PARCC

Spring 2015 Digital Devices Comparability Research Study

Jeffrey Steedle, Malena McBride, Marc Johnson, and Leslie Keng

Pearson

Contents

Executive Summary

The first operational administration of the Partnership for Assessment of Readiness for College and Careers (PARCC) took place during the 2014–2015 school year. In addition to the traditional paper-and-pencil format, the assessments were available for administration on a variety of electronic devices, including desktop computers, laptop computers, and tablets. This report describes a study of the comparability of PARCC assessments administered on tablets and non-tablet devices, such as notebook and desktop computers. The goal of this study was to investigate whether assessment tasks were of similar difficulty on tablets and non-tablet devices, whether the psychometric properties of test scores were similar when comparing testing on tablets and non-tablet devices, and whether overall test performance was similar for students testing on tablets and non-tablet devices.

This study examined performance on eight PARCC assessments: grade 5 mathematics, grade 7 mathematics, Algebra 1, Geometry, Algebra 2, grade 3 English language arts/literacy (ELA/L), grade 7 ELA/L, and grade 9 ELA/L. Any student who took one of the study forms on a tablet or non-tablet device in spring 2015 was eligible for inclusion in this study. Students were matched on demographic variables to create samples that would be comparable, thereby providing an unbiased comparison of performance on tablet and non-tablet devices.

Overall, this study revealed consistent evidence of comparability between testing on tablets and non-tablet devices. Specifically, the task means and IRT difficulty estimates suggested that tasks were similarly difficult on tablets and non-tablet devices. A small number of tasks were flagged for device effects, and nearly all of them were part of high school mathematics assessments, particularly Geometry. However, no specific task type was found to consistently exhibit device effects. Consistent with findings about task difficulty, an examination of raw score and scale score distributions indicated similar overall performance on the performance-based assessment (PBA) and end-of-year (EOY) components of the 2015 PARCC assessments. Moreover, item response theory (IRT) true-score equating indicated that students who tested on non-tablet devices would be expected to score similarly had they taken the PARCC assessments on tablets.

Additional analyses compared the psychometric properties of scores from students testing on tablets and non-tablet devices. Correlations between the EOY and PBA components were similar in the tablet and non-tablet conditions for all eight assessments analyzed in this study, and scores from assessments administered on tablets and non-tablet devices were similarly reliable. Validity coefficients (i.e., correlations with a prior test score in the same content area) were also similar for tablet and non-tablet testers. An analysis of person-fit statistics did not reveal notable differences in the distributions of person-fit statistics for students testing on tablets and non-tablet devices.

This study generated robust evidence of comparability on the analyzed assessments. When there was statistical evidence of device effects, the magnitudes of those effects were small. Given the potential presence of device effects, task development and interface design need to take into consideration the familiarity that students have with nontraditional task types

administered online. The generalizability of findings from this study may be limited as a result of the select sample of PARCC assessments analyzed, but large sample sizes and apparently high-quality matching support the internal validity of this study's findings.

Section 1: Background

1.1 Introduction

The Partnership for Assessment of Readiness for College and Careers (PARCC) assessments were implemented operationally for the first time during the 2014–2015 school year. In addition to being available through traditional paper-based testing (PBT), assessments for all grades and courses were also available for administration via computer-based testing (CBT). Students were able to take the PARCC assessment on desktop computers, laptops, tablets, and other digital devices.

Professional testing standards (AERA, APA, & NCME, 1999, 2014) stipulate that whenever a test is administered in multiple modes, such as PBT and CBT, score comparability must be examined to support the validity of test score interpretations. Because of the variety of digital devices that may be used to take the PARCC assessments, it is also important to evaluate score comparability across CBT devices, particularly between traditional computing devices, such as desktop and laptop computers, and newer computing devices, such as tablets.

