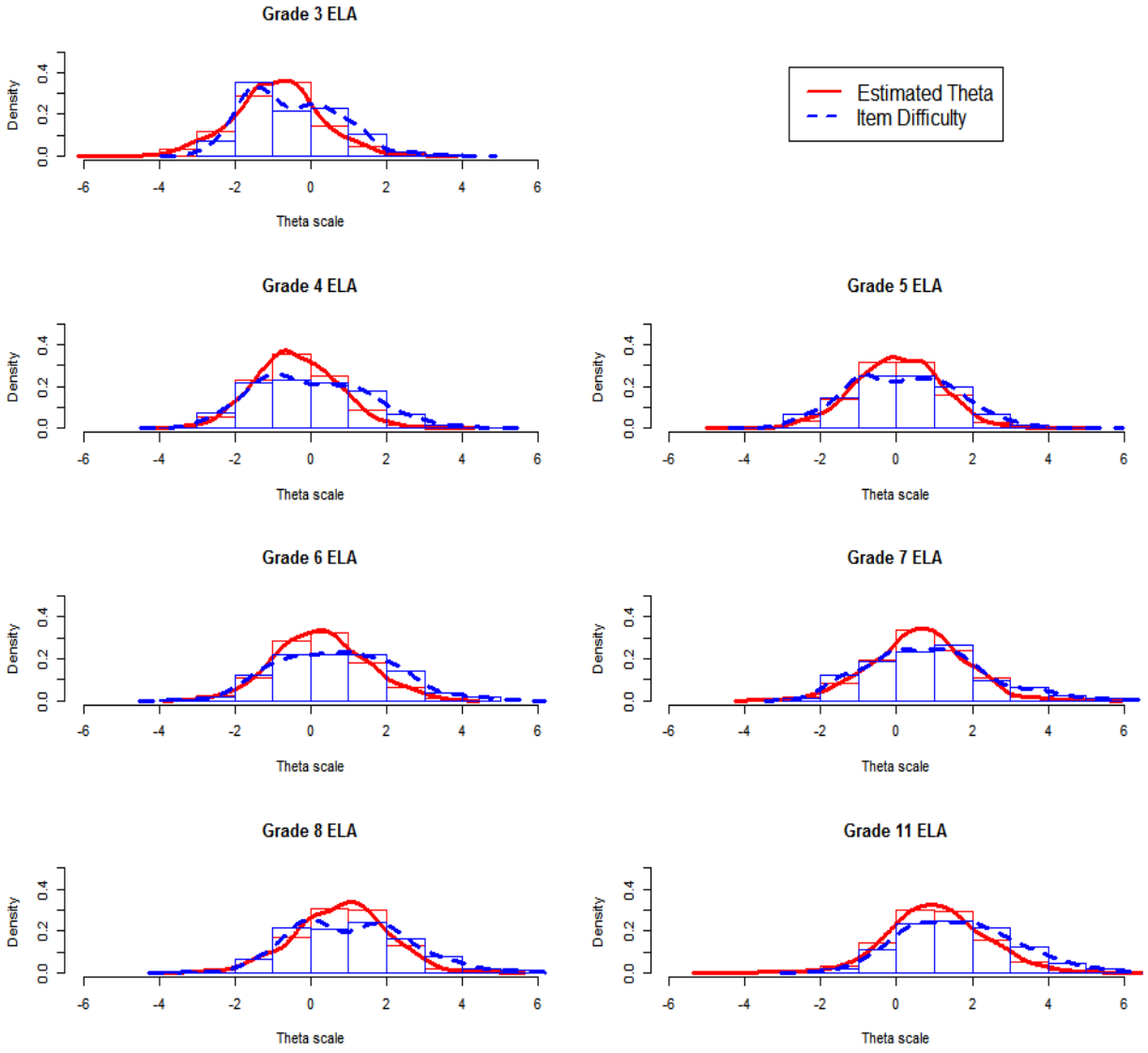


Figure 3. ELA/L Student Ability–Item Difficulty Distribution (Braille)



3.2 Mathematics Adaptive Item Pool

The adaptive item pool for mathematics consists of discrete items only. The adaptive pools in English, Braille, and Spanish are summarized in Table 6. The Braille and Spanish pools consist of a portion of the English item pool, ranging from 20% to 37% for Braille and from 26% to 36% for Spanish. The number of off-grade items is summarized in Table 7. The average difficulties of off-grade items are provided in Appendix B. Figure 4 exhibits the distribution of item difficulties for the English, Braille, and Spanish pools.

Table 6. Number of Items in the Mathematics Adaptive Operational Item Pool

Grade	Calculator	Total	Claim 1	Claim 2	Claim 3	Claim 4
			English			
3	No Calculator	1,196	785	98	197	116
4	No Calculator	1,306	865	114	206	121
5	No Calculator	1,267	833	91	201	142
6	Calculator	566	261	83	142	80
	No Calculator	533	515		18	
7	Calculator	681	378	80	134	89
	No Calculator	294	294			
8	Calculator	646	396	57	131	62
	No Calculator	216	216			
11	Calculator	1,745	1,009	164	381	191
	No Calculator	890	841		49	
			Braille			
3	No Calculator	385	248	42	55	40
4	No Calculator	359	239	39	41	40
5	No Calculator	381	234	38	56	53
6	Calculator	190	90	38	40	22
	No Calculator	195	193		2	
7	Calculator	255	151	34	46	24
	No Calculator	106	106			
8	Calculator	206	134	14	45	13
	No Calculator	84	84			
11	Calculator	351	175	37	94	45
	No Calculator	173	161		12	
			Spanish			
3	No Calculator	375	231	48	52	44
4	No Calculator	388	232	48	59	49
5	No Calculator	406	229	44	70	63
6	Calculator	196	87	33	49	27
	No Calculator	199	193		6	
7	Calculator	244	141	27	49	27
	No Calculator	100	100			
8	Calculator	233	141	15	50	27
	No Calculator	76	76			
11	Calculator	456	252	50	102	52
	No Calculator	234	219		15	

Table 7. Number of Off-Grade Items in the Mathematics Adaptive Operational Item Pool

Grade	English		Braille		Spanish	
	Above Grade	Below Grade	Above Grade	Below Grade	Above Grade	Below Grade
3	4	n/a	1	n/a	1	n/a
4	n/a	27	n/a	20	n/a	19
5	n/a	57	n/a	30	n/a	30
6	n/a	31	n/a	20	n/a	20
7	5	27	n/a	17	n/a	14
8	3	16	n/a	11	n/a	8
11	n/a	13	n/a	7	n/a	5

Figure 4. Mathematics Item Difficulty Distribution for English, Braille and Spanish Pools

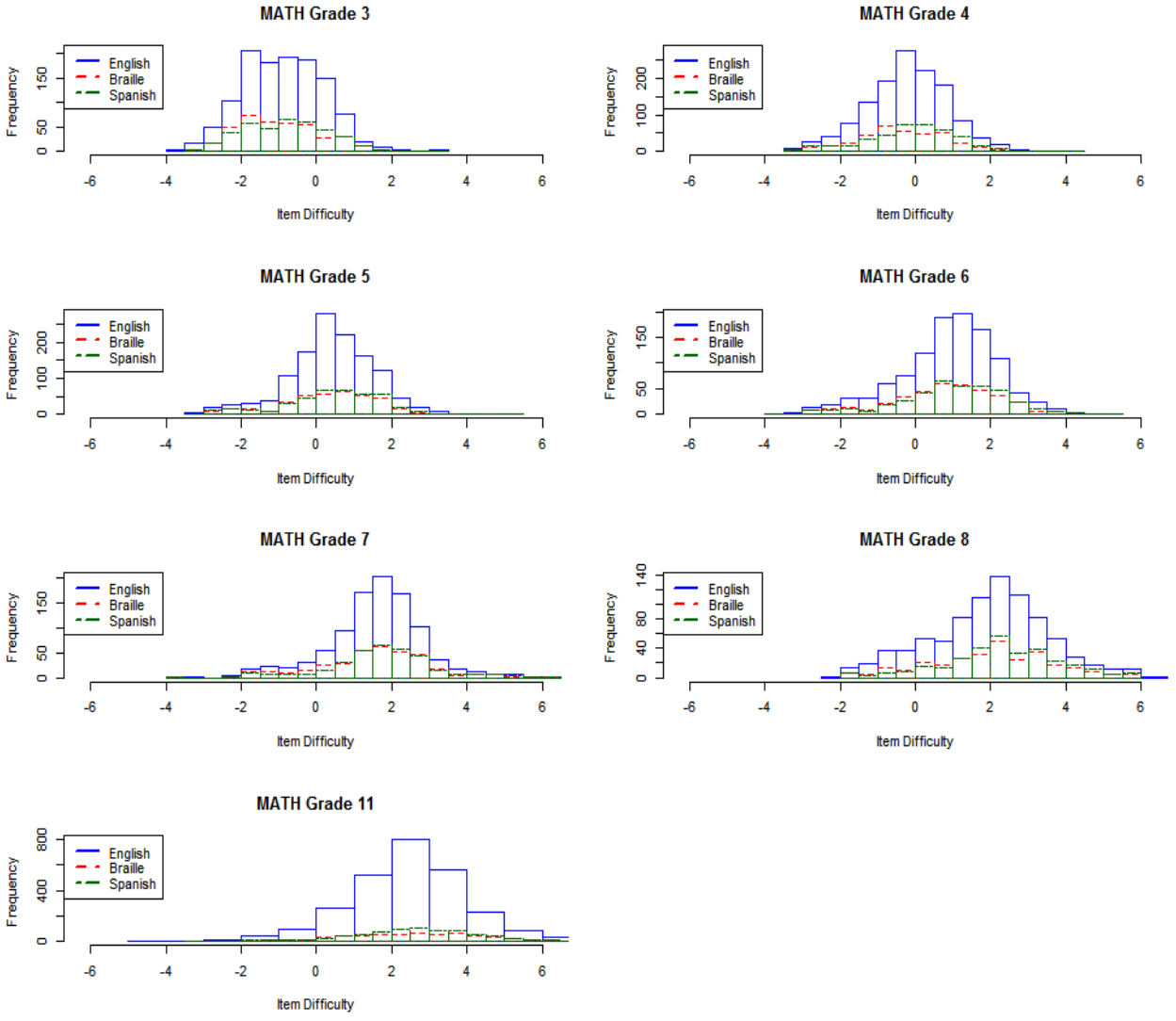


Table 8 provides the average item difficulty for the adaptive operational item pool and the average estimated ability for the simulated students for mathematics. The average item difficulties are higher than the average student abilities, especially in grades 7–8 and 11, which makes difficult to select items to maximize assessment information near the student's estimated ability in the lower end of the ability range while meeting the blueprint requirements. The distribution of item difficulties and estimated abilities is overlaid in Figures 5–7 for the English, Braille, and Spanish pools. The difference between the average item difficulties and the average student abilities, however, decreased relative to the difference in the 2015-16 pool because easy items were added to the 2016–17 operational item pool. The distribution of item difficulties in the 2015–16 and 2016–17 adaptive pools can be found in Appendix A.

Table 8. Average Difficulty of the Adaptive Operational Item Pool and Average Observed Student Ability Estimates for Simulated Test Administrations for Mathematics

Grade	English				Braille				Spanish			
	Items		Ability		Items		Ability		Items		Ability	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
3	-0.884	1.044	-0.946	1.063	-0.964	1.084	-1.012	1.060	-0.824	1.097	-0.971	1.058
4	-0.217	1.061	-0.454	1.021	-0.357	1.150	-0.415	1.064	-0.100	1.177	-0.435	1.059
5	0.384	1.149	-0.079	1.187	0.314	1.293	-0.120	1.178	0.443	1.309	-0.151	1.194
6	0.877	1.315	0.061	1.346	0.787	1.472	0.143	1.339	0.940	1.480	0.104	1.401
7	1.577	1.272	0.409	1.389	1.552	1.482	0.303	1.386	1.734	1.438	0.331	1.368
8	1.965	1.566	0.525	1.626	2.012	1.655	0.431	1.607	2.315	1.518	0.531	1.548
11	2.447	1.513	0.823	1.642	2.618	1.708	0.699	1.748	2.730	1.551	0.773	1.651

Figure 5. Mathematics Student Ability–Item Difficulty Distribution (English)

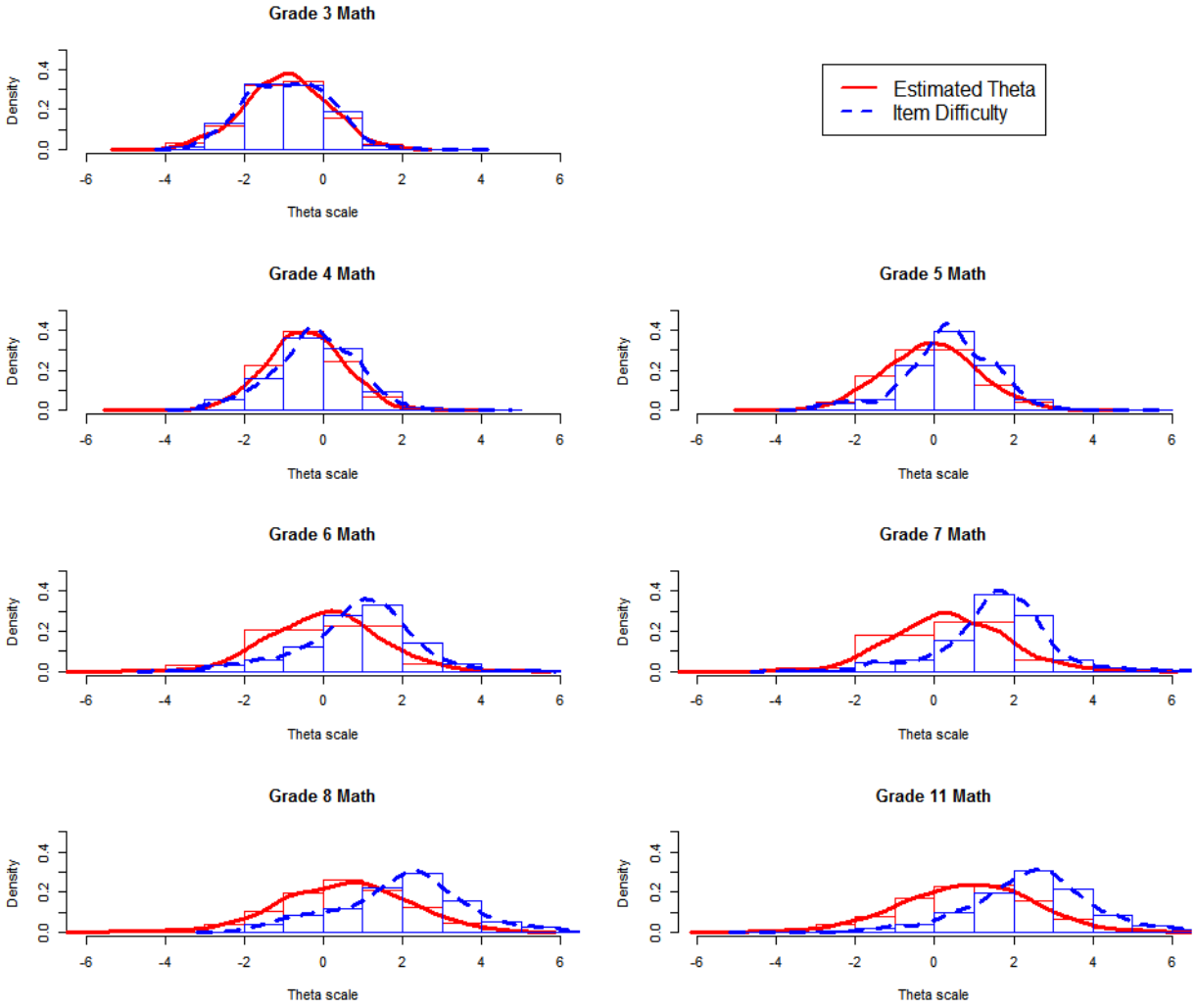


Figure 6. Mathematics Student Ability–Item Difficulty Distribution (Braille)

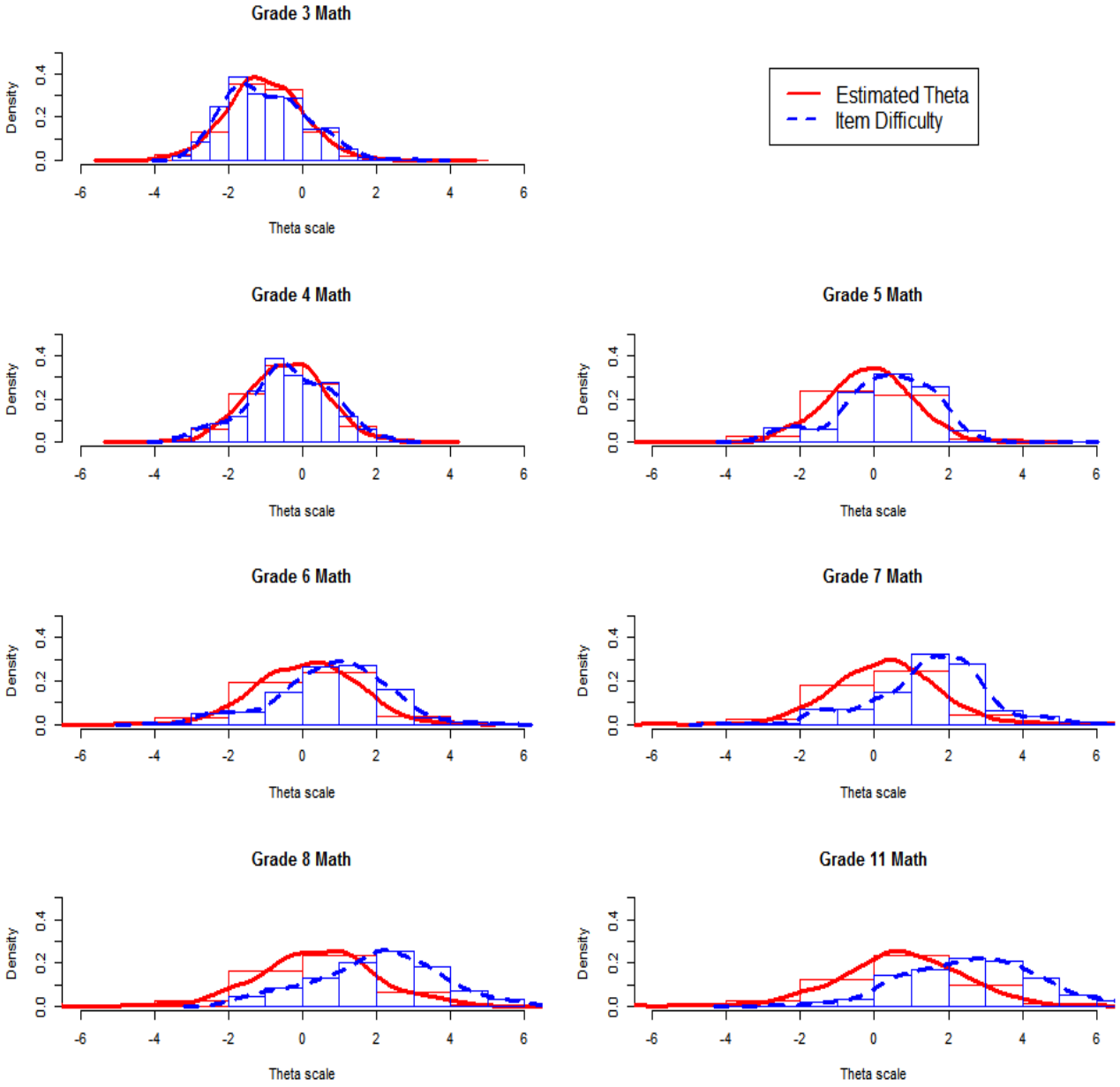
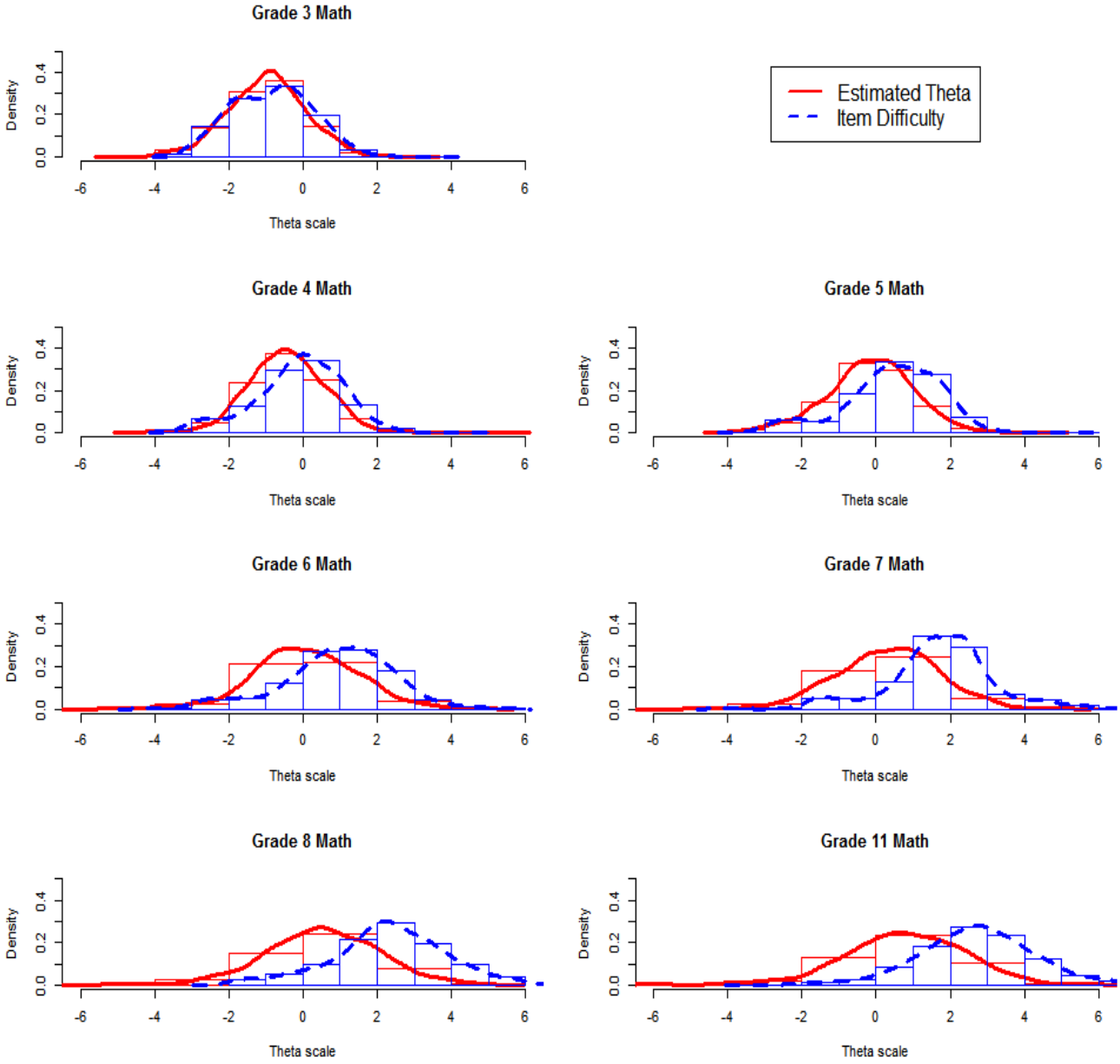


Figure 7. Mathematics Student Ability–Item Difficulty Distribution (Spanish)



4 STATISTICAL SUMMARY INDICES

The statistics computed include the following: the statistical bias of the estimated theta; mean squared error (MSE); significance of the bias; average standard error of the estimated theta; standard error of theta at the 5th, 25th, 75th, and 95th percentiles; and the percentage of students' estimated theta falling outside the 95% and 99% confidence intervals. Statistical bias refers to whether test scores systematically underestimate or overestimate the student's true ability.

Computational details of each statistic are provided below:

$$bias = N^{-1} \sum_{i=1}^{N} (\theta_i - \hat{\theta}_i) \qquad (1)$$

$$MSE = N^{-1} \sum_{i=1}^{N} (\theta_i - \hat{\theta}_i)^2$$

where θ_i is the true theta and $\hat{\theta}_i$ is the estimated theta for individual *i*. For the variance of the bias, a first-order Taylor series of equation (1) is used as:

$$var(bias) = \sigma^2 * \ g'(\hat{\theta}_i)^2 = \frac{1}{N(N-1)} \sum_{i=1}^{N} (\theta_i - \bar{\hat{\theta}}_i)^2$$

where, $\bar{\hat{\theta}}_i$ is an average of the estimated theta.

Significance of the bias is then tested as:

$$z = bias / \sqrt{var(bias)}$$

A *p*-value for the significance of the bias is reported from the above *z* test.

The average standard error of the estimated theta is computed as:

$$mean\ (se) = \sqrt{N^{-1} \sum_{i=1}^{N} se(\hat{\theta}_i)^2}$$

where $se(\hat{\theta}_i)$ is the standard error of the estimated theta (θ) for individual *i*.

To determine the percentage of students' estimated theta falling outside the 95% and 99% confidence interval coverage, a *t*-statistic is performed as follows:

$$t = \frac{\theta_i - \hat{\theta}_i}{se(\hat{\theta}_i)}$$

where $\hat{\theta}_i$ is the estimated theta for individual *i*, and θ_i is the true theta for individual *i*. The percentage of students' estimated theta falling outside the coverage is determined by comparing the absolute value of the *t*-statistic to a critical value of 1.96 for the 95% coverage and to 2.58 for the 99% coverage.

5 SUMMARY OF STATISTICAL ANALYSES

This section summarizes the results of the statistics computed to examine the robustness of the item-selection algorithm. For each grade and subject, 1,000 tests were simulated. For the item exposure rates, however, 3,000 tests in grades 3–8 and 5,000 tests in grade 11 were simulated because more items get used as the sample size increases.

5.1 Summary Statistics on Test Blueprints

The Smarter Balanced blueprints (Smarter Balanced Assessment Consortium, 2015) specify a range of items to be administered in each claim, content domain/standards, and targets. Moreover, blueprints constrain DoK, item types, and passage types. In blueprints, all content blueprint elements are configured to obtain a strictly enforced range of items administered. The algorithm also seeks to satisfy target-level constraints, but these ranges are not strictly enforced. In ELA/L, the blueprints also specify the number of passages in the reading (Claim 1) and listening (Claim 3) claims.

5.1.1 Summary Statistics on Test Blueprint Match for ELA/L

Tables 9–10 present the percentages of tests aligned with the test blueprints for ELA/L in English and Braille. The blueprint match rates are summarized for item and passage requirements by claim. For the tests in English, all tests met the blueprint constraints for claims and passages, while some tests in Braille received one more item or passage than the blueprint constraints in claim 1 reading and/or claim 3 listening.

For DoK and item type constraints, the constraints with a blueprint match rate less than 100% are summarized in Table 11. All violations, except for three or four Braille tests in grades 4, 6, and 11, are due to one item fewer being administered than the minimum requirement.

Table 9. Percentage of ELA/L Test Administrations Meeting Blueprint Requirements for Each Claim

Grade	Claim	Min.	Max.	%BP Match for Item Requirement	%BP Match Passage Requirement
3	1-LT	7	8	100%	100%
3	1-IT	7	8	100%	100%
3	2-W	10	10	100%	
3	3-L	8	8	100%	100%
3	4-CR	6	6	100%	
4	1-LT	7	8	100%	100%
4	1-IT	7	8	100%	100%
4	2-W	10	10	100%	
4	3-L	8	8	100%	100%
4	4-CR	6	6	100%	
5	1-LT	7	8	100%	100%
5	1-IT	7	8	100%	100%
5	2-W	10	10	100%	
5	3-L	8	9	100%	100%
5	4-CR	6	6	100%	
6	1-LT	4	4	100%	100%
6	1-IT	10	12	100%	100%
6	2-W	10	10	100%	
6	3-L	8	9	100%	100%

Grade	Claim	Min.	Max.	%BP Match for Item Requirement	%BP Match Passage Requirement
6	4-CR	6	6	100%	
7	1-LT	4	4	100%	100%
7	1-IT	10	12	100%	100%
7	2-W	10	10	100%	
7	3-L	8	9	100%	100%
7	4-CR	6	6	100%	
8	1-LT	4	4	100%	100%
8	1-IT	12	12	100%	100%
8	2-W	10	10	100%	
8	3-L	8	9	100%	100%
8	4-CR	6	6	100%	
11	1-LT	4	4	100%	100%
11	1-IT	11	12	100%	100%
11	2-W	10	10	100%	
11	3-L	8	9	100%	100%
11	4-CR	6	6	100%	

Legend:
IT: Reading with Informational Text, 1-LT: Reading with Literary Text, 2-W: Writing, 3-L: Listening, and 4-CR: Research

Table 10. ELA/L: Percentage of Test Administrations Meeting Blueprint Requirements for Claims and Passages (Braille)

Grade	Claim	Item Requirement	Passage Requirement	Grade	Claim	Item Requirement	Passage Requirement
3	1-LT	100%	97.3%	7	1-LT	100%	98.8%
3	1-IT	100%	98.0%	7	1-IT	100%	100%
3	2-W	100%		7	2-W	100%	
3	3-L	100%	100%	7	3-L	100%	100%
3	4-CR	100%		7	4-CR	100%	
4	1-LT	100%	100%	8	1-LT	100%	100%
4	1-IT	100%	100%	8	1-IT	100%	99.6%
4	2-W	100%		8	2-W	100%	
4	3-L	100%	100%	8	3-L	100%	99.8%
4	4-CR	100%		8	4-CR	100%	
5	1-LT	100%	100%	11	1-LT	100%	100%
5	1-IT	100%	100%	11	1-IT	100%	100%
5	2-W	100%		11	2-W	100%	
5	3-L	100%	97.6%	11	3-L	100%	95.6%
5	4-CR	100%		11	4-CR	100%	
6	1-LT	94.3%	93.6%				
6	1-IT	100%	100%				
6	2-W	100%					
6	3-L	100%	75.2%				
6	4-CR	100%					

Legend:
1-IT: Reading with Informational Text, 1-LT: Reading with Literary Text, 2-W: Writing, 3-L: Listening, and 4-CR: Research

Table 11. ELA/L: Percentage of Test Administrations Meeting Blueprint Requirements for DoK and Item Type

Grade	DoK and Item Type Constraints	Min	Max	%BP Match
	English			
3	Claim 2 DOK2	4	6	90.6
6	Claim 2 DOK2	4	6	70.6
	Braille			
4	Brief Write	1	1	94.3
5	Brief Write	1	1	30.1
6	Brief Write	1	1	99.9
11	Brief Write	1	1	99.8
4	Claim 1 DOK2	7	10	99.9
6	Claim 1 DOK2	7	10	99.9
3	Claim 2 DOK2	4	6	82.9
4	Claim 2 DOK2	4	6	76.1
5	Claim 2 DOK2	4	6	99.2
6	Claim 2 DOK2	4	6	61.2
8	Claim 2 DOK2	4	6	87.4
11	Claim 2 DOK2	4	6	53.1
4	Claim 2 DOK3 or higher	1	4	94.3
5	Claim 2 DOK3 or higher	1	4	30.1
6	Claim 2 DOK3 or higher	1	4	99.9
11	Claim 2 DOK3 or higher	1	4	99.8
3	Claim 2 Evidence & Elaboration in Target 1,3,6	1	2	93.1
4	Claim 2 Evidence & Elaboration in Target 1,3,6	1	2	99.6

The Smarter Balanced blueprints for ELA/L do not require every target to be covered in a claim; therefore, all targets listed in the blueprint are not expected to be covered in every test. Table 12 summarizes the number of unique targets administered in each simulated test by claim. The table includes the number of targets specified in the blueprints and the item pool, and the mean and range of the number of targets administered to students.

In Claim 1 and Claim 2, the target coverage varies somewhat across individual tests; however, all targets are covered at an aggregate level, across all simulated tests combined. Targets in claims 3 and 4 are covered in every test, except for a few tests in Braille.

Table 12. ELA/L: Number of Unique Targets Assessed Within Each Claim

Grade	Total Targets Specified in BP				Mean				Range (Minimum-Maximum)			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
English												
3	14	5	1	3	10.2	4.0	1	3	8-13	3-5	1-1	3-3
4	14	5	1	3	10.3	4.1	1	3	8-13	3-5	1-1	3-3
5	14	5	1	3	10.1	4.7	1	3	7-12	3-5	1-1	3-3
6	14	5	1	3	9.3	4.1	1	3	8-11	3-5	1-1	3-3
7	14	5	1	3	9.4	4.9	1	3	8-11	4-5	1-1	3-3
8	14	5	1	3	9.4	4.0	1	3	8-11	3-4	1-1	3-3
11	14	5	1	3	9.2	4.0	1	3	6-11	3-4	1-1	3-3
Braille												
3	14	5	1	3	10.5	4.2	1	3	8-13	3-5	1-1	3-3
4	14	5	1	3	10.6	4.5	1	3	7-13	3-5	1-1	2-3
5	14	5	1	3	10.7	5.0	1	3	7-12	5-5	1-1	3-3
6	14	5	1	3	9.8	4.1	1	3	8-12	3-5	1-1	3-3
7	14	5	1	3	10.3	4.7	1	3	8-11	4-5	1-1	2-3
8	12	5	1	3	9.9	4.0	1	3	8-11	3-5	1-1	2-3
11	14	5	1	3	9.2	4.2	1	3	8-11	3-5	1-1	3-3

5.1.2 Summary Statistics on Test Blueprint Match for Mathematics

Tables 13–21 present the percentages of tests aligned with the test blueprints for mathematics in English, Braille, and Spanish. The blueprint match rates are summarized by claims and content domains within each claim. All tests met the blueprint requirements for claims, but there were a few exceptions in a few content domains. A few tests administered one item more than the maximum item requirements for content domains.

Table 22 provides a summary of the percentages of tests meeting the DoK and target constraints. Only the constraints with blueprint violations are included in the summary. The DoK and target constraints are met in all tests in English while some tests in Braille and Spanish are administered with one item fewer than the minimum item requirement.

Table 13. Percentage of Test Administrations Meeting Blueprint Requirements for Each Claim and Content Domains: Mathematics Grades 3–5 (English)

Claim	Content Domain	Grade 3			Grade 4			Grade 5		
		Min.	Max.	%BP Match	Min.	Max.	%BP Match	Min.	Max.	%BP Match
1	ALL	20	20	100%	20	20	100%	20	20	100%
1	P	15	15	100%	15	15	100%	15	15	100%
1	S	5	5	100%	5	5	100%	5	5	100%
2	ALL	3	3	100%	3	3	100%	3	3	100%
2	G	0	2	100%	0	2	100%	0	2	100%
2	MD	0	2	100%	0	2	100%	0	2	100%
2	NBT	0	2	100%	0	2	100%	0	2	100%
2	NF	0	2	100%	1	3	100%	1	3	100%
2	OA	0	2	100%	0	2	100%	0	2	100%
3	All	8	8	100%	8	8	100%	8	8	100%
3	G							0	3	100%
3	MD	0	4	100%				0	4	100%
3	NBT				0	4	100%	0	4	100%
3	NF	2	6	100%	2	6	100%	2	6	100%
3	OA	0	4	100%	0	4	100%			
3	Other				0	2	100%			
4	All	3	3	100%	3	3	100%	3	3	100%
4	G	0	1	100%	0	1	100%	0	1	100%
4	MD	1	2	100%	0	2	100%	1	2	100%
4	NBT	0	1	100%	0	1	100%	0	1	100%
4	NF	0	1	100%	0	2	100%	1	2	100%
4	OA	1	2	100%	0	2	100%	0	1	100%

Legend:

ALL	Total item requirement in a claim		
1-P	Primary target set in claim 1	NBT	Number and operations in Base ten
1-S	Secondary target set in claim 1	NF	Number and operations—fractions
G	Geometry	OA	Operations and algebraic thinking
MD	Measurement and data	OTHER	Other content domains

Table 14. Percentage of Test Administrations Meeting Blueprint Requirements for Each Claim and Content Domains: Mathematics Grades 6–7 (English)

Claim	Content Domain	Segment	Grade 6			Grade 7		
			Min.	Max.	%BP Match	Min.	Max.	%BP Match
1	ALL	Calc	6	6	100%	10	10	100%
1	P	Calc	3	3	100%	6	6	100%
1	S	Calc	3	3	100%	4	4	100%
1	ALL	NoCalc	13	13	100%	10	10	100%
1	P	NoCalc	11	11	100%	9	9	100%
1	S	NoCalc	2	2	100%	1	1	100%
2	ALL	Calc	3	3	100%	3	3	100%
2	EE	Calc	0	2	100%	0	2	100%
2	G	Calc	0	2	100%	0	2	100%
2	NS	Calc	0	2	100%	0	2	100%
2	RP	Calc	0	2	100%	0	2	100%
2	SP	Calc	0	2	100%	0	2	100%
2	OTHER	Calc	0	2	100%	0	2	100%
3	All	Calc	7	7	100%	8	8	100%
3	EE	Calc	0	5	100%	1	5	100%
3	NS	Calc	2	6	100%	1	5	100%
3	RP	Calc	0	5	100%	1	5	100%
3	All	NoCalc	1	1	100%			
3	EE	NoCalc	0	1	100%			
3	NS	NoCalc	0	1	100%			
3	RP	NoCalc	0	1	100%			
4	All	Calc	3	3	100%	3	3	100%
4	EE	Calc	0	1	100%	0	1	100%
4	G	Calc	0	1	100%	0	1	100%
4	NS	Calc	0	1	98.6%	0	1	100%
4	RP	Calc	0	1	96.2%	0	1	100%
4	SP	Calc	0	1	100%	0	1	100%
4	OTHER	Calc	0	1	100%	0	1	100%

Legend:

ALL	Total item requirement in a claim	N	Number and quantity
1-P	Primary target set in claim 1	NBT	Number and operations in Base ten
1-S	Secondary target set in claim 1	NF	Number and operations—fractions
A	Algebra	NS	Number system
EE	Expressions and equations	OA	Operations and algebraic thinking
F	Functions	OTHER	Other content domains
G	Geometry	RP	Ratios and proportional relationships
MD	Measurement and data	SP	Statistics and probability
Calc	Segment with calculator use	NoCalc	Segment without calculator use

Table 15. Percentage of Test Administrations Meeting Blueprint Requirements for Each Claim and Content Domains: Mathematics Grades 8 and 11 (English)

Grade 8						Grade 11					
Claim	**Content Domain**	**Segment**	**Min**	**Max**	**%BP Match**	**Claim**	**Content Domain**	**Segment**	**Min**	**Max**	**%BP Match**
1	ALL	Calc	14	14	100%	1	ALL	Calc	11	11	100%
1	P	Calc	11	11	100%	1	P	Calc	8	8	100%
1	S	Calc	3	3	100%	1	S	Calc	3	3	100%
1	ALL	NoCalc	6	6	100%	1	ALL	NoCalc	11	11	100%
1	P	NoCalc	4	4	100%	1	P	NoCalc	8	8	100%
1	S	NoCalc	2	2	100%	1	S	NoCalc	3	3	100%
2	ALL	Calc	3	3	100%	2	ALL	Calc	3	3	100%
2	EE	Calc	0	2	100%	2	A	Calc	1	2	100%
2	F	Calc	0	2	100%	2	F	Calc	0	2	100%
2	G	Calc	0	2	100%	2	G	Calc	0	2	100%
2	NS	Calc	0	2	100%	2	N	Calc	0	2	100%
2	SP	Calc	0	2	100%	2	S	Calc	0	2	100%
2	OTHER	Calc	0	2	100%	2	O	Calc	0	2	100%
3	ALL	Calc	8	8	100%	3	All	Calc	7	7	100%
3	EE	Calc	1	5	100%	3	A	Calc	1	4	100%
3	F	Calc	1	5	100%	3	F	Calc	0	4	100%
3	G	Calc	1	5	100%	3	G	Calc	1	4	100%
						3	N	Calc	0	4	100%
						3	All	NoCalc	1	1	100%
						3	A	NoCalc	0	1	100%
						3	F	NoCalc	0	1	100%
						3	G	NoCalc	0	1	100%
						3	N	NoCalc	0	1	100%
4	ALL	Calc	3	3	100%	4	All	Calc	3	3	100%
4	EE	Calc	1	2	100%	4	A	Calc	0	2	99.9%
4	F	Calc	0	1	100%	4	F	Calc	0	1	99.7%
4	G	Calc	0	1	100%	4	G	Calc	0	1	91.1%
4	NS	Calc	0	1	100%	4	N	Calc	0	2	100%
4	SP	Calc	0	1	97.1%	4	S	Calc	0	2	100%
4	OTHER	Calc	0	1	100%	4	O	Calc	0	1	100%