Although PARCC adopted technology requirements that mandate the use of external keyboards with tablets, this requirement is not sufficient to ensure comparability with tests delivered via other CBT devices. On tablets, students use the touch screen interface for cursor placement and text selection for all tasks involving text entry, and they engage with the touch screen to manipulate objects and input their answers for other tasks types. Additionally, the screen size of allowable tablets (9.5 inch or greater) is smaller than most desktop and laptop monitors and therefore creates possible challenges with regard to screen real estate. Students using tablets may therefore use more scrolling to access the same task content as their peers testing on desktops or laptops. Lastly, tablet-specific features (e.g., screen rotation, pinch and zoom, etc.) may create additional differences between student experiences on traditional computers and tablets, such that test score interchangeability cannot be assumed for examinees testing on different device types.

PARCC planned a set of research studies to evaluate the comparability of different assessment delivery methods. The overall purpose of these studies was to ensure that the PARCC assessments measured the Common Core State Standards with fidelity, that the methods of delivery and analysis were innovative and of the highest quality, and that PARCC would be poised to support inferences about student achievement with compelling and comprehensive evidence in 2014–2015 and beyond. PARCC's comparability research agenda includes two studies: a study to evaluate the score comparability between the PBT and CBT modes of administration and a study to evaluate score comparability between tests administered on different types of digital devices. It should be noted that there is a difference in the degree of comparability expected in these two studies. As stated in PARCC's request for proposal, "Note that strict comparability (i.e., score interchangeability) across CBT and PBT **is not** a goal for PARCC; however, score interchangeability across CBT devices and input types **is** a goal" (State of Indiana, 2012).

1.2 Spring 2014 Study Summary

A first round of comparability studies was conducted using data from the spring 2014 field test administration of the PARCC assessments (Keng, Davis, McBride, Glaze, & Steedle, 2015). The goal of the field-test device comparability study was to investigate whether assessment tasks were similarly difficult on tablets and non-tablet devices (i.e., notebook and desktop computers), whether the psychometric properties of test scores were similar when comparing testing on tablets and non-tablet devices, and whether overall test performance was similar for students testing on tablets and non-tablet devices. This study examined performance on six PARCC assessments: grade 4 English language arts/literacy (ELA/L), grade 4 mathematics, grade 8 ELA/L, grade 8 mathematics, grade 10 ELA/L, and Geometry.

The data analyzed for this study came primarily from Burlington Public Schools, a school district in Massachusetts that has invested substantial resources and effort toward integrating electronic devices into regular classroom practice. Fourth-graders from Burlington who took the PARCC assessment on tablets were matched with other students in Massachusetts who took the assessment on non-tablet devices. Burlington students in grades 8 and 10 were randomly assigned by classroom to take the PARCC assessment on tablets or non-tablet devices. Several analyses were conducted to assess the comparability of testing on tablets and non-tablet devices. The task-level analysis revealed that task difficulty was generally similar on tablets and non-tablet devices, with the exception of the grade 4 mathematics assessment. Correlations between the performance-based assessment (PBA) and end-of-year (EOY) components were comparable on tablets and non-tablet devices, with the exception of the grade 8 mathematics assessment. An analysis of reliability coefficients indicated that test scores from tablet and non-tablet administrations were similarly reliable, although small significant differences in reliability were detected for the grade 8 mathematics and grade 10 ELA/L assessments. The validity of scores from tablet and non-tablet administrations was examined by correlating the PARCC assessments with other measures of the same construct. The analysis indicated that validity coefficients for tablet and non-tablet testing were comparable, with the exception of the grade 4 ELA/L assessment. Finally, test-level score interpretation analysis revealed that raw scores from tests administered on tablets and non-tablet devices were comparable, with the exception of the grade 4 ELA/L assessment.

The 2014 device comparability study generated evidence consistent with comparability on most of the analyzed assessments. Based on results from the 2014 study, PARCC planned to treat scores for students testing on tablets and computers as comparable during the first operational PARCC assessment administration in 2014–2015. However, the generalizability of findings from the 2014 study was limited as a result of the select sample of PARCC assessments analyzed and the small, non-representative sample of students who participated. Given these limitations, PARCC opted to conduct a follow-up study using operational data from the spring 2015 administration.

1.3 Spring 2015 Study

This report summarizes the findings of the digital devices comparability study conducted using data from the spring 2015 operational administration of the PARCC assessments. The specific research questions investigated included:

1. Is student performance on individual tasks similar when comparing tasks administered on tablets and non-tablet devices?
2. Are the psychometric properties of test scores similar when comparing tests administered on tablets and non-tablet devices?
3. Are students' levels of overall test performance similar when comparing tests administered on tablets and non-tablet devices?

Section 2: Method

2.1 PARCC Assessments

Students in grade 3 through high school take PARCC assessments in mathematics and English language arts/literacy (ELA/L). In 2014–2015, the PARCC assessment at each grade level and content area included two components: performance based assessment (PBA) and end-of-year (EOY). In spring 2015, PBA and EOY were administered during two separate testing windows between March and May. PBA was to be administered after approximately 75% of the academic year, and EOY was to be administered after approximately 90% of the academic year.

A subset of the PARCC assessments was selected for the comparability research study concerning digital devices. The assessments included in the study were:

- Grades 5 and 7 Mathematics
- Algebra 1, Geometry, and Algebra 2
- Grades 3, 7, and 9 ELA/L

These assessments were selected so that the study would include assessments representative of each content area at each level of schooling: elementary school, middle school, and high school. For each grade and content area, one PBA form and one EOY form were chosen as the study forms for this research. In selecting the study forms, evaluations of the task types were conducted so that, to the extent practicable, a variety of innovative and technology-enhanced items (TEIs) were included in the research. Tasks omitted from the spring 2015 analyses and scoring were also omitted from this study.

2.2 Sampling and Recruitment

Unlike the 2014 study, there was no special sampling or recruitment for the 2015 study. Any student who took one of the study forms during the spring 2015 PARCC assessment administration was eligible for inclusion in the study. Data from the following states were considered for inclusion in this study: Arkansas, Colorado, District of Columbia, Illinois, Massachusetts, Maryland, Mississippi, New Jersey, New Mexico, Ohio, and Rhode Island.

However, some states were occasionally excluded from analyses because they had zero students testing on tablets for certain grades and subjects.

One of the aforementioned states (no longer a PARCC member) was excluded from all analyses because of highly atypical differences between the performance of students who tested on tablets and non-tablet devices. When analyses included data from this state, extensive evidence of device effects was observed on nearly every assessment. Consider, for example, that after creating groups matched on demographics, students in that state who tested on tablets scored an average of 10 raw score points lower on Algebra 1 PBA and 14 raw score points lower on Algebra 1 EOY. With these data included, the overall differences in performance between the matched tablet and non-tablet groups on the Algebra 1 PBA and EOY tests were 0.40 and 0.42 standard deviations, respectively (with tablet testers performing worse). Without that state, those differences were only 0.03 and –0.08 standard deviations. With the available data, it is not possible to discern exactly what caused this state's tablet testing population to be aberrant. Perhaps there was large district with widespread use of tablets whose students were unusually low-performing or unfamiliar with testing on tablets.

2.3 Matching

Students were not randomly assigned to take the PARCC assessment on tablets or non-tablet devices. Thus, if differences in performance were detected between students who took the PARCC assessment under those two conditions, those differences could not be attributed only to a device effect. Perhaps, for example, schools enrolling students with high prior achievement were more likely to administer via tablets. In that case, the assessment would appear to be easier for students testing on tablets versus non-tablet devices, but that apparent easiness could not be attributed to a device effect. To address this difficulty, matched samples of students taking the study forms on tablets and non-tablet devices were created to support better inferences about student performance in the tablet and non-tablet conditions.

In general, the number of students testing on tablets was smaller than the number of students testing on non-tablet devices. Consequently, the tablet group was treated as the "base" sample, and the non-tablet group was treated as the "matching" sample. That is, the matching procedure attempted to identify matches for the tablet testers in the larger group of non-tablet testers.