Legend:

ALL	Total item requirement in a claim	N	Number and quantity
1-P	Primary target set in claim 1	NBT	Number and operations in Base ten
1-S	Secondary target set in claim 1	NF	Number and operations—fractions
A	Algebra	NS	Number system
EE	Expressions and equations	OA	Operations and algebraic thinking
F	Functions	OTHER	Other content domains
G	Geometry	RP	Ratios and proportional relationships
MD	Measurement and data	SP	Statistics and probability
Calc	Segment with calculator use	NoCalc	Segment without calculator use

Table 16. Mathematics Grades 3–5:
Percentage of Test Administrations Meeting Blueprint Requirements
for Claims and Content Domains (Braille)

Claim	Content Domain	Grade 3			Grade 4			Grade 5		
		Min.	Max.	%BP Match	Min.	Max.	%BP Match	Min.	Max.	%BP Match
1	ALL	20	20	100%	20	20	100%	20	20	100%
1	P	15	15	100%	15	15	100%	15	15	100%
1	S	5	5	100%	5	5	100%	5	5	100%
2	ALL	3	3	100%	3	3	100%	3	3	100%
2	G							0	2	100%
2	MD	0	2	100%	0	2	100%	0	2	100%
2	NBT	0	2	100%	0	2	100%	0	2	100%
2	NF	0	2	100%	1	3	100%	1	3	100%
2	OA	0	2	100%	0	2	100%			
3	All	8	8	100%	8	8	100%	8	8	100%
3	G							0	3	100%
3	MD	0	4	100%				0	4	100%
3	NBT				0	4	100%	0	4	100%
3	NF	2	6	100%	2	6	99.7%	2	6	98.7%
3	OA	0	4	100%	0	4	100%			
3	Other									
4	All	3	3	100%	3	3	100%	3	3	100%
4	G				0	1	100%	0	1	100%
4	MD	1	2	100%	0	2	100%	1	2	100%
4	NBT	0	1	100%	0	1	100%	0	1	100%
4	NF	0	1	100%	0	2	100%	1	2	100%
4	OA	1	2	100%	0	2	100%	0	1	100%

Legend:

ALL	Total item requirement in a claim		
1-P	Primary target set in claim 1	NBT	Number and operations in Base ten
1-S	Secondary target set in claim 1	NF	Number and operations—fractions
G	Geometry	OA	Operations and algebraic thinking
MD	Measurement and data	OTHER	Other content domains

Table 17. Mathematics Grades 6–7:
Percentage of Test Administrations Meeting Blueprint Requirements
for Claims and Content Domains (Braille)

Claim	Content Domain	Segment	Grade 6			Grade 7		
			Min.	Max.	%BP Match	Min.	Max.	%BP Match
1	ALL	Calc	6	6	100%	10	10	100%
1	P	Calc	3	3	100%	6	6	100%
1	S	Calc	3	3	100%	4	4	100%
1	ALL	NoCalc	13	13	100%	10	10	100%
1	P	NoCalc	11	11	96.0%	9	9	100%
1	S	NoCalc	2	2	96.0%	1	1	100%
2	ALL	Calc	3	3	100%	3	3	100%
2	EE	Calc	0	2	100%	0	2	100%
2	G	Calc	0	2	100%	0	2	100%
2	NS	Calc	0	2	100%	0	2	100%
2	RP	Calc	0	2	99.9%	0	2	100%
2	SP	Calc	0	2	100%	0	2	100%
2	OTHER	Calc						
3	All	Calc	7	7	100%	8	8	100%
3	EE	Calc	0	5	100%	1	5	100%
3	NS	Calc	2	6	100%	1	5	100%
3	RP	Calc	0	5	100%	1	5	100%
3	All	NoCalc	1	1	100%			
3	EE	NoCalc	0	1	100%			
3	NS	NoCalc	0	1	100%			
3	RP	NoCalc	0	1	100%			
4	All	Calc	3	3	100%	3	3	100%
4	EE	Calc	0	1	98.5%	0	1	90.8%
4	G	Calc	0	1	100%	0	1	100%
4	NS	Calc	0	1	94.2%	0	1	100%
4	RP	Calc	0	1	100%	0	1	99.6%
4	SP	Calc	0	1	100%	0	1	95.3%
4	OTHER	Calc						

Legend:

ALL	Total item requirement in a claim	N	Number and quantity
1-P	Primary target set in claim 1	NBT	Number and operations in Base ten
1-S	Secondary target set in claim 1	NF	Number and operations—fractions
A	Algebra	NS	Number system
EE	Expressions and equations	OA	Operations and algebraic thinking
F	Functions	OTHER	Other content domains
G	Geometry	RP	Ratios and proportional relationships
MD	Measurement and data	SP	Statistics and probability
Calc	Segment with calculator use	NoCalc	Segment without calculator use

Table 18. Mathematics Grades 8, 11:
Percentage of Test Administrations Meeting Blueprint Requirements
for Claims and Content Domains (Braille)

Grade 8						Grade 11					
Claim	**Content Domain**	**Segment**	**Min.**	**Max.**	**%BP Match**	**Claim**	**Content Domain**	**Segment**	**Min.**	**Max.**	**%BP Match**
1	ALL	Calc	14	14	100%	1	ALL	Calc	11	11	100%
1	P	Calc	11	11	100%	1	P	Calc	8	8	100%
1	S	Calc	3	3	100%	1	S	Calc	3	3	100%
1	ALL	NoCalc	6	6	100%	1	ALL	NoCalc	11	11	100%
1	P	NoCalc	4	4	100%	1	P	NoCalc	8	8	83.5%
1	S	NoCalc	2	2	100%	1	S	NoCalc	3	3	83.5%
2	ALL	Calc	3	3	100%	2	ALL	Calc	3	3	100%
2	EE	Calc	0	2	100%	2	A	Calc	1	2	99.8%
2	F	Calc	0	2	100%	2	F	Calc	0	2	100%
2	G	Calc	0	2	100%	2	G	Calc	0	2	100%
2	NS	Calc				2	N	Calc	0	2	100%
2	SP	Calc				2	S	Calc	0	2	100%
2	OTHER	Calc				2	O	Calc			
3	ALL	Calc	8	8	100%	3	All	Calc	7	7	100%
3	EE	Calc	1	5	100%	3	A	Calc	1	4	100%
3	F	Calc	1	5	100%	3	F	Calc	0	4	100%
3	G	Calc	1	5	100%	3	G	Calc	1	4	100%
						3	N	Calc	0	4	100%
						3	All	NoCalc	1	1	100%
						3	A	NoCalc	0	1	100%
						3	F	NoCalc	0	1	100%
						3	G	NoCalc	0	1	100%
						3	N	NoCalc	0	1	100%
4	ALL	Calc	3	3	100%	4	All	Calc	3	3	100%
4	EE	Calc	1	2	88.9%	4	A	Calc	0	2	100%
4	F	Calc	0	1	35.1%	4	F	Calc	0	1	98.5%
4	G	Calc	0	1	100%	4	G	Calc	0	1	99.5%
4	NS	Calc				4	N	Calc	0	2	100%
4	SP	Calc	0	1	100%	4	S	Calc	0	2	100%
4	OTHER	Calc				4	O	Calc			

Legend:

ALL	Total item requirement in a claim.	N	Number and quantity
1-P	Primary target set in claim 1	NBT	Number and operations in Base ten
1-S	Secondary target set in claim 1	NF	Number and operations—fractions
A	Algebra	NS	Number system
EE	Expressions and equations	OA	Operations and algebraic thinking
F	Functions	OTHER	Other content domains
G	Geometry	RP	Ratios and proportional relationships
MD	Measurement and data	SP	Statistics and probability
Calc	Segment with calculator use	NoCalc	Segment without calculator use

Table 19. Mathematics Grades 3–5:
Percentage of Test Administrations Meeting Blueprint Requirements
for Claims and Content Domains (Spanish)

Claim	Content Domain	Grade 3			Grade 4			Grade 5		
		Min.	Max.	%BP Match	Min.	Max.	%BP Match	Min.	Max.	%BP Match
1	ALL	20	20	100%	20	20	100%	20	20	100%
1	P	15	15	100%	15	15	100%	15	15	100%
1	S	5	5	100%	5	5	100%	5	5	100%
2	ALL	3	3	100%	3	3	100%	3	3	100%
2	G							0	2	100%
2	MD	0	2	100%	0	2	100%	0	2	100%
2	NBT	0	2	100%	0	2	100%	0	2	100%
2	NF	0	2	100%	1	3	100%	1	3	100%
2	OA	0	2	100%	0	2	100%			
3	All	8	8	100%	8	8	100%	8	8	100%
3	G							0	3	100%
3	MD	0	4	100%				0	4	100%
3	NBT				0	4	99.3%	0	4	100%
3	NF	2	6	100%	2	6	99.8%	2	6	99.4%
3	OA	0	4	100%	0	4	100%			
3	Other									
4	All	3	3	100%	3	3	100%	3	3	100%
4	G				0	1	100%	0	1	100%
4	MD	1	2	100%	0	2	100%	1	2	100%
4	NBT	0	1	100%	0	1	100%	0	1	100%
4	NF	0	1	100%	0	2	100%	1	2	100%
4	OA	1	2	100%	0	2	100%	0	1	100%

Legend:

ALL	Total item requirement in a claim.		
1-P	Primary target set in claim 1	NBT	Number and operations in Base ten
1-S	Secondary target set in claim 1	NF	Number and operations—fractions
G	Geometry	OA	Operations and algebraic thinking
MD	Measurement and data	OTHER	Other content domains

Table 20. Mathematics Grades 6–7:
Percentage of Test Administrations Meeting Blueprint Requirements
for Claims and Content Domains (Spanish)

Claim	Content Domain	Segment	Grade 6			Grade 7		
			Min.	Max.	%BP Match	Min.	Max.	%BP Match
1	ALL	Calc	6	6	100%	10	10	100%
1	P	Calc	3	3	99.9%	6	6	100%
1	S	Calc	3	3	99.9%	4	4	100%
1	ALL	NoCalc	13	13	100%	10	10	100%
1	P	NoCalc	11	11	95.6%	9	9	100%
1	S	NoCalc	2	2	95.6%	1	1	100%
2	ALL	Calc	3	3	100%	3	3	100%
2	EE	Calc	0	2	100%	0	2	100%
2	G	Calc	0	2	100%	0	2	100%
2	NS	Calc	0	2	100%	0	2	100%
2	RP	Calc	0	2	100%	0	2	100%
2	SP	Calc	0	2	100%	0	2	100%
2	OTHER	Calc	0	2	100%			
3	All	Calc	7	7	100%	8	8	100%
3	EE	Calc	0	5	100%	1	5	100%
3	NS	Calc	2	6	100%	1	5	100%
3	RP	Calc	0	5	100%	1	5	100%
3	All	NoCalc	1	1	100%			
3	EE	NoCalc	0	1	100%			
3	NS	NoCalc	0	1	100%			
3	RP	NoCalc	0	1	100%			
4	All	Calc	3	3	100%	3	3	100%
4	EE	Calc	0	1	96.2%	0	1	99.1%
4	G	Calc	0	1	100%	0	1	100%
4	NS	Calc	0	1	98.0%	0	1	100%
4	RP	Calc	0	1	92.4%	0	1	95.7%
4	SP	Calc	0	1	100%	0	1	99.8%
4	OTHER	Calc						

Legend:

ALL	Total item requirement in a claim	N	Number and quantity
1-P	Primary target set in claim 1	NBT	Number and operations in Base ten
1-S	Secondary target set in claim 1	NF	Number and operations—fractions
A	Algebra	NS	Number system
EE	Expressions and equations	OA	Operations and algebraic thinking
F	Functions	OTHER	Other content domains
G	Geometry	RP	Ratios and proportional relationships
MD	Measurement and data	SP	Statistics and probability
Calc	Segment with calculator use	NoCalc	Segment without calculator use

Table 21. Mathematics Grades 8, 11:
Percentage of Test Administrations Meeting Blueprint Requirements
for Claims and Content Domains (Spanish)

Grade 8						Grade 11					
Claim	**Content Domain**	**Segment**	**Min.**	**Max.**	**%BP Match**	**Claim**	**Content Domain**	**Segment**	**Min.**	**Max.**	**%BP Match**
1	ALL	Calc	14	14	100%	1	ALL	Calc	11	11	100%
1	P	Calc	11	11	100%	1	P	Calc	8	8	100%
1	S	Calc	3	3	100%	1	S	Calc	3	3	100%
1	ALL	NoCalc	6	6	100%	1	ALL	NoCalc	11	11	100%
1	P	NoCalc	4	4	100%	1	P	NoCalc	8	8	90.7%
1	S	NoCalc	2	2	100%	1	S	NoCalc	3	3	90.7%
2	ALL	Calc	3	3	100%	2	ALL	Calc	3	3	100%
2	EE	Calc	0	2	99.9%	2	A	Calc	1	2	99.9%
2	F	Calc	0	2	100%	2	F	Calc	0	2	100%
2	G	Calc	0	2	100%	2	G	Calc	0	2	100%
2	NS	Calc				2	N	Calc	0	2	100%
2	SP	Calc	0	2	100%	2	S	Calc	0	2	100%
2	OTHER	Calc				2	O	Calc			
3	ALL	Calc	8	8	100%	3	All	Calc	7	7	100%
3	EE	Calc	1	5	95.1%	3	A	Calc	1	4	100%
3	F	Calc	1	5	100%	3	F	Calc	0	4	100%
3	G	Calc	1	5	100%	3	G	Calc	1	4	100%
						3	N	Calc	0	4	100%
						3	All	NoCalc	1	1	100%
						3	A	NoCalc	0	1	100%
						3	F	NoCalc	0	1	100%
						3	G	NoCalc	0	1	100%
						3	N	NoCalc	0	1	100%
4	ALL	Calc	3	3	100%	4	All	Calc	3	3	100%
4	EE	Calc	1	2	98.6%	4	A	Calc	0	2	99.9%
4	F	Calc	0	1	100%	4	F	Calc	0	1	97.7%
4	G	Calc	0	1	100%	4	G	Calc	0	1	97.8%
4	NS	Calc				4	N	Calc	0	2	100%
4	SP	Calc	0	1	97.7%	4	S	Calc	0	2	100%
4	OTHER	Calc				4	O	Calc			

Legend:

ALL	Total item requirement in a claim	N	Number and quantity
1-P	Primary target set in claim 1	NBT	Number and operations in Base ten
1-S	Secondary target set in claim 1	NF	Number and operations—fractions
A	Algebra	NS	Number system
EE	Expressions and equations	OA	Operations and algebraic thinking
F	Functions	OTHER	Other content domains
G	Geometry	RP	Ratios and proportional relationships
MD	Measurement and data	SP	Statistics and probability
Calc	Segment with calculator use	NoCalc	Segment without calculator use

Table 22. Mathematics:
Percentage of Test Administrations Meeting Blueprint Requirements
for DoK and Targets

Grade	Additional Constraints	Min	Max	%BP Match
	English			
8	Claim 2/4 DOK3 or higher	2	5	99.9
	Braille			
3	Claim 2/4 DOK3 or higher	2	5	99.9
4	Claim 2/4 DOK3 or higher	2	5	99.9
5	Claim 2/4 DOK3 or higher	2	5	99.1
6	Claim 1 DOK1	2	2	95.5
6	Claim 2/4 DOK3 or higher	2	5	97.6
6	Claim 3 Target A,D	3	3	41.3
6	Claim 3 Target C,F,G	2	2	96.5
8	Claim 2/4 DOK3 or higher	2	5	87.7
	Spanish			
5	Claim 2/4 DOK3 or higher	2	5	98.6
6	Claim 1 DOK1	2	2	82.1
6	Claim 2/4 DOK3 or higher	2	5	98.3
6	Claim 3 Target A,D	3	3	54.5
6	Claim 3 Target C,F,G	2	2	98.0
8	Claim 2/4 DOK3 or higher	2	5	99.6
11	Claim 3 Target B,E	3	3	99.9

The Smarter Balanced blueprints for mathematics do not require every target to be covered in a claim and content domain; therefore, all targets listed in the blueprint are not expected to be covered in every test. Table 23 summarizes the number of unique targets administered in each simulated test by claim. The table includes the number of targets specified in the blueprints and in the item pool, and the mean and range of the number of targets administered to the simulated test administrations. Individual tests do not cover all targets specified in the blueprint, but all targets are covered at an aggregate level, across all simulated tests combined.

Table 23. Mathematics: Number of Unique Targets Assessed Within Each Claim

Grade	Total Targets Specified in BP				Mean				Range (Minimum-Maximum)			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
English												
3	11	4	6	6	10.8	2	5.6	3	10-11	2-2	4-6	3-3
4	12	4	6	6	10.0	2	5.5	3	10-10	2-2	4-6	3-3
5	11	4	6	6	9.0	2	5.3	3	9-9	2-2	3-6	3-3
6	10	4	7	6	10.0	2	4.8	3	9-10	2-2	3-7	3-3
7	9	4	7	6	8.0	2	5.0	3	8-8	2-2	3-6	3-3
8	10	4	7	6	10.0	2	4.8	3	10-10	2-2	3-6	3-3
11	16	4	7	6	14.8	2	5.0	3	14-15	2-2	3-7	3-3
Braille												
3	11	4	6	6	9.9	2	4.9	3	9-10	2-2	3-6	3-3
4	12	4	6	6	10.0	2	5.3	3	10-10	2-2	3-6	3-3
5	10	4	6	6	9.0	2	5.3	3	9-9	2-2	3-6	3-3
6	10	3	7	6	9.7	2	4.0	3	8-10	2-2	3-6	3-3
7	9	3	7	5	8.0	2	5.0	3	8-8	2-2	3-6	3-3
8	10	3	7	6	10.0	2	4.6	3	10-10	2-2	3-6	3-3
11	16	4	7	6	14.8	2	4.6	3	13-16	2-2	3-6	3-3
Spanish												
3	11	4	6	6	9.9	2	5.1	3	9-10	2-2	3-6	3-3
4	12	4	6	6	10.0	2	5.5	3	10-10	2-2	3-6	3-3
5	11	4	6	6	9.0	2	5.2	3	9-9	2-2	3-6	3-3
6	10	3	7	6	9.7	2	4.3	3	8-10	2-2	3-6	3-3
7	9	3	7	5	8.0	2	5.1	3	8-8	2-2	3-6	3-3
8	10	4	7	6	10.0	2	5.3	3	10-10	2-2	4-6	3-3
11	16	4	7	6	14.9	2	4.9	3	13-16	2-2	3-7	3-3

5.2 Summary Statistics of the Ability Estimation

Each simulated test includes an initial ability, a true ability, and an estimated ability based on the adaptive test administration. Three pairs of correlations are examined: (1) correlation between the initial ability and the first item difficulty, (2) the correlation between the true ability and the estimated ability, and (3) the correlation between the estimated ability and the average item difficulty (form difficulty) for each simulated test.

The correlation between the initial ability and the first item difficulty is expected to be close to zero or very small since the first item was set to be selected randomly from the pool in order to mitigate item exposure rates. The correlations between estimated ability and true ability (a reliability index) are expected to be high, indicating that the adaptive test administrations reliably estimate student ability. The correlations between the estimated ability and the average difficulty (form difficulty) of the test administered to each student are also expected to be high. The higher the correlations are, the more adaptive the assessment is. The high correlations demonstrate that the algorithm adapt to student ability efficiently while matching the blueprint specifications.

5.2.1 Summary Statistics of the Ability Estimation for ELA/L

Table 24 shows the correlation between the initial ability and the first item difficulty, the correlation between the true ability and the estimated ability, and the correlation between the estimated ability and the average item difficulty (form difficulty) for each simulated test.

The correlation between the initial ability and the first item difficulty is close to zero, as expected, because the pool size was used for the parameter *k* to mitigate the number of unused items. Although the first item was selected randomly, the selection of the first item doesn't have a large effect on the final score because the algorithm recovers students' ability quickly as students take more items. This is evidenced in the high correlations between the estimated ability and the average item difficulty (form difficulty) in Table 24, ranging from 0.91 to 0.94 in English and from 0.85 to 0.94 in Braille. The lower correlations in grade 11 relative to the other assessments, is likely due to a mismatch between the difficulty of items in the pool and the ability of the student population, resulting in less information for the estimation of achievement for low-ability students, and thus less reliable ability estimation for those students. The correlations between estimated ability and true ability, a reliability index, are high, indicating that the adaptive test administrations reliably estimate student ability.

Table 24. ELA/L Correlations Between First Item Difficulty and Initial Ability, Between True Ability and Estimated Ability, and Between Estimated Ability and Average Item Difficulty for Simulated Test Administrations

Grade	First Item Difficulty and Initial Ability	True Ability and Estimated Ability	Estimated Ability and Average Item Difficulty
English			
3	-0.02	0.95	0.91
4	0.03	0.95	0.94
5	0.06	0.95	0.93
6	0.01	0.95	0.94
7	-0.04	0.95	0.94
8	0.05	0.95	0.93
11	0.02	0.95	0.92
Braille			
3	0.04	0.96	0.90
4	0.03	0.95	0.92
5	0.04	0.96	0.92
6	-0.02	0.95	0.92
7	0.01	0.95	0.94
8	-0.01	0.94	0.90
11	0.02	0.93	0.85

Table 25 presents statistical summaries of the ability estimation (the mean of the biases) which is the average of the biases of estimated abilities across all students; the standard error of the mean bias; and the *p*-value for the significance of the estimated bias reported from the *z* test. The mean square error and the percentage of students' estimated theta falling outside the 95% and 99% confidence interval coverage are also provided in the table. All statistics computed in Table 25 are described in section 4 - Statistical Summary Indices.

The average bias of the estimated abilities across all students is very small and statistically insignificant, providing evidence that the true score is adequately recovered in the estimated score, except for some tests

in the low and high ends of the ability range. The distributions of bias across the estimated ability range are provided in Figures 8–9. The vertical dashed lines indicate the achievement standards (cuts).

Table 25. ELA/L: Mean Bias of the Ability Estimates (True Score – Observed Score)

Grade	Mean of the Biases	SE of the Biases	*P*-value for the *Z*-Test	MSE	95% Coverage	99% Coverage
English						
3	-0.01	0.01	0.35	0.10	4.5	0.7
4	-0.01	0.01	0.26	0.12	4.7	1.0
5	-0.01	0.01	0.32	0.12	4.4	1.1
6	-0.02	0.01	0.15	0.13	4.5	0.6
7	-0.01	0.01	0.48	0.15	5.7	1.4
8	0.02	0.01	0.13	0.14	4.9	0.8
11	0.01	0.01	0.27	0.16	4.1	0.7
Braille						
3	-0.01	0.01	0.50	0.11	4.4	1.2
4	-0.02	0.01	0.07	0.12	4.4	0.6
5	0.00	0.01	0.98	0.11	4.2	0.8
6	-0.02	0.01	0.06	0.14	5.1	0.9
7	-0.01	0.01	0.44	0.15	4.7	0.5
8	0.00	0.01	0.96	0.16	5.1	0.8
11	0.00	0.01	0.93	0.21	6.0	1.3

Figure 8. ELA/L Distribution of Bias Across Estimated Abilities (English)

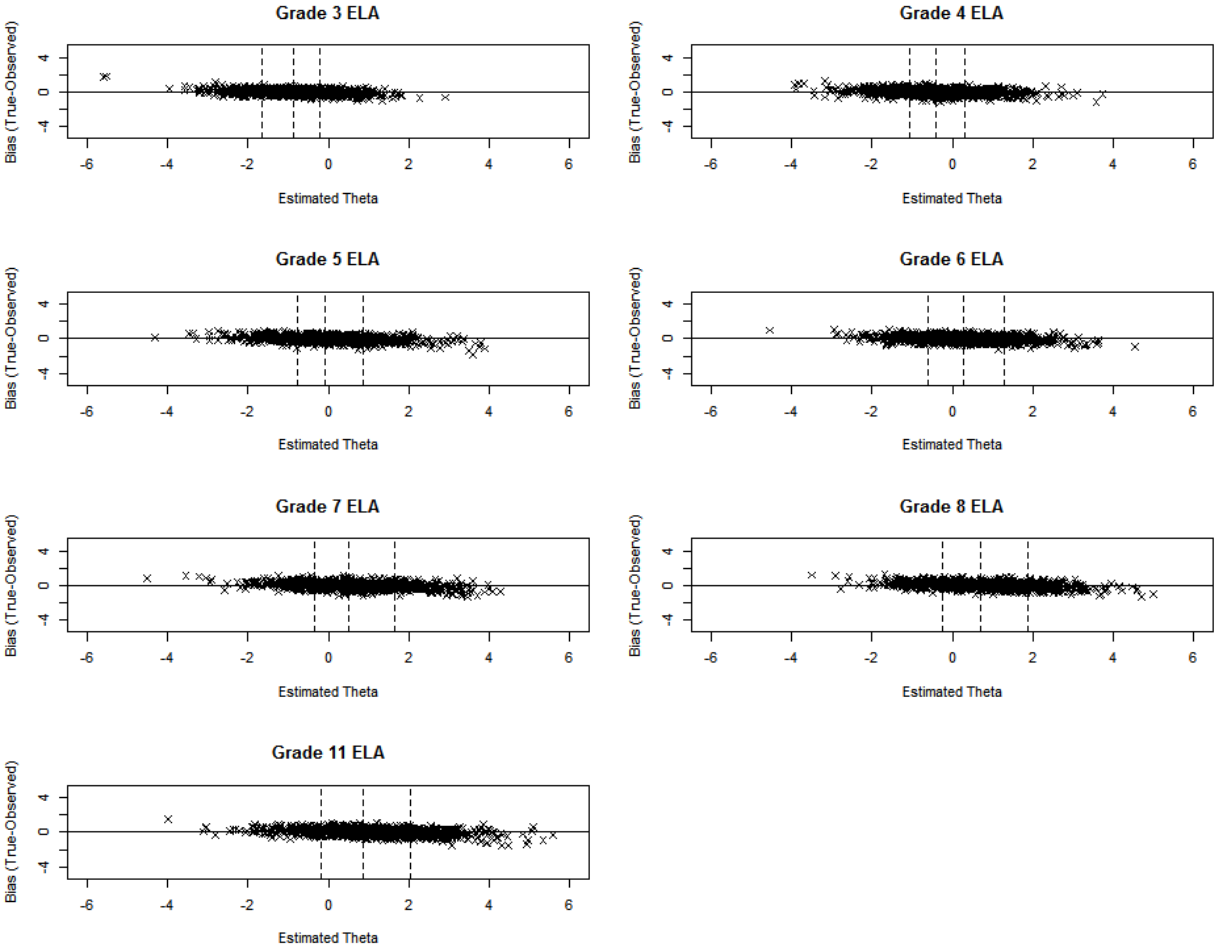


Figure 9. ELA/L Grades 3–6: Distribution of Bias Across Estimated Abilities (Braille)

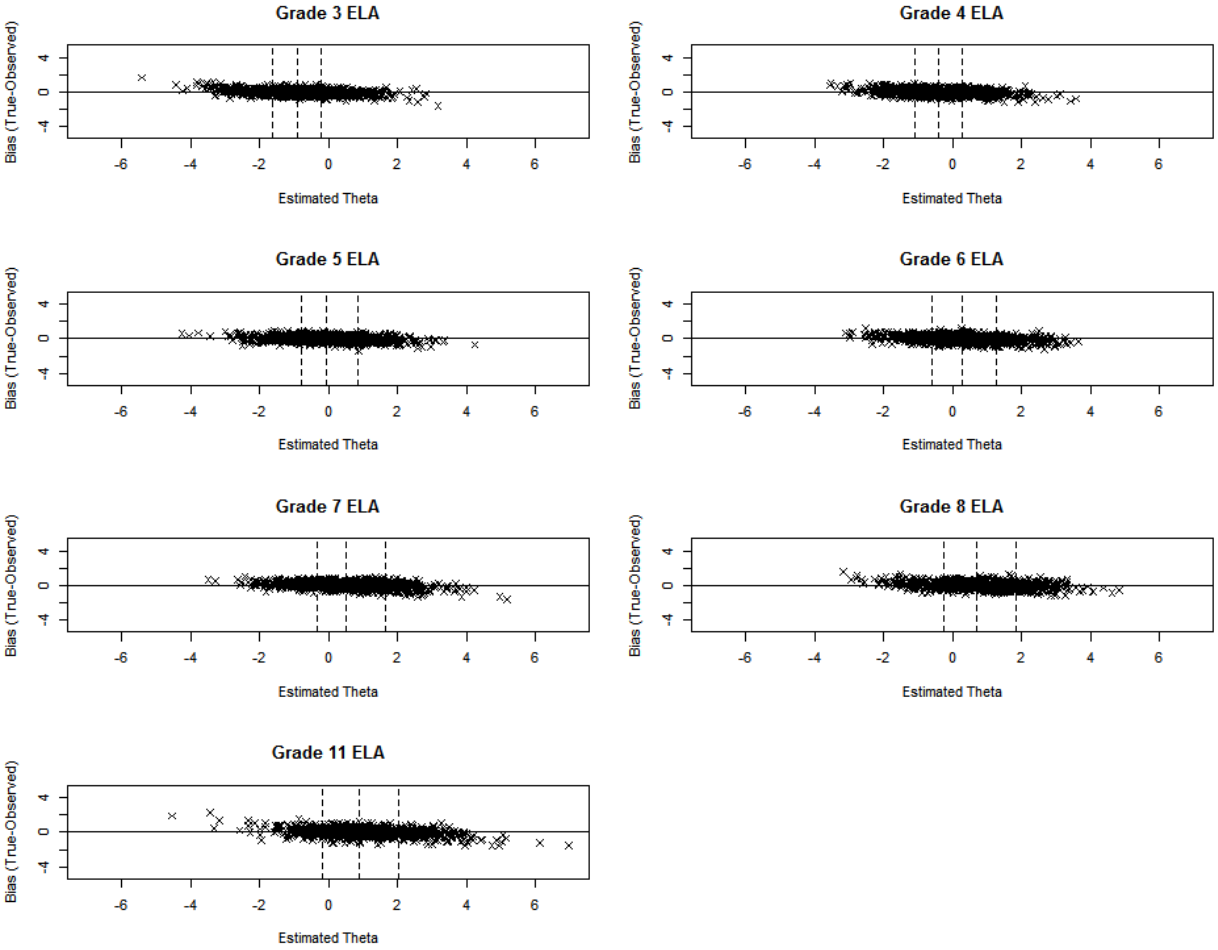


Table 26 presents the average standard error of the ability estimates across simulated test administrations, as well as the standard error at 5th, 25th, 75th and 95th percentiles of the ability distribution. The standard error is largest at the high end of the ability range in both the English and Braille pools, except for Braille grade 11. The average standard errors are slightly larger in Braille tests than English tests, which can be attributed to the smaller item pool in Braille. The standard error curves are shown in Figures 10–11.

Table 26. ELA/L: Mean Standard Error of the Ability Estimates Across the Ability Distribution

Grade	Average SE	SE at 5th Percentile	SE at Bottom Quartile (25th)	SE at Top Quartile (75th)	SE at 95th Percentile
English					
3	0.31	0.32	0.28	0.28	0.35
4	0.35	0.37	0.33	0.35	0.38
5	0.34	0.34	0.33	0.36	0.36
6	0.36	0.36	0.35	0.34	0.39
7	0.37	0.38	0.36	0.35	0.41
8	0.38	0.38	0.37	0.35	0.43
11	0.41	0.42	0.35	0.40	0.43
Braille					
3	0.33	0.37	0.28	0.32	0.36
4	0.35	0.36	0.35	0.34	0.37
5	0.34	0.34	0.31	0.36	0.35
6	0.37	0.37	0.33	0.38	0.40
7	0.39	0.38	0.36	0.39	0.43
8	0.40	0.43	0.35	0.37	0.45
11	0.45	0.57	0.41	0.41	0.48

Figure 10. ELA/L Standard Error of Measurements Across Estimated Theta Range (English)

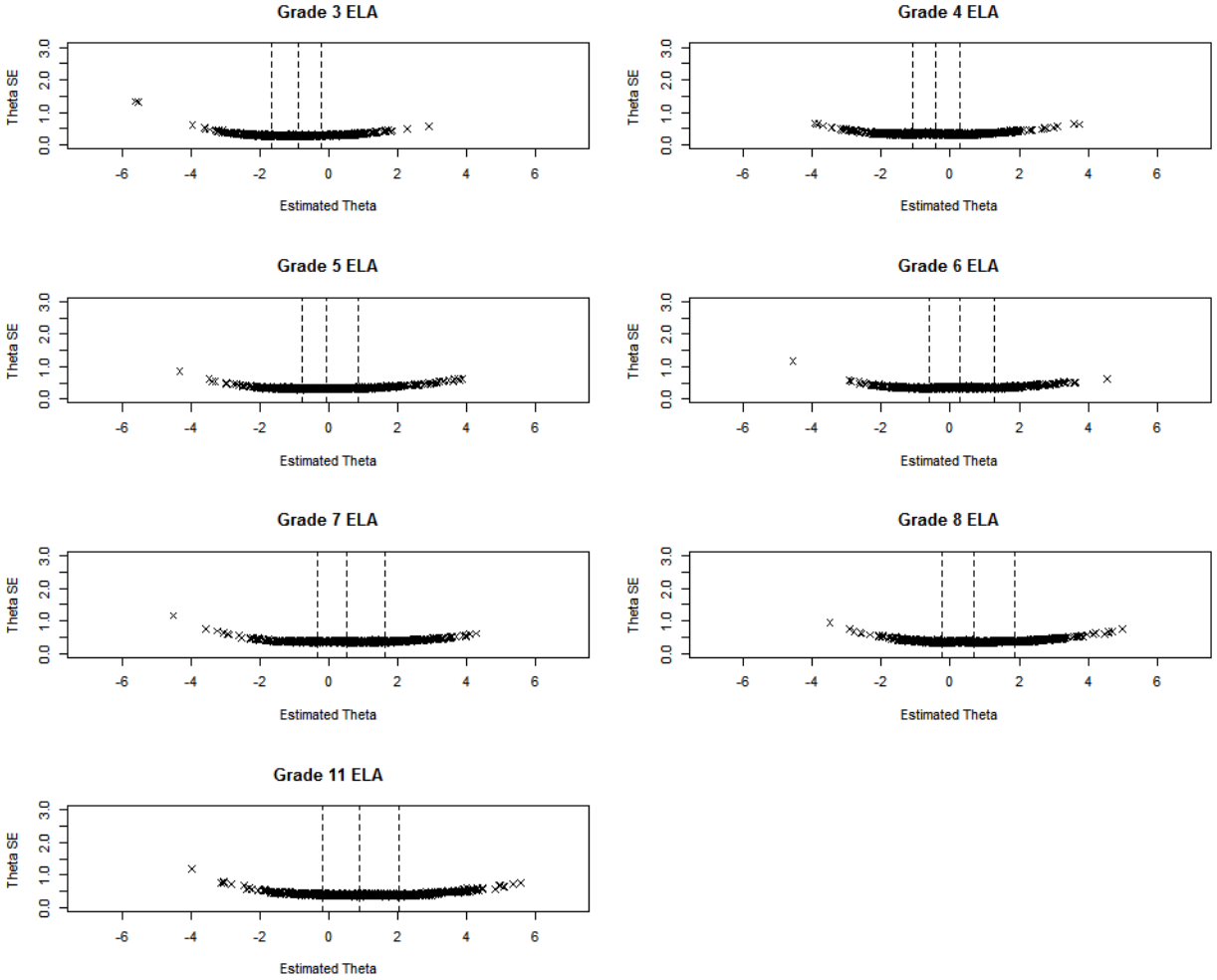
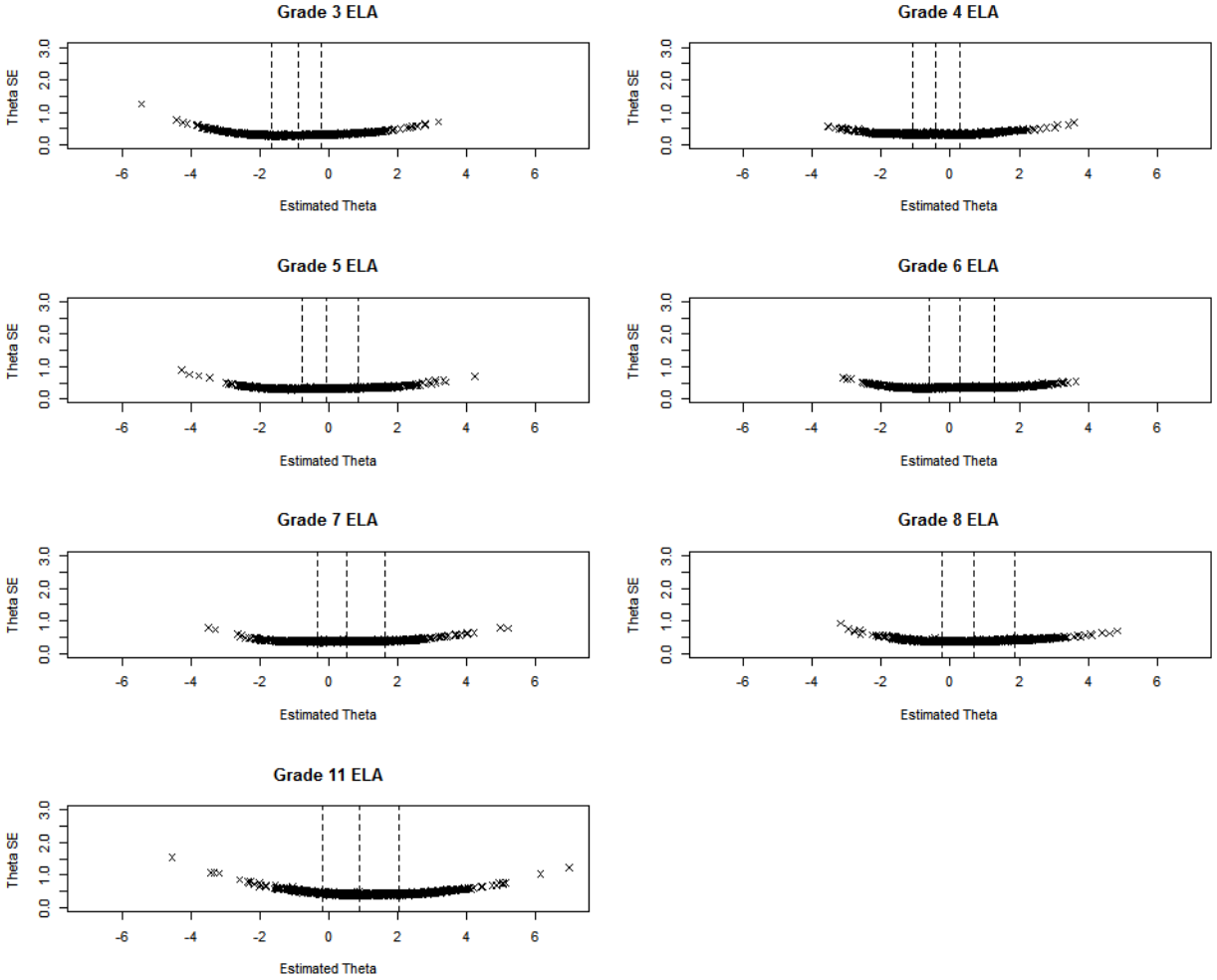


Figure 11. ELA/L Standard Error of Measurements Across Estimated Theta Range (Braille)



5.2.2 Summary Statistics of the Ability Estimation for Mathematics

Table 27 presents the correlation between the initial ability and the first item difficulty, the correlation between the true ability and the estimated ability, and the correlation between the estimated ability and the average item difficulty (form difficulty) across simulated tests.