Within each state, students in the tablet and non-tablet groups were matched on the following variables (when available):

- Sex
- Ethnicity
- Economic Disadvantage Status (Yes or No)
- English Learner Status (Yes or No)
- Student with Disabilities Status (Yes or No)
- Testing Language (English or Spanish)

- Accommodation (None or Text-to-Speech)
- Grade Level When Assessed

Not every state provided data for every variable, so variables were occasionally excluded from matching analyses. Moreover, the data were sparse for some variables in certain states (e.g., less than 50% of students with non-blank records). In such cases, the variable was excluded from the matching analysis in order to preserve sample size. Otherwise, a large number of students with missing data would have been excluded.

Matching was conducted separately for each state, and the matched samples were later combined into a single matched data set for each assessment. The reason for matching by state was to control for differences between states in their progress toward full implementation of instructional and assessment practices aligned to the Common Core State Standards. Although students will take the PARCC assessments in a single administration in 2016 (not as separate PBA and EOY components), many of the analyses in the study were conducted for PBA and EOY separately. Preservation of sample size (and the resulting statistical precision of results) was the reason for this decision. Each PBA and EOY study form was administered to tens of thousands of students, but the number of students who took the exact combination of PBA and EOY study forms was much smaller, and the subset of students who took that exact combination on a tablet was even smaller. When possible, results are reported for this small subset of students (labeled "PBA+EOY").

2.3.1 Propensity Score Matching

The original plan for this study called for the use of propensity score matching (PSM). In PSM, students are matched based on their "propensity" to be in the treatment condition. In this case, the "propensity score" is a student's estimated probability of testing on a tablet, based on a logistic regression model that includes the matching variables as predictors. Each student in the tablet condition was matched to a student in the non-tablet condition with a similar propensity to test on a tablet. In this study, optimal PSM was conducted using the *MatchIt* package for R (Ho, Imai, King, & Stuart, 2011).

"Balance" between the tablet condition (TC) and non-tablet condition (NTC) was examined to evaluate the quality of matching results. That is, the distributions of the demographic variables were compared between the TC and NTC testers before and after matching. If the demographic distributions were more similar after matching, this suggested that the matching procedure was successful. As an example, the first columns of Table 1 ("Unmatched") show the percentages of NTC and TC students in different demographic categories for Algebra 1 PBA. Notice that some differences are relatively large (e.g., the percentage of Black/African American students). After matching, the differences should be reduced nearly to zero. In the case of Algebra 1 PBA, none of the percentage differences exceeded 1%. A similar pattern of results is apparent in Table 2, which shows balance for Grade 7 ELA/L EOY.

With regard to missing variables, PSM depends on the strong ignorability assumption (Rosenbaum & Rubin, 1983), which is that all covariates related with group selection (TC or

NTC) and potential outcomes (e.g., PARCC assessment performance) are included in the propensity score model. One additional variable—prior achievement on a state assessment in the same content area—was made available by three of the analyzed states, although one state's data were not useful for matching because there were zero tablet testers with prior achievement. Prior achievement is known to be a strong predictor of current achievement, so its omission from the propensity score model could have introduced unknown biases in results. However, results in the "Prior Achievement" rows of Tables 1 and 2 suggest that the NTC and TC groups matched only on demographic variables were reasonably well matched in terms of student proficiency (differences of –0.11 and 0.15 standard deviations based only on matched students who both had prior achievement data). Matching with prior achievement is described in Section 2.3.3.