The correlation between the initial ability and the first item difficulty is close to zero, as expected, because the pool size was used for the parameter *k* to mitigate the number of unused items. Although the first item was selected randomly, the selection of the first item doesn't have a large effect on the final score because the algorithm recovers students' ability quickly, as students take more items. This is evidenced by the high correlations between the estimated ability and the average item difficulty (form difficulty) in Table 27, ranging from 0.92 to 0.96 in English, from 0.80 to 0.93 in Braille, and from 0.81 to 0.92 in Spanish.

The lower correlations in grades 7, 8, and 11, relative to other grades, are likely due to a mismatch between the difficulty of items in the pool and the ability of the student population, resulting in less information for the estimation of achievement for low-performing students, and thus a less reliable ability estimation for those students. The correlations between estimated ability and true score (a reliability index) are high, indicating that the adaptive test administrations reliably estimate student ability.

Table 27. Mathematics: Correlations Between First Item and Initial Ability, Between True Ability and Estimated Ability, and Between Estimated Ability and Average Item Difficulty for Simulated Test Administrations

Grade	First Item and Initial Ability	True Ability and Estimated Ability	Estimated Ability and Average Item Difficulty
English			
3	0.02	0.97	0.96
4	-0.03	0.97	0.94
5	-0.06	0.97	0.95
6	0.03	0.97	0.95
7	-0.01	0.97	0.94
8	0.00	0.97	0.92
11	-0.01	0.97	0.92
Braille			
3	0.00	0.96	0.92
4	0.00	0.96	0.92
5	0.02	0.96	0.92
6	0.04	0.97	0.93
7	-0.05	0.96	0.91
8	-0.01	0.96	0.88
11	0.00	0.95	0.80
Spanish			
3	0.00	0.97	0.92
4	-0.03	0.96	0.89
5	-0.04	0.96	0.91
6	-0.04	0.97	0.92
7	0.01	0.96	0.89
8	0.00	0.95	0.89
11	-0.02	0.94	0.81

Table 28 presents statistical summaries of the ability estimation (the mean of the biases) which is the average of the biases of estimated abilities across all students; the standard error of the mean bias; and the *p*-value for the significance of the estimated bias reported from the *z* test. The mean square error and the percentage of students' estimated theta falling outside the 95% and 99% confidence interval coverage are also summarized.

The average bias of the estimated abilities across all students is very small and statistically insignificant, providing evidence that the true score is adequately recovered in the estimated score, except for the low end of the ability range, especially in upper grades. The distributions of bias across the estimated ability range are provided in Figures 12–14. The vertical dashed lines indicate the achievement standards (cut scores).

Table 28. Mathematics: Mean Bias of the Ability Estimates (True Score – Observed Score)

Grade	Mean of the Biases	SE of the Biases	*P*-value for the *Z*-Test	MSE	95% Coverage	99% Coverage
			English			
3	0.00	0.01	0.70	0.06	5.1	0.8
4	0.01	0.01	0.16	0.07	4.0	1.0
5	-0.01	0.01	0.43	0.09	5.7	1.4
6	0.01	0.01	0.17	0.11	4.8	0.7
7	0.00	0.01	0.91	0.13	5.9	1.0
8	0.02	0.01	0.15	0.18	5.4	1.3
11	0.03	0.01	0.06	0.18	4.5	1.1
			Braille			
3	0.00	0.01	0.82	0.08	4.8	1.0
4	0.01	0.01	0.29	0.08	4.9	1.2
5	0.01	0.01	0.29	0.10	3.5	0.5
6	0.01	0.01	0.42	0.12	4.8	1.1
7	0.01	0.01	0.24	0.15	4.8	1.2
8	0.05	0.01	0.00	0.22	4.1	1.1
11	0.08	0.02	0.00	0.34	2.7	0.2
			Spanish			
3	0.01	0.01	0.29	0.07	3.3	0.4
4	0.00	0.01	0.70	0.08	5.0	0.7
5	0.04	0.01	0.00	0.12	4.9	1.1
6	0.02	0.01	0.05	0.13	3.6	0.7
7	0.02	0.01	0.16	0.15	5.3	1.2
8	0.04	0.02	0.01	0.23	6.0	0.6
11	0.07	0.02	0.00	0.34	5.9	0.9

Figure 12. Mathematics Distribution of Bias Across Estimated Abilities (English)

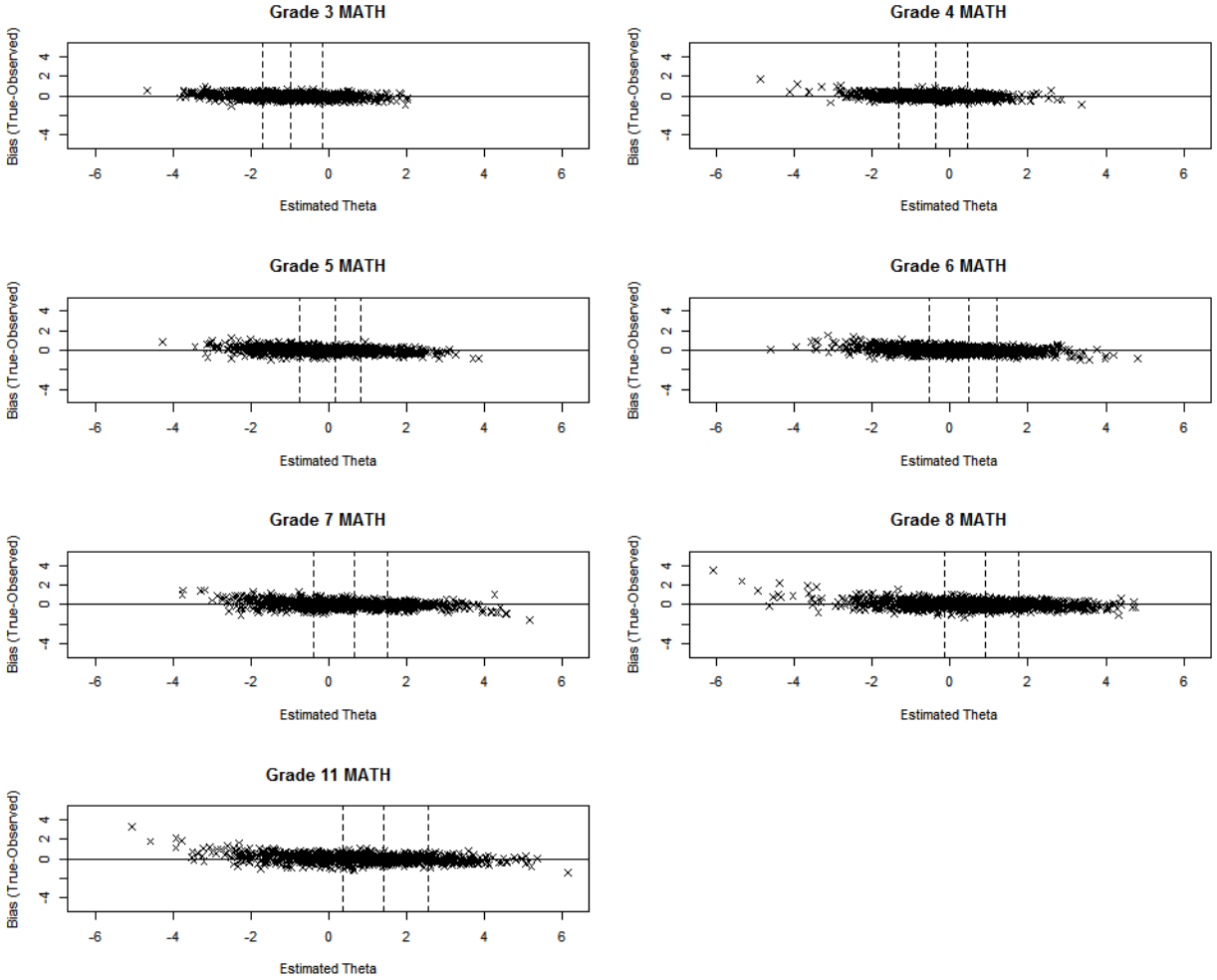


Figure 13. Mathematics Distribution of Bias Across Estimated Abilities (Braille)

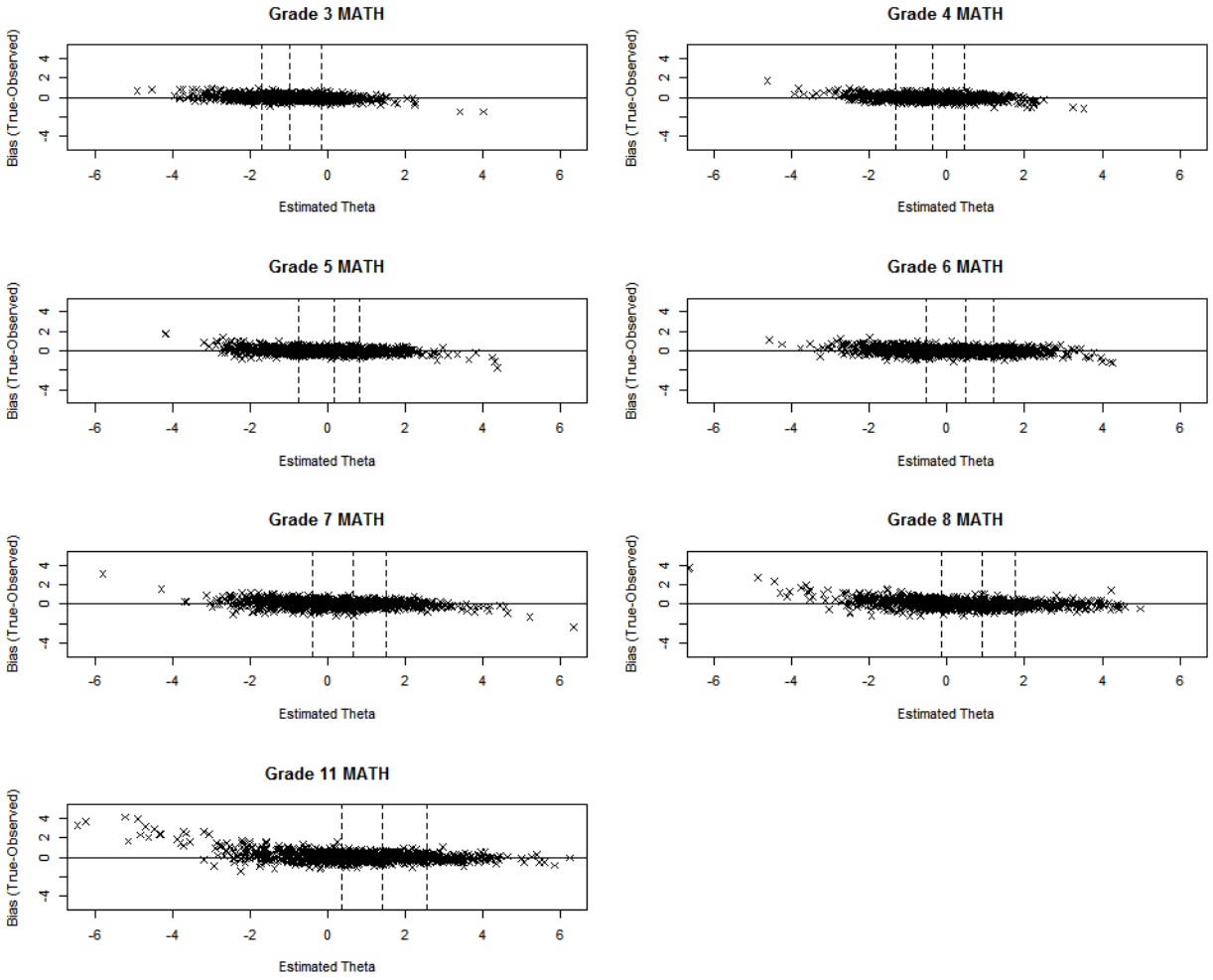


Figure 14. Mathematics Distribution of Bias Across Estimated Abilities (Spanish)

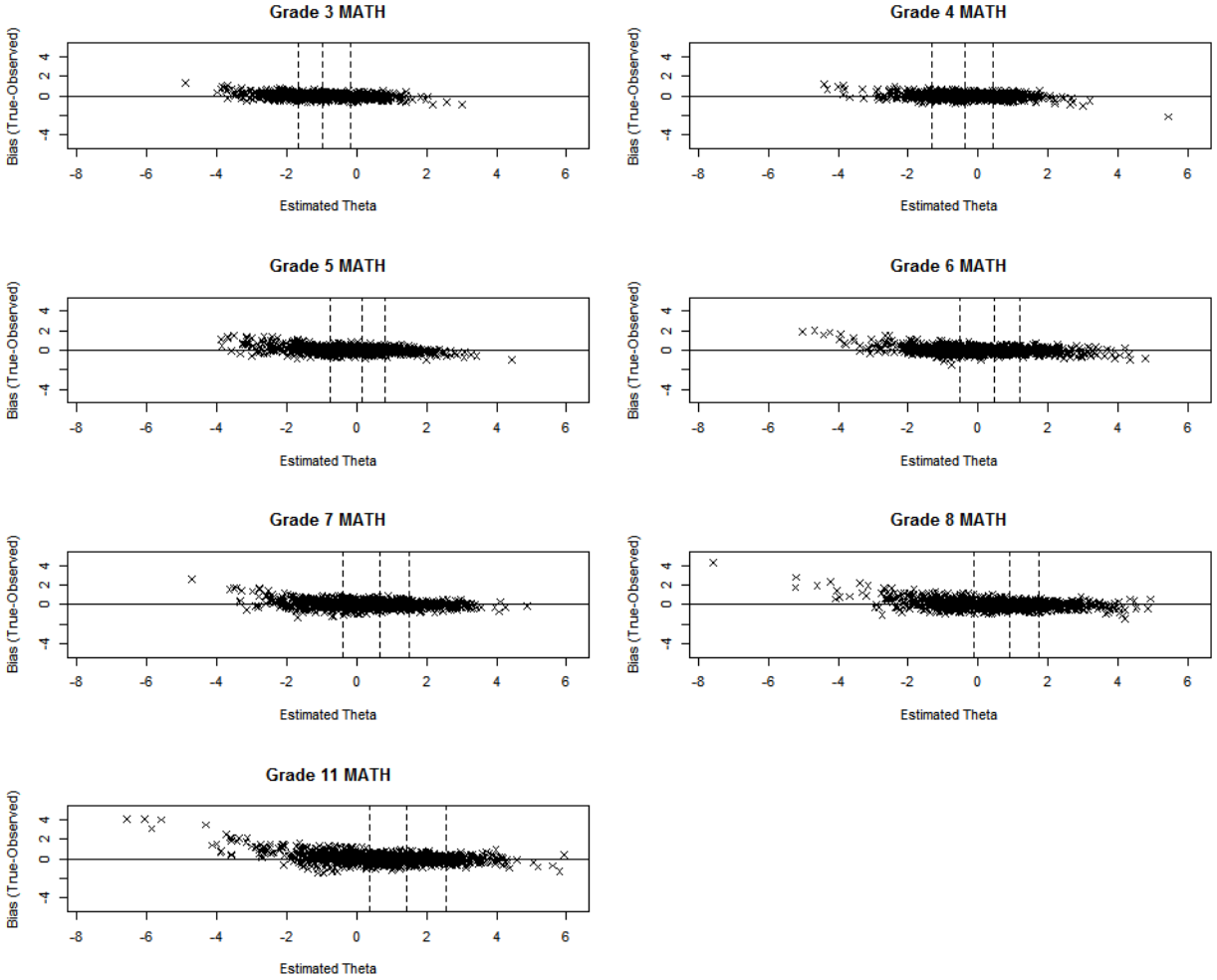


Table 29 presents the average standard error of the ability estimates across simulated test administrations, as well as the standard error at 5th, 25th, 75th, and 95th percentiles of the ability distribution. The standard error is greatest at the low end of the ability range, and becomes smaller as the ability group goes up, which indicates a shortage of easy items for low-performing students. Similar to ELA/L, the average standard errors are larger for the mathematics tests in Braille and Spanish than for the tests in English, due to the smaller item pools in Braille and Spanish. The standard error curves are provided in Figures 15–17.

Table 29. Mathematics: Mean Standard Error of the Ability Estimates Across the Ability Distribution

Grade	Average SE	SE at 5th Percentile	SE at Bottom Quartile (25th)	SE at Top Quartile (75th)	SE at 95th Percentile
English					
3	0.25	0.31	0.25	0.23	0.23
4	0.25	0.29	0.24	0.24	0.25
5	0.29	0.40	0.33	0.23	0.24
6	0.32	0.42	0.34	0.27	0.29
7	0.34	0.43	0.37	0.26	0.29
8	0.38	0.54	0.40	0.30	0.28
11	0.40	0.64	0.43	0.32	0.33
Braille					
3	0.27	0.31	0.28	0.22	0.25
4	0.28	0.34	0.29	0.25	0.27
5	0.31	0.44	0.33	0.29	0.26
6	0.34	0.46	0.34	0.28	0.35
7	0.37	0.55	0.45	0.28	0.27
8	0.42	0.58	0.47	0.33	0.28
11	0.51	0.99	0.55	0.36	0.43
Spanish					
3	0.26	0.32	0.25	0.24	0.28
4	0.28	0.33	0.27	0.26	0.26
5	0.32	0.56	0.32	0.26	0.26
6	0.35	0.50	0.38	0.27	0.27
7	0.37	0.53	0.40	0.29	0.28
8	0.42	0.69	0.48	0.31	0.28
11	0.48	0.77	0.53	0.33	0.29

Figure 15. Mathematics Standard Error of Measurements Across Estimated Theta Range (English)