Table 1

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Algebra 1 PBA

	Unmatched			PSM			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	62039	1489		1489	1489		1445	1445		166	166	
American Indian/Alaska Native	1.0	0.7	-0.3	0.7	0.7	0.1	0.5	0.5	0.0	0.0	0.0	0.0
Asian	5.2	7.5	2.3	8.1	7.5	-0.6	7.0	7.0	0.0	2.4	2.4	0.0
Black/African American	24.1	14.4	-9.7	13.6	14.4	0.7	14.5	14.5	0.0	19.9	19.9	0.0
Hispanic/Latino Ethnicity	28.0	30.6	2.6	31.5	30.6	-0.9	31.0	31.0	0.0	31.9	31.9	0.0
Hawaiian/Pacific Islander	0.3	0.0	-0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
White	39.4	44.9	5.5	44.1	44.9	0.7	45.4	45.4	0.0	44.0	44.0	0.0
Two or More Races	1.4	1.1	-0.2	1.1	1.1	0.1	1.1	1.1	0.0	1.8	1.8	0.0
Ethnicity Not Provided	0.6	0.7	0.2	0.9	0.7	-0.1	0.6	0.6	0.0	0.0	0.0	0.0
Female	47.6	46.8	-0.8	46.7	46.8	0.1	47.1	47.1	0.0	50.0	50.0	0.0
Economic Disadvantage	44.3	43.7	-0.6	44.6	43.7	-0.9	44.2	44.2	0.0	42.8	42.8	0.0
English Learner	9.8	10.1	0.4	9.9	10.1	0.3	9.0	9.0	0.0	8.4	8.4	0.0
Student with Disability	10.1	10.5	0.4	10.1	10.5	0.4	10.4	10.4	0.0	4.2	4.2	0.0
Prior Achievement*	0.05	0.05	0	0.26	0.15	-0.11	0.17	0.12	-0.04	0.15	0.11	-0.04
Grade 6	0.1	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Grade 7	3.8	0.6	-3.2	0.7	0.6	-0.1	0.6	0.6	0.0	0.0	0.0	0.0
Grade 8	18.4	22.4	4.0	23.0	22.4	-0.5	22.2	22.2	0.0	49.4	49.4	0.0
Grade 9	68.2	67.5	-0.8	68.2	67.5	-0.7	68.8	68.8	0.0	48.8	48.8	0.0
Grade 10	8.1	8.1	-0.1	7.1	8.1	1.0	7.6	7.6	0.0	1.8	1.8	0.0
Grade 11	1.2	1.3	0.0	1.0	1.3	0.3	0.8	0.8	0.0			
Grade 12	0.2	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table 2

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 7 ELA/L EOY

	Unmatched			PSM			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	69840	3039		3039	3039		3031	3031		635	635	
American Indian/Alaska Native	1.0	0.5	-0.5	0.5	0.5	0.0	0.4	0.4	0.0	0.0	0.0	0.0
Asian	5.2	7.3	2.1	7.4	7.3	0.0	7.4	7.4	0.0	7.1	7.1	0.0
Black/African American	17.5	10.2	-7.3	9.9	10.2	0.3	10.2	10.2	0.0	12.9	12.9	0.0
Hispanic/Latino Ethnicity	25.5	21.4	-4.2	22.1	21.4	-0.8	21.4	21.4	0.0	13.2	13.2	0.0
Hawaiian/Pacific Islander	0.3	0.1	-0.2	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0
White	48.0	57.8	9.8	57.2	57.8	0.6	57.8	57.8	0.0	63.8	63.8	0.0
Two or More Races	1.4	0.8	-0.6	0.8	0.8	0.0	0.8	0.8	0.0	1.4	1.4	0.0
Ethnicity Not Provided	1.0	1.9	0.9	2.1	1.9	-0.2	1.9	1.9	0.0	1.6	1.6	0.0
Female	48.5	47.5	-1.0	48.6	47.5	-1.1	47.5	47.5	0.0	48.0	48.0	0.0
Economic Disadvantage	47.9	37.8	-10.1	38.3	37.8	-0.5	37.7	37.7	0.0	29.0	29.0	0.0
English Learner	5.9	5.0	-0.9	4.7	5.0	0.4	4.9	4.9	0.0	4.4	4.4	0.0
Student with Disability	10.9	9.5	-1.4	9.5	9.5	0.1	9.3	9.3	0.0	14.7	14.7	0.0
Prior Achievement*	-0.07	0.11	0.18	-0.04	0.11	0.15	0.01	0.14	0.14	0.11	0.12	0.01
Grade 7	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

2.3.2 Coarsened Exact Matching

Coarsened exact matching (CEM; Iacus, King, & Porro, 2012) was applied to the data in an effort to improve upon the PSM results. PSM depends on associations between the matching variables and group membership. When those associations are weak, PSM results may not represent an improvement over the unmatched data. Although the PSM results across all states appear to be reasonable (Tables 1 and 2), within certain states the demographic variables were not useful predictors of group membership. Predictive utility was evaluated using McFadden's pseudo-R^2 for logistic regression models (McFadden, 1974). As examples, pseudo-R^2 ranged from .02 to .28 (median .13) for Algebra 1 PBA and from .01 to .65 (median .03) for grade 7 ELA/L. (Note: the state with pseudo-R^2 of .65 had only 3 tablet testers.)

CEM does not depend on associations between the matching variables and group membership. CEM simply identifies "strata" of students with the same demographic characteristics. For a given strata, some students will be tablet testers and some will be non-tablet testers, and students within a strata are treated as matched. The "CEM Matched" columns in Table 1 and Table 2 show example results from CEM. Notice that the CEM samples sizes are slightly smaller than the PSM sample sizes because some students cannot be matched. For Algebra 1 PBA (Table 1) and for grade 7 ELA/L EOY (Table 2), CEM reduced all of the demographic differences between the TC and NTC to 0%. Note also that the average difference in prior achievement was reduced by 0.07 standard deviations for Algebra 1 PBA (based on 165 matched pairs). Given the low pseudo-R^2 values from PSM and the slight improvements in balance offered by CEM, CEM was selected as the matching method for the device comparability study. In this study, CEM was conducted using the *cem* package for R (Iacus, King, & Porro, 2009).

Appendix A provides the complete set of CEM results for all assessments analyzed in this study. For all assessments, demographic differences were reduced to zero, but differences in prior achievement were not. Some differences appear quite large (e.g., 0.27 standard deviations for Geometry PBA and 0.32 for Algebra 2 PBA), but those differences are based on small samples (9 and 27 pairs of students, respectively), which provide less estimation precision.

2.3.2 Matching with Prior Achievement

As noted and Section 2.3.1, the omission of relevant variables from the matching procedure can potentially bias results. As shown in Tables 1 and 2 (and throughout Appendix A), average differences in prior achievement (expressed on a *z*-score metric) between the NTC and TC groups were generally reduced by matching on demographic variables only. When data were available, CEM using demographics and prior achievement was applied to the data. Sample results are provided in the "CEM with Prior Ach." columns of Tables 1 and 2. Notice that for Algebra 1 matching on prior achievement did not improve the balanced of the matched samples, but for grade 7 ELA/L the average difference in prior achievement between the matched tablet and non-tablet groups was reduced from .14 to .01 standard deviations. Thus, for some assessments, the omission of prior achievement from matching may have influenced results. Of course, the sample sizes were dramatically reduced when matching on prior achievement because those data were available only from two states. This report will focus on results from analyses based on CEM without prior achievement.

2.4 Task-Level Analysis

Two sets of analyses were conducted at the task level: a comparison of task means across conditions and a comparison of item response theory (IRT) difficulty estimates across conditions. The analyses were replicated for each grade/subject, and results were aggregated to draw inferences about differences in task difficulty between the tablet and computer conditions.

2.4.1 Comparison of Task Means

The relative difficulty of tasks in the NTC and TC conditions was investigated by computing mean scores on the tasks for the matched groups. Tasks with lower mean scores are more difficult than tasks with higher mean scores. Note that when a task is scored dichotomously (i.e., 0 = incorrect and 1 = correct), the mean is referred to as the "p-value," and it reflects the proportion of examinees who answered correctly. Each task's mean was computed for the NTC and TC conditions. A statistical test revealed whether differences in task means were statistically significant, and an effect size indicated practically significant differences. In order to maximize sample size, the comparison of task means was conducted separately for the PBA and EOY components (not for the small sample of students who took both the PBA and EOY study forms). This comparison involved the following steps:

1. Calculate the mean score for each task for each condition (i.e., p_{TC} and p_{NTC}) and calculate the difference between the task means $D_{TC–NTC} = p_{TC} – p_{NTC}$.

2. For each dichotomous task, run McNemar's test for dependent proportions to determine whether the p-values were significantly different (McNemar, 1947). This step involves setting up a 2×2 frequency table for each task and calculating a chi-squared test statistic according to equation (1).

	Item Score	NTC 1	NTC 0
TC	1	*a*	*b*
	0	*c*	*d*

a = number of students in the TC condition who got item *i* correct and whose match in the NTC condition got item *i* correct
b = number of students in the TC condition who got item *i* correct and whose match in the NTC condition got item *i* incorrect
c = number of students in the TC condition who got item *i* incorrect and whose match in the NTC condition got item *i* correct
d = number of students in the TC condition who got item *i* incorrect and whose match in the NTC condition got item *i* incorrect

$$\chi^2 = \frac{(b-c)^2}{b+c} \quad (1)$$

The test statistic is distributed approximately $\chi^2_{df=1}$, so if $\chi^2 > 3.841459$, the difference between p-values was considered statistically significant.

3. For each polytomous task, run a paired-sample *t*-test to determine whether task means were significantly different.

4. For each task, calculate an effect size for the difference between the task means using the method for paired samples (Dunlap, Cortina, Vaslow, & Burke, 1996).

$$var(D_{TC-NTC}) = \frac{1}{N}(s_{TC}^2 + s_{NTC}^2 - 2s_{TC}s_{NTC}r_{TC,NTC}) \quad (2)$$

$$t = \frac{D_{TC-NTC}}{\sqrt{var(D_{TC-NTC})}} \quad (3)$$

$$\text{Effect size} = t\sqrt{\frac{2(1-r_{TC,NTC})}{N}} \quad (4)$$

5. Identify the tasks with significant differences between task means according to Steps 2 and 3. Such tasks were flagged as having significantly different task means between the two device conditions.

6. Identify the tasks with significant differences between task means according to Steps 2 and 3 and for which the absolute value of the effect size was greater than 0.2 (Cohen, 1988). Such tasks were flagged as displaying a device effect for task means.

2.4.2 Comparison of IRT Difficulty Estimates

Like the comparison of task means, this analysis sought to compare tasks in terms of their difficulty across conditions. However, this analysis used IRT estimates of difficulty, which, unlike task means, are interval-scaled when the data fit the IRT model. As in the spring 2014 device comparability study (and as proposed for the spring 2015 study), item and step difficulty parameters for the Rasch partial credit model (RPCM) were estimated for each task for the computer and tablet conditions separately (Masters, 1982). Assuming that the matching procedure generated randomly equivalent groups, the IRT parameters from the separate calibrations should be on the same scale, and differences in IRT difficulty can be attributed to device effects.

In the RPCM, dichotomous tasks have a single measure of item difficulty, but polytomous items have several "step difficulties." For polytomous items, the mean step difficulty was treated as the measure of overall task difficulty. The standard errors of estimation from the Winsteps (Linacre, 2015) calibration were used to evaluate whether the IRT difficulty estimates were significantly different between the NTC and TC conditions. In order to maximize sample size,

the comparison of task means was conducted separately for the PBA and EOY components (not for the small sample of students who took both the PBA and EOY study forms). This comparison involved the following steps:

1. Calibrate the tasks using data from the TC condition.

2. Calibrate the tasks using data from the NTC condition.

3. For each task for each condition, calculate the difference between the IRT difficulties $D_{TC–NTC} = b_{TC} – b_{NTC}$.