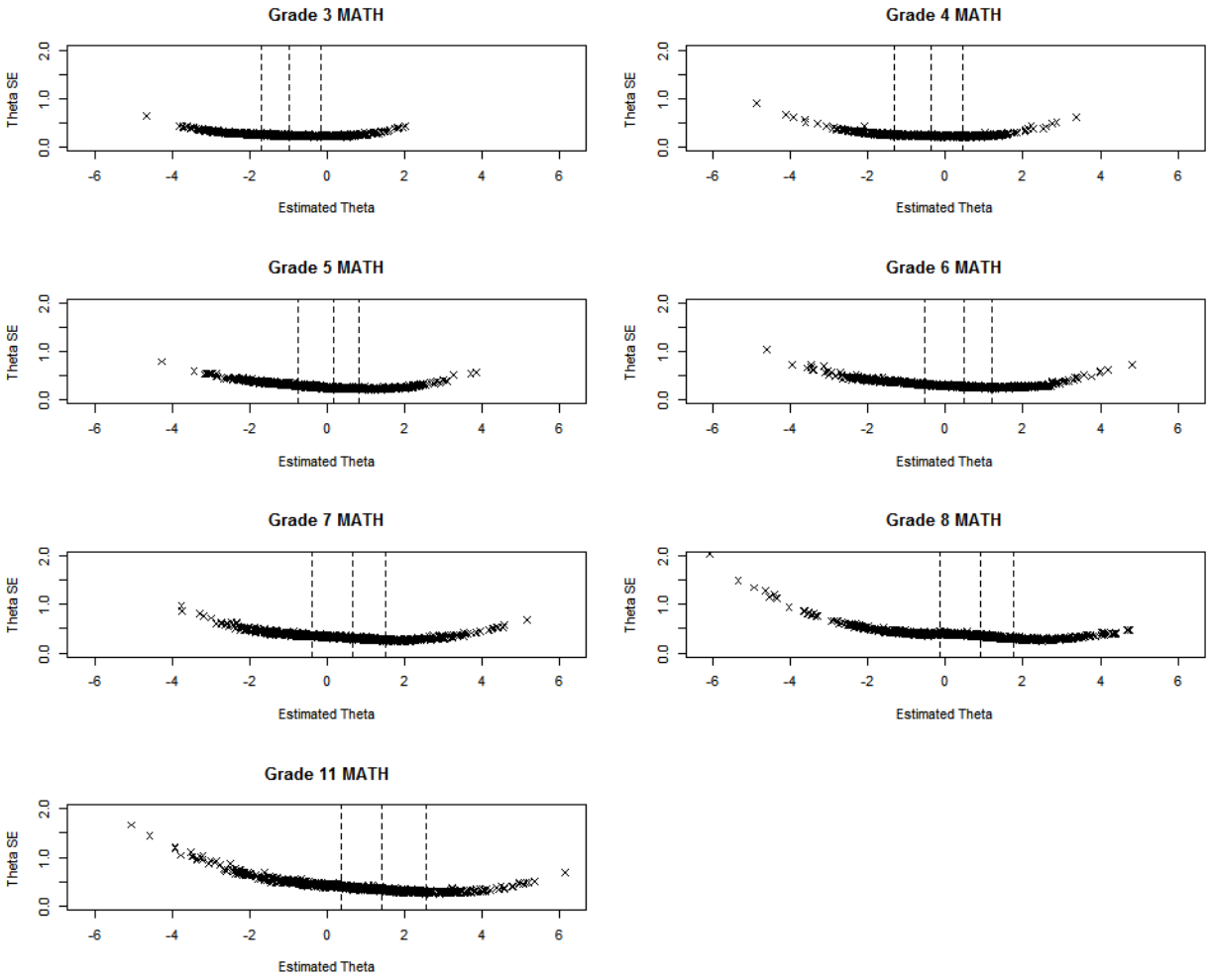


Figure 16. Mathematics Standard Error of Measurements Across Estimated Theta Range (Braille)

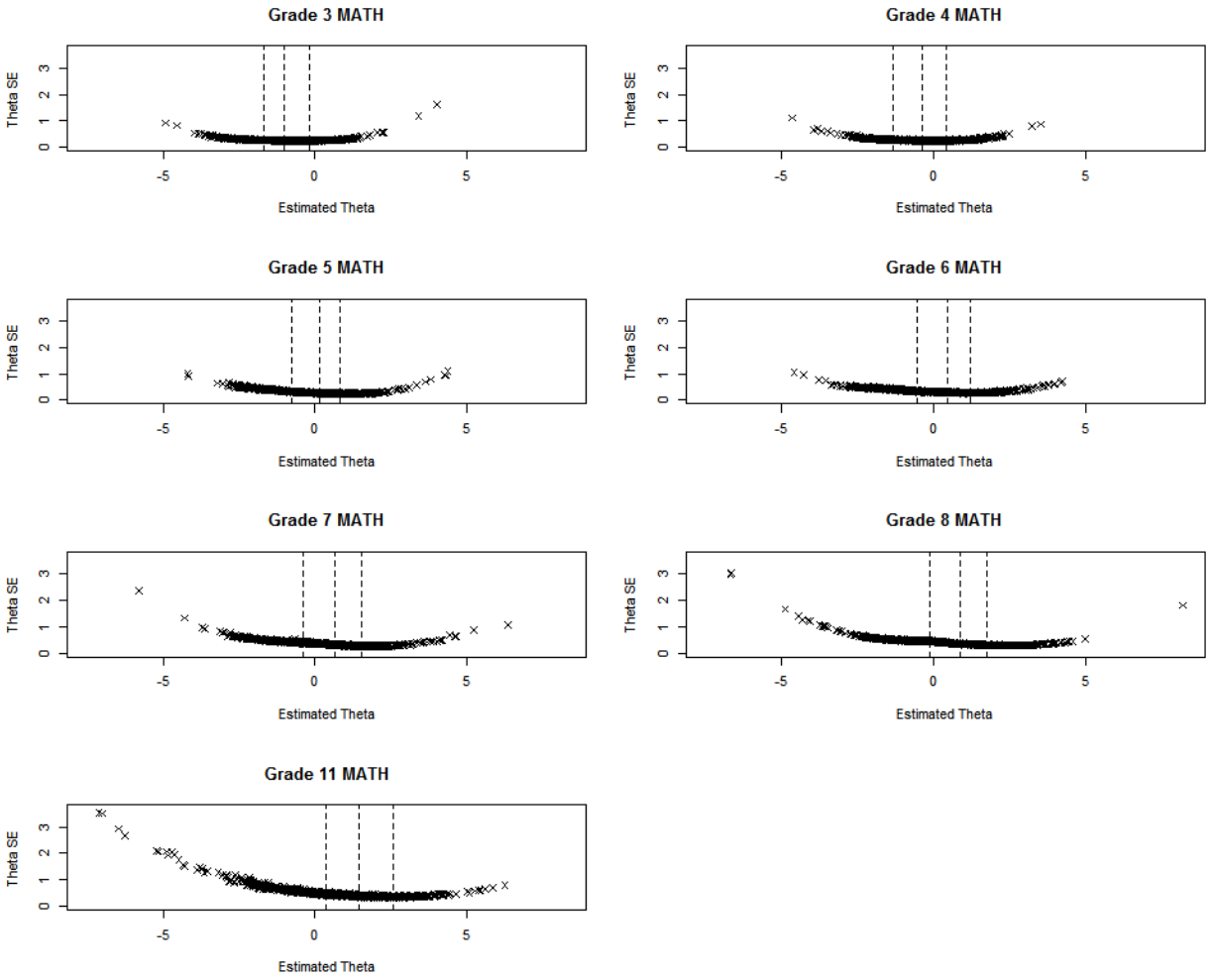
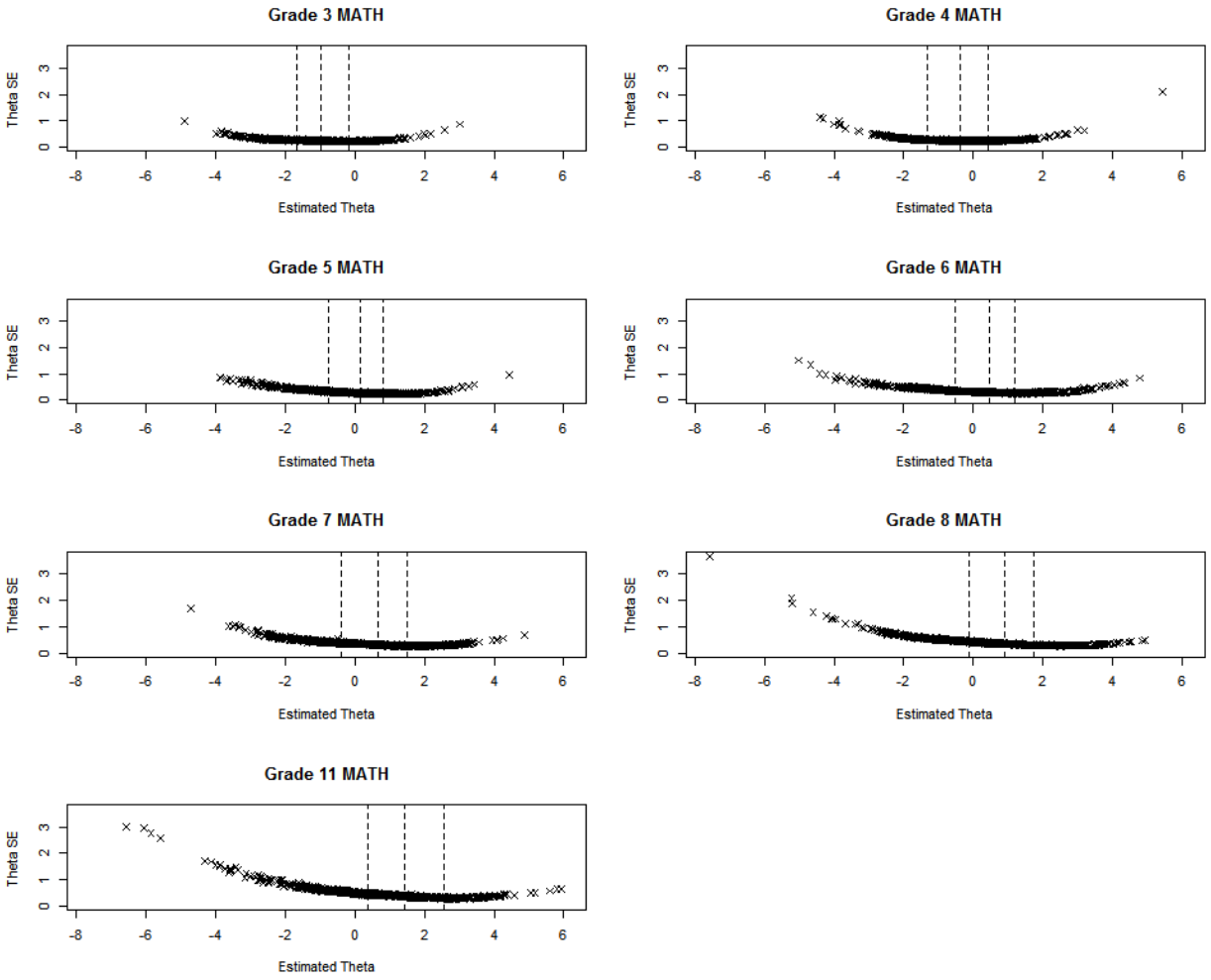


Figure 17. Mathematics Standard Error of Measurements Across Estimated Theta Range (Spanish)



5.3 Item Exposure Rates

The simulator output also reports the degree to which the constraints set forth in the blueprints may yield greater exposure of items to students. This is reported by examining the percentage of test administrations in which an item appears. In an adaptive test with a sufficiently large item pool, where the items are distributed in proportion to the blueprint constraints, we would expect most of the items to appear in only a relatively small percentage of the test administrations. When this condition holds, it suggests that test administrations between students are unique.

The item exposure rate for each item is computed by dividing the total number of test administrations in which an item appears, by the total number of tests administered. The distribution of the item exposure rates (r) is summarized in six bins: $r = 0\%$ (unused), $0\% < r \leq 20\%$, $20\% < r \leq 40\%$, $40\% < r \leq 60\%$, $60\% < r \leq 80\%$ and $80\% < r \leq 100\%$. If item exposure is minimal, we would expect the largest portion of items to appear in the $0\% < r \leq 20\%$ bin, an indication that most of the items appear on a very small percentage of the test forms.

For English tests, the exposure rates are computed for 3,000 simulated tests in grades 3–8, and 5,000 simulated tests in grade 11, accommodating large pool sizes. For Braille tests, the exposure rates are computed for 1,000 simulated tests in all grades. The distribution of exposure rates is as expected, given the item distributions in the blueprint constraints. Table 30 presents the percentage of items that fall into each exposure bin in ELA/L for the English and Braille pools. The unused item rates range from 2.9% to 8.7% in English tests, and from 3.0% to 9.8% in Braille tests. For the used items, most test items are administered in 20% or less of test administrations. There are a few items in Braille with exposure rates 81%–100% because the pool has too few items to meet some blueprint constraints.

Table 30. ELA/L: Percent of Pool Items Classified at each Exposure Rate

Grade	Total Items	Exposure Rate (%)					
		Unused	0–20	21–40	41–60	61–80	81–100
English							
3	890	3.2	93.3	3.0	0.6	0.0	0.0
4	873	2.9	93.7	3.1	0.3	0.0	0.0
5	886	5.8	90.5	2.9	0.8	0.0	0.0
6	826	6.1	90.6	2.4	0.4	0.6	0.0
7	763	8.7	85.6	4.5	1.3	0.0	0.0
8	808	4.1	89.5	5.7	0.7	0.0	0.0
11	2463	5.4	94.5	0.0	0.1	0.0	0.0
Braille							
3	300	3.0	73.3	19.3	1.3	2.0	1.0
4	306	6.9	70.3	19.3	2.3	1.3	0.0
5	337	4.5	75.1	15.4	4.8	0.3	0.0
6	306	9.8	68.0	14.7	6.2	0.7	0.7
7	287	9.4	65.2	19.5	3.8	2.1	0.0
8	305	3.0	71.2	19.7	6.2	0.0	0.0
11	540	6.5	85.0	8.5	0.0	0.0	0.0

Table 31 presents the percentage of items that fall into each exposure bin in mathematics for the English, Braille, and Spanish pools. Almost all items are used, with small unused item rates, ranging from 0% to 2.2% in English, from 0% to 4.7% in Braille, and 0% to 4.4% in Spanish, across all grades and segments. For the used items, most test items are administered in 20% or less of test administrations. A few items in

Braille and Spanish have exposure rates of 81%–100% because the pool has too few items to meet some blueprint constraints.

To further investigate the item usage across testers, the number of unique items administered by item position for simulated examinees are presented in Appendix C.

Table 31. Mathematics: Percent of Pool Items Classified at each Exposure Rate

Grade	Calculator	Total Items	Exposure Rate (%)					
			Unused	0–20	21–40	41–60	61–80	81–100
English								
3	N	1,196	0.5	99.0	0.4	0.1	0.0	0.0
4	N	1,306	1.2	98.7	0.1	0.0	0.0	0.0
5	N	1,267	2.1	97.5	0.4	0.0	0.0	0.0
6	All	1,099	0.0	98.4	1.5	0.2	0.0	0.0
	Y	566	0.0	98.2	1.8	0.0	0.0	0.0
	N	533	0.0	98.5	1.1	0.4	0.0	0.0
7	All	975	1.6	94.9	3.3	0.2	0.0	0.0
	Y	681	2.2	94.3	3.5	0.0	0.0	0.0
	N	294	0.3	96.3	2.7	0.7	0.0	0.0
8	All	862	0.5	95.8	2.9	0.5	0.4	0.0
	Y	646	0.6	94.7	3.6	0.6	0.5	0.0
	N	216	0.0	99.1	0.9	0.0	0.0	0.0
11	All	2,635	0.8	98.9	0.2	0.0	0.0	0.0
	Y	1745	1.0	98.9	0.1	0.0	0.0	0.0
	N	890	0.5	99.1	0.5	0.0	0.0	0.0
Braille								
3	N	385	1.0	85.2	13.5	0.3	0.0	0.0
4	N	359	3.3	85.8	7.8	3.1	0.0	0.0
5	N	381	4.7	87.1	7.6	0.5	0.0	0.0
6	All	385	0.0	89.6	8.3	1.8	0.3	0.0
	Y	190	0.0	83.2	14.2	2.6	0.0	0.0
	N	195	0.0	95.9	2.6	1.0	0.5	0.0
7	All	361	2.2	82.6	10.8	2.8	1.4	0.3
	Y	255	3.1	82.4	10.6	2.8	0.8	0.4
	N	106	0.0	83.0	11.3	2.8	2.8	0.0
8	All	290	0.3	83.5	8.6	3.8	2.4	1.4
	Y	206	0.5	80.1	8.7	5.3	3.4	1.9
	N	84	0.0	91.7	8.3	0.0	0.0	0.0
11	All	524	1.2	92.8	5.3	0.6	0.0	0.2
	Y	351	0.6	92.9	6.3	0.3	0.0	0.0
	N	173	2.3	92.5	3.5	1.2	0.0	0.6
Spanish								
3	N	375	0.8	85.9	12.3	1.1	0.0	0.0
4	N	388	3.6	84.8	9.8	1.8	0.0	0.0
5	N	406	4.4	89.4	5.9	0.3	0.0	0.0
6	All	395	0.0	91.7	5.6	2.0	0.8	0.0
	Y	196	0.0	88.3	6.6	4.1	1.0	0.0
	N	199	0.0	95.0	4.5	0.0	0.5	0.0
7	All	344	2.0	83.1	10.5	2.6	1.7	0.0
	Y	244	2.9	82.8	10.3	2.1	2.1	0.0
	N	100	0.0	84.0	11.0	4.0	1.0	0.0
8	All	309	0.3	83.5	9.7	3.9	1.0	1.6
	Y	233	0.4	80.3	10.7	5.2	1.3	2.2
	N	76	0.0	93.4	6.6	0.0	0.0	0.0

11	All	690	1.7	93.8	4.4	0.0	0.0	0.1
	Y	456	1.8	93.2	5.0	0.0	0.0	0.0
	N	234	1.7	94.9	3.0	0.0	0.0	0.4

5.4 Off-Grade Item Selection

As described in section 2.3, the off-grade items were added to the on-grade item pool when a student reached two-thirds of the adaptive test length. Tables 32–33 provide the number of off-grade items that were administered, the number of students who responded to off-grade items, the number of proficient students who took above-grade items, and the number of not-proficient students who took below-grade items for ELA/L and mathematics, respectively. All administered off-grade items, except for one Braille test in grade 4, were administered as specified in the algorithm: below-grade items are administered to students who are not proficient on their overall test performance, and above-grade items are administered to students who are proficient on their overall test performance.

Table 32. ELA/L: Number of Off-Grade Items Administered and Number of Tests with Off-Grade Items Administered

Grade	Total Off-Grade Items	Number of Administered Off-Grade Items	Number of Students who Responded to Off-Grade Items	Number of Proficient Students with Above Grade Items	Number of *not-Proficient* Students with *Below* Grade Items
English					
3	11	5	163	163	n/a
4	38	25	311	153	158
5	58	32	252	44	208
6	57	34	199	38	161
7	64	30	515	166	349
8	57	33	343	46	297
11	28	20	75	n/a	75
Braille					
3	8	6	255	255	n/a
4	27	11	373	188	184
5	38	26	322	32	290
6	36	19	276	0	276
7	42	24	700	354	346
8	23	20	398	48	350
11	12	8	61	n/a	61

Table 33. Mathematics: Number of Off-Grade Items Administered and Number of Tests with Off-Grade Items Administered

Grade	Total Off-Grade Items	Number of Administered Off-Grade Items	Number of Students who Responded to Off-Grade Items	Number of Proficient Students with Above Grade Items	Number of *not-Proficient* Students with *Below* Grade Items
English					
3	4	1	10	10	n/a
4	27	12	87	n/a	87
5	57	30	166	n/a	166
6	31	29	213	n/a	213
7	32	13	468	144	324
8	19	13	391	0	391
11	13	7	266	n/a	266
Braille					
3	1	0	0	0	n/a
4	20	8	271	n/a	271
5	30	12	173	n/a	173
6	20	20	228	n/a	228
7	17	9	423	n/a	423
8	11	10	429	n/a	429
11	7	2	246	n/a	246
Spanish					
3	1	0	0	0	n/a
4	19	8	271	n/a	105
5	30	12	190	n/a	190
6	20	20	228	n/a	228
7	14	7	400	n/a	400
8	8	7	331	n/a	331
11	5	1	147	n/a	147

6 FIELD-TEST ITEMS

In the 2016–17 adaptive pool in English, Smarter Balanced included 271 field-test items in ELA/L and 239 field-test items in mathematics.

In each simulated test, field-test items are embedded with the following rules:

- Field-test items appear at any position between after/at the fifth item on the test and before/at the fifth-from-last item on the test.
- Within the allowable field-test positions, each item will be administered in randomly selected positions.
- The number of field-test items administered to individual students will never exceed the intended maximum or fall short of the intended minimum.

The number of field-test items administered per student is same as the previous years to keep the same test length. Each student will be administered exactly two field-test items in mathematics and 3-6 field-test items in ELA/L, embedded in the allowable field-test positions. Table 34 provides the number of field-test items in the pool and the average number of field-test items administered in each test.

Table 34. Embedded Field-Test Items

Grade	ELA/L		Mathematics	
	Average Number of FT Items Administered per Student	Total Field-Test Items	Average Number of FT Items Administered per Student	Total Field-Test Items
3	3.4	38	2	29
4	3.9	36	2	26
5	3.7	41	2	22
6	4.2	38	2	38
7	3.8	52	2	45
8	3.6	66	2	49
11	0	0	2	30

7 SUMMARY

The Smarter Balanced adaptive test delivery system administers assessments with items representing the breadth and depth identified in the test specifications and content standards. The overall blueprint match results demonstrate that all test forms conform to the same content coverage, thus providing evidence of content comparability. In other words, while each form is unique with respect to its items, all forms align with the same curricular expectations set forth in the test blueprints.

The summary statistics of the estimated abilities show that for all examinees in all grades, the item selection algorithm is choosing items that are optimized, conditional on each examinee's ability. Essentially, this shows that the examinee-ability estimates generated on the basis of the items chosen are optimal in the sense that the final score for each examinee recovers the true score across ability ranges - an indication that the algorithm is working exactly as expected for a computer-adaptive test.

Overall, the diagnostics on the item-selection algorithm provide evidence to support the following: scores are comparable with respect to the targeted content; scores are measured with good precision across the range of proficiency, given the item contents and the item difficulty distributions in the pool; item exposure is minimized given the parameter chosen; and off-grade items are administered according to the criteria. Moreover, the field-test items are distributed equally as intended.

REFERENCES

Cohen, J., & Albright, L. (2014a). *Smarter Balanced adaptive item selection algorithm design report*, Washington, D.C, http://www.smarterapp.org/documents/AdaptiveAlgorithm-Preview-v3.pdf.

Cohen, J., & Albright, L. (2014b). *Talking points for out of grade level testing*, Washington, D.C.

Smarter Balanced Assessment Consortium. (2015). *ELA/Literacy Smarter Balanced Summative Assessment Blueprint*, http://www.smarterbalanced.org/wordpress/wp-content/uploads/2015/02/ELA_Blueprint.pdf.

Smarter Balanced Assessment Consortium. (2015). *Mathematics Smarter Balanced Summative Assessment Blueprint*, http://www.smarterbalanced.org/wordpress/wp-content/uploads/2015/02/Mathematics_Blueprint.pdf.

Appendix A

Distribution of Item Difficulties in the 2015–16 and 2016–17 Item Pools

Figure A1. ELA/L Grades 3–6: Item Difficulty Distributions

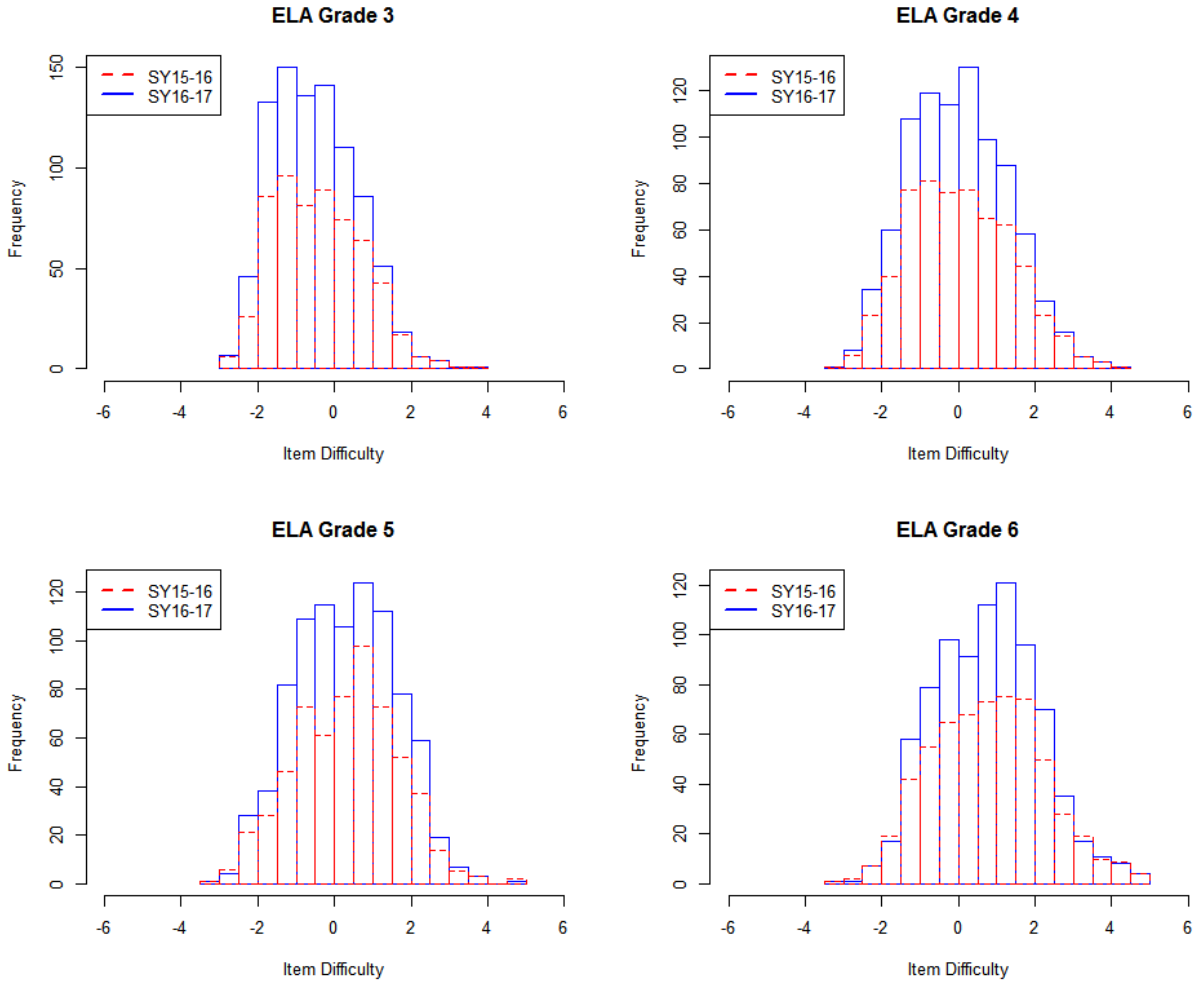


Figure A2. ELA/L Grades 7–8, 11: Item Difficulty Distributions

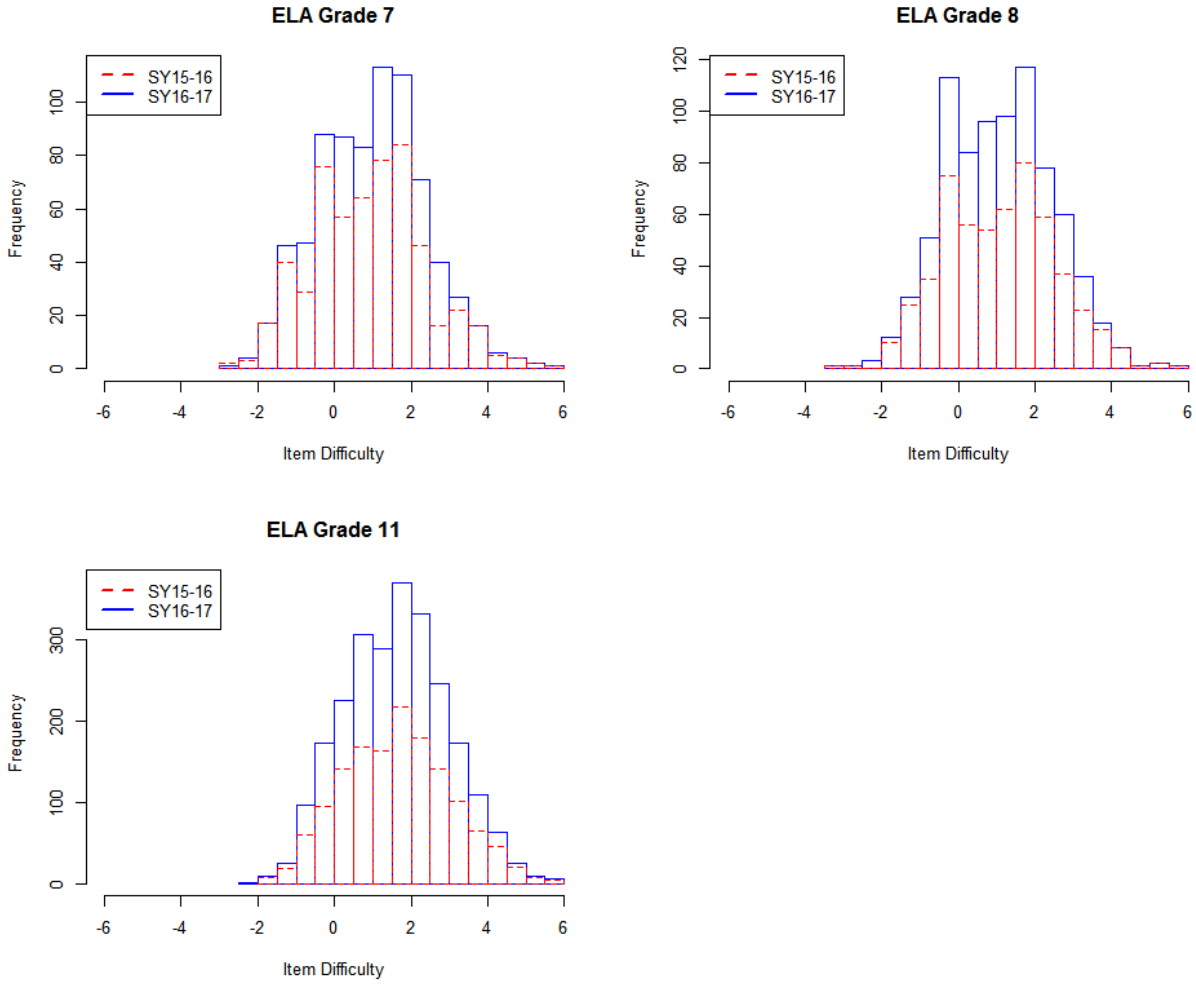


Figure A3. Math Grades 3–6: Item Difficulty Distributions

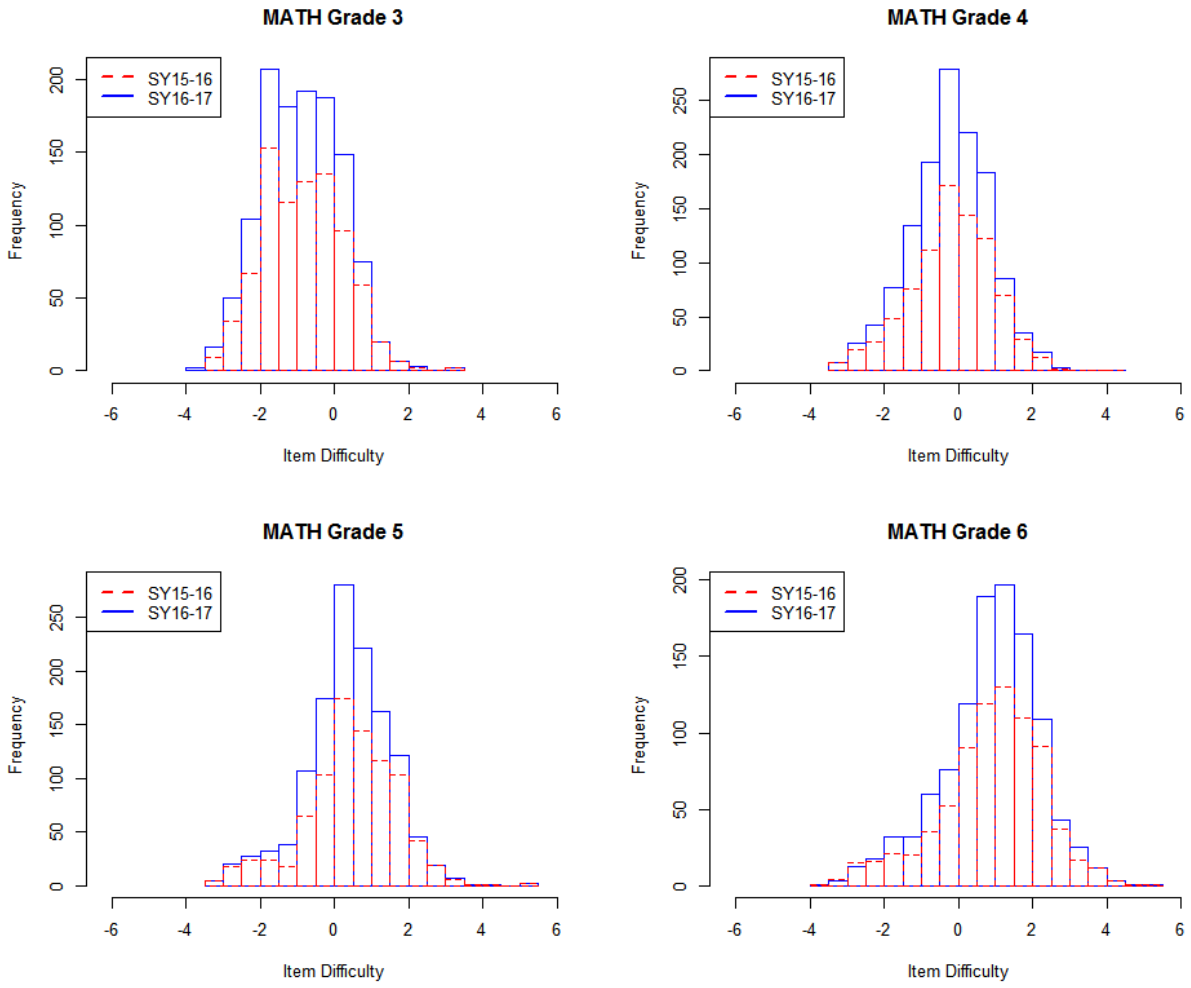
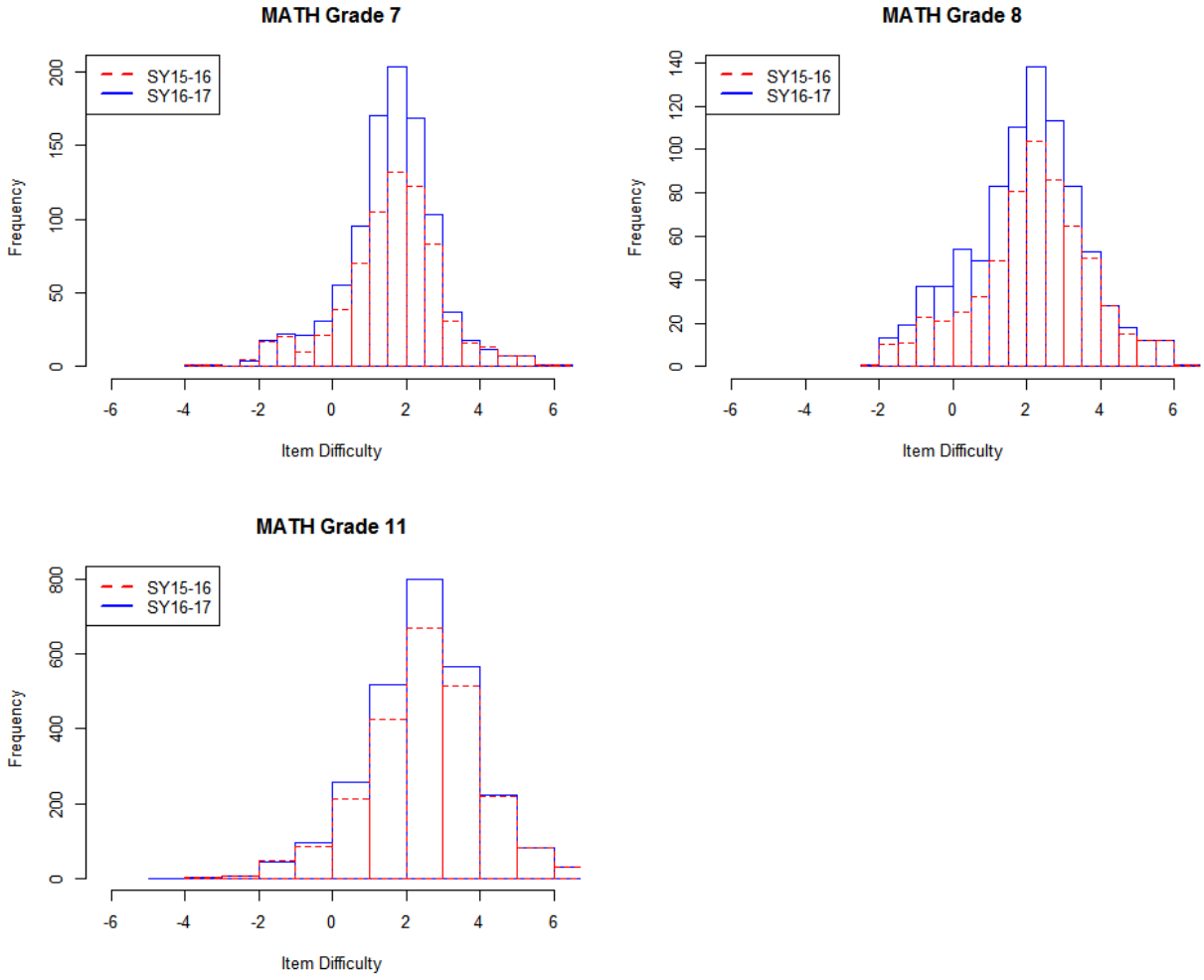


Figure A4. Math Grades 7, 8, 11: Item Difficulty Distributions



Appendix B

Average Difficulty for the On-Grade and Off-Grade Items

Table B1. ELA/L: Average Difficulty for the On-Grade and Off-Grade Item Pools

Grade	Above/Below Grade	Item Difficulty				
		Number of Items	Min.	Max.	Average	SD
English						
3	Above	11	-1.42	1.38	-0.14	1.05
	On	879	-2.92	3.82	-0.52	1.09
4	Above	18	-1.53	1.71	-0.01	1.04
	On	835	-3.25	4.25	0.05	1.25
	Below	20	-2.90	-1.62	-2.23	0.37
5	Above	15	-1.62	2.75	0.40	1.23
	On	828	-2.40	4.81	0.41	1.21
	Below	43	-3.25	-1.04	-1.98	0.52
6	Above	16	-1.44	2.60	0.49	1.10
	On	769	-2.72	4.92	0.88	1.26
	Below	41	-3.25	-0.98	-1.60	0.46
7	Above	19	-0.67	3.58	0.80	1.30
	On	699	-2.02	5.52	1.11	1.30
	Below	45	-2.72	-0.28	-1.26	0.53
8	Above	15	-0.89	2.48	0.77	0.92
	On	751	-3.01	5.57	1.20	1.31
	Below	42	-2.72	-0.28	-1.07	0.55
11	On	2435	-2.09	5.93	1.66	1.32
	Below	28	-1.98	-0.50	-1.09	0.38
Braille						
3	Above	8	-1.42	1.38	-0.15	1.05
	On	292	-2.90	3.82	-0.49	1.23
4	Above	12	-1.53	1.71	0.02	1.20
	On	279	-3.25	4.25	0.12	1.40
	Below	15	-2.90	-1.62	-2.28	0.39
5	Above	6	-1.28	2.75	0.25	1.43
	On	299	-2.40	4.81	0.38	1.26
	Below	32	-3.25	-1.04	-2.05	0.54
6	Above	5	-0.88	1.8	0.26	1.07
	On	270	-1.77	4.92	0.94	1.33
	Below	31	-3.25	-0.98	-1.62	0.49
7	Above	11	-0.67	3.01	0.57	1.25
	On	245	-1.88	5.52	1.09	1.40
	Below	31	-2.18	-0.30	-1.30	0.46
8	Above	3	-0.23	1.11	0.50	0.68
	On	282	-3.01	5.57	1.22	1.42
	Below	20	-1.77	-0.30	-1.03	0.45
11	On	528	-1.61	5.93	1.73	1.43
	Below	12	-1.90	-0.72	-1.16	0.37

Table B2. Mathematics: Average Difficulty for the On-Grade and Off-Grade Item Pools (English)