4. For each task for each condition, use the standard error of the difficulty estimate to create a 95% confidence interval.

$$\begin{matrix} \hat{b}_{TC} \pm 1.96 se(\hat{b}_{TC}) \\ \hat{b}_{NTC} \pm 1.96 se(\hat{b}_{NTC}) \end{matrix} \quad (5)$$

5. Identify the tasks for which $\hat{b}_{TC}$ falls outside the range of the confidence interval for $\hat{b}_{NTC}$ and for which $\hat{b}_{NTC}$ falls outside the range of the confidence interval for $\hat{b}_{TC}$. Such tasks were flagged as having significantly different IRT difficulties between the two device conditions.

6. Identify the tasks with significant differences between IRT difficulties according to Step 3 and for which the absolute value of $D_{TC–NTC}$ was greater than 0.3 (Miller, Rotou, & Twing, 2004). Such tasks were flagged as displaying a device effect for IRT difficulties.

For ELA PBA tests, extra steps were necessary before conducting Steps 1–6 above. It was discovered that for the prose constructed response (PCR) tasks on the ELA/L PBA tests, zero or nearly zero students obtained the maximum possible score as well as certain scores below the maximum, which would result in irregular raw-score-to-theta tables and challenges in comparing results between administration modes. To address this issue, the score points were collapsed (as in the 2014 device comparability study). The score ranges 0–3, 4–7, 8–11, 12–15, and 16–19 were transformed to 0, 1, 2, 3, and 4, respectively.

2.5 Test-Level Analysis

The test-level comparability analysis focused on the properties of student proficiency estimates on the separate PARCC components (PBA and EOY) and on the full summative assessment (PBA+EOY). Five different factors were considered in evaluating comparability at the test level: correlations between component scores, raw and scale score distributions, reliability, validity, and score interpretations.

2.5.1 Correlations between Component Scores

The correlation between PBA and EOY component scores was computed separately for students in the tablet and computer conditions. A weaker correlation might have been

expected in the tablet condition because of potentially greater challenges in responding to constructed-response tasks in the PBA component (e.g., due to the smaller screen size and keyboard limitations). A Fisher *z* transformation (Fisher, 1921) was applied to the correlations so that differences between tablet and computer conditions could be evaluated for statistical significance. Cohen's *q*, which equals the difference between the *z*-transformed correlations (Cohen, 1988), was used to measure the effect size reflected by the difference between correlations.

The specific steps for conducting the component-level analyses for each subject were:

1. Compute the sample correlation between the PBA ability estimates and EOY ability estimates for the TC group (r_{TC}) and for the NTC group (rNTC). Use the ability estimates obtained in Steps 4a and 4b of Section 2.5.2 for this calculation.

2. Transform the sample correlations to get z_{TC} and z_{NTC} for the TC and NTC groups, respectively.

$$z = \frac{1}{2}\ln(\frac{1+r}{1-r}) \quad (6)$$

3. Calculate the effect size (Cohen's *q*) for the difference in correlations between conditions.

$$q = z_{TC} - z_{NTC} = D_{TC-NTC} \quad (7)$$

   To maintain consistency with other analyses, *q* can also be labeled D_{TC-NTC}.

4. Calculate the standard error of the difference between correlations

$$\sigma_{z_{TC}-z_{NTC}} = \sqrt{\frac{1}{N_{TC}-3} + \frac{1}{N_{NTC}-3}} \quad (8)$$

5. Calculate the *z* test statistic (z_{diff}).

$$z_{diff} = \frac{D_{TC-NTC}}{\sigma_{z_{TC}-z_{NTC}}} \quad (9)$$

6. Identify the tasks for which the absolute value of z_{diff} was greater than 1.96. Such tasks were flagged as having significantly different correlations between the device conditions.

7. Identify the tasks for which the absolute value of z_{diff} was greater than 1.96 and the absolute value of Cohen's *q* was greater than 0.1. Such tasks were flagged as displaying a device effect for the correlation between PBA and EOY.

2.5.2 Raw Score and Scale Score Distributions

An examination of score distributions indicates whether task-level device effects have any notable impact on the overall distribution of scores. For example, if the task-level device effects do not systematically favor a particular device, the device effects