Grade	Calculator (Y/N)	Above/Below Grade	Number of Items	Item Difficulty			
				Min	Max	Average	SD
3	N	Above	4	-2.00	-1.88	-1.93	0.05
	N	On	1192	-3.57	3.46	-0.88	1.04
4	N	On	1279	-3.26	4.11	-0.16	1.01
	N	Below	27	-3.15	-2.15	-2.72	0.26
5	N	On	1210	-2.79	5.28	0.51	1.00
	N	Below	57	-3.26	-1.59	-2.37	0.45
6	Y	On	566	-3.93	5.10	1.07	1.32
	N	On	502	-2.61	4.32	0.86	1.06
	N	Below	31	-3.14	-1.21	-2.36	0.45
7	Y	Above	1	-2.19	-2.19	-2.19	–
	Y	On	666	-1.79	6.17	1.67	1.17
	Y	Below	14	-3.93	-0.90	-1.66	0.73
	N	Above	4	-1.55	-1.14	-1.33	0.18
	N	On	277	-1.28	5.64	1.73	1.06
	N	Below	13	-3.14	-0.93	-1.71	0.59
8	Y	Above	3	-2.19	-1.60	-1.83	0.32
	Y	On	638	-1.87	6.70	2.05	1.52
	Y	Below	5	-1.79	-1.09	-1.44	0.30
	N	On	205	-1.30	6.32	2.02	1.42
	N	Below	11	-1.70	-0.93	-1.33	0.30
11	Y	On	1736	-3.36	7.30	2.66	1.53
	Y	Below	9	-2.98	-0.85	-1.35	0.65
	N	On	886	-4.43	6.55	2.09	1.34
	N	Below	4	-1.70	-0.93	-1.40	0.33

Table B3. Mathematics: Average Difficulty for the On-Grade and Off-Grade Item Pools (Braille)

Grade	Calculator (Y/N)	Above/Below Grade	Number of Items	Item Difficulty			
				Min	Max	Average	SD
3	N	Above	1	-2.00	-2.00	-2.00	–
	N	On	384	-3.15	3.08	-0.96	1.08
4	N	On	339	-3.26	2.57	-0.22	1.02
	N	Below	20	-3.15	-2.15	-2.71	0.28
5	N	On	351	-2.53	5.04	0.54	1.07
	N	Below	30	-3.26	-1.59	-2.34	0.45
6	Y	On	190	-3.93	5.10	1.17	1.41
	N	On	175	-1.70	4.32	0.73	1.14
	N	Below	20	-3.09	-1.72	-2.37	0.36
7	Y	On	247	-1.79	6.17	1.66	1.30
	Y	Below	8	-3.93	-0.90	-1.70	0.94
	N	On	97	-1.28	5.64	1.84	1.39
	N	Below	9	-2.35	-0.93	-1.53	0.41
8	Y	On	204	-1.54	5.93	2.14	1.55
	Y	Below	2	-1.79	-1.65	-1.72	0.10
	N	On	75	-0.77	5.75	2.17	1.49
	N	Below	9	-1.70	-0.93	-1.38	0.31
11	Y	On	348	-3.01	7.21	2.82	1.76
	Y	Below	3	-1.39	-0.93	-1.23	0.26
	N	On	169	-0.94	5.43	2.36	1.36
	N	Below	4	-1.70	-0.93	-1.40	0.33

Table B4. Mathematics: Average Difficulty for the On-Grade and Off-Grade Item Pools (Spanish)

Grade	Calculator (Y/N)	Above/Below Grade	Number of Items	Item Difficulty			
				Min	Max	Average	SD
3	N	Above	1	-2.00	-2.00	-2.00	–
	N	On	374	-3.15	3.46	-0.82	1.10
4	N	On	369	-3.26	4.11	0.04	1.03
	N	Below	19	-3.15	-2.15	-2.76	0.26
5	N	On	376	-2.53	5.04	0.67	1.07
	N	Below	30	-3.26	-1.59	-2.38	0.44
6	Y	On	196	-3.93	5.10	1.35	1.39
	N	On	179	-1.70	4.32	0.86	1.14
	N	Below	20	-3.00	-1.72	-2.38	0.37
7	Y	On	237	-1.79	6.17	1.81	1.29
	Y	Below	7	-3.93	-0.90	-1.70	1.02
	N	On	93	-0.94	5.64	2.04	1.24
	N	Below	7	-2.35	-0.93	-1.55	0.45
8	Y	On	231	-1.54	5.93	2.49	1.42
	Y	Below	2	-1.79	-1.65	-1.72	0.10
	N	On	70	-0.77	4.75	2.18	1.35
	N	Below	6	-1.70	-0.93	-1.37	0.35
11	Y	On	455	-3.01	7.30	2.99	1.54
	Y	Below	1	-1.12	-1.12	-1.12	–
	N	On	230	-1.28	5.43	2.30	1.35
	N	Below	4	-1.70	-0.93	-1.40	0.33

Appendix C

Number of Unique Items Administered by Item Position

Figure C1. ELA/L Number of Unique Items Administered by Item Position (English)

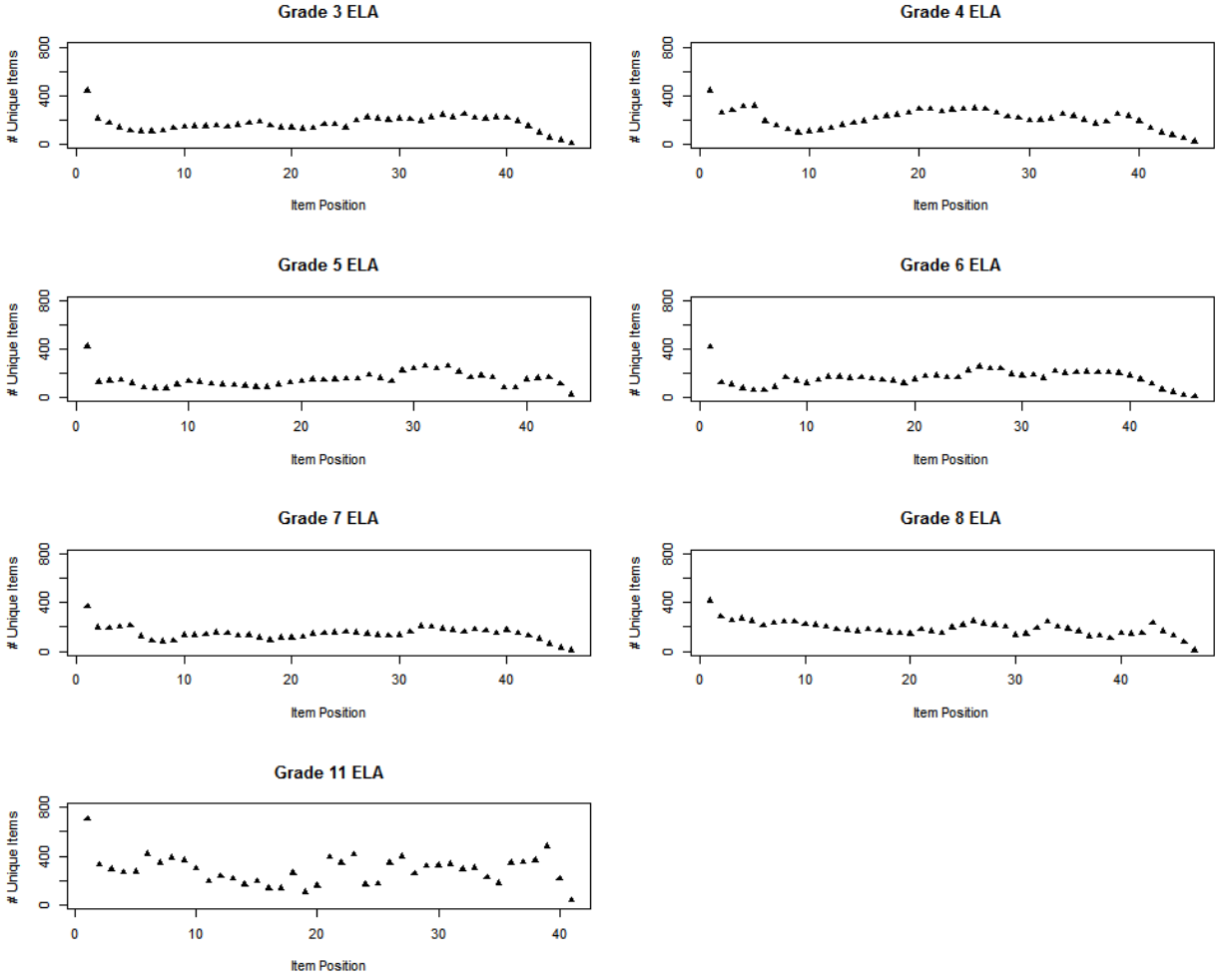


Figure C2. ELA/L Number of Unique Items Administered by Item Position (Braille)

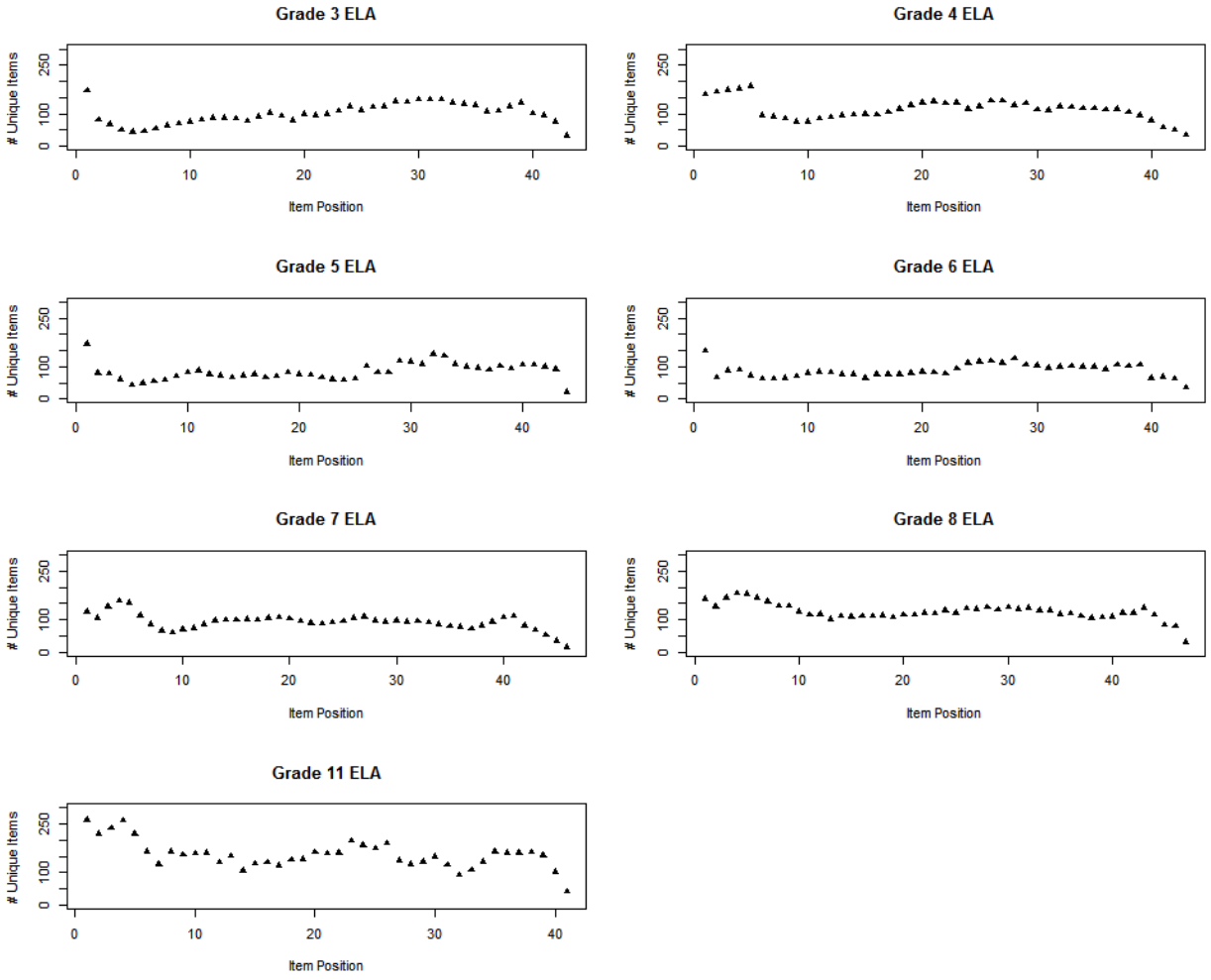


Figure C3. Mathematics Number of Unique Items Administered by Item Position (English)

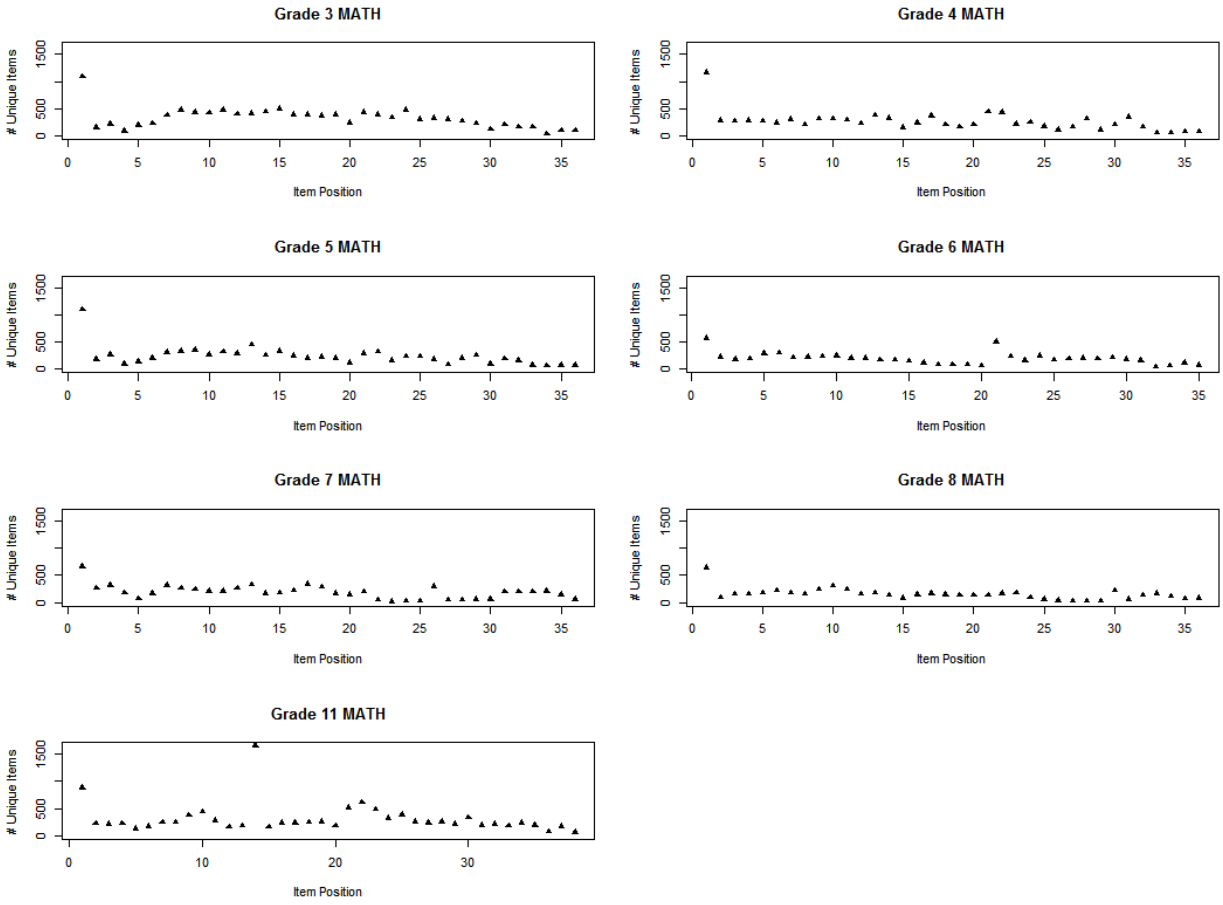


Figure C4. Mathematics Number of Unique Items Administered by Item Position (Braille)

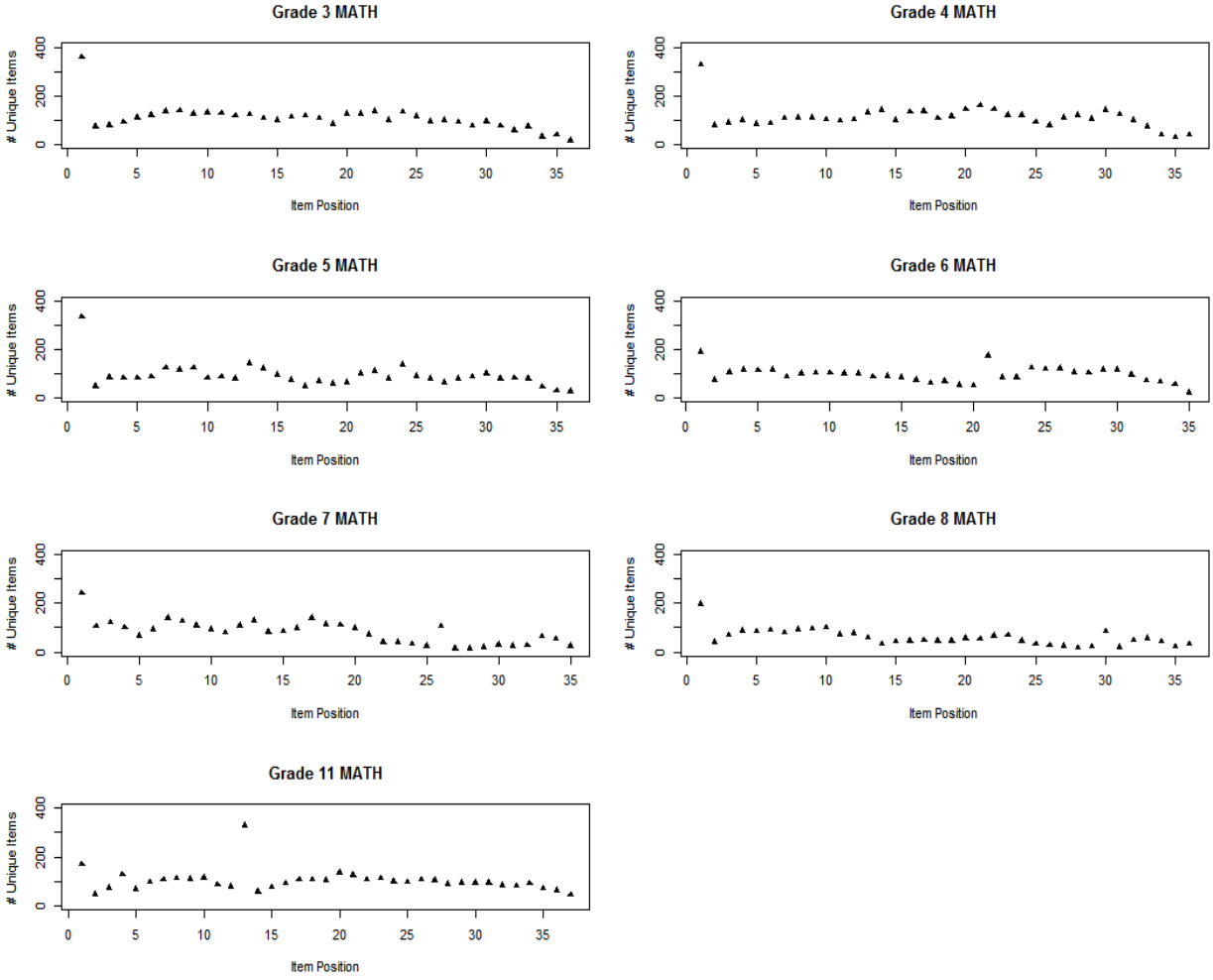


Figure C5. Mathematics Number of Unique Items Administered by Item Position (Spanish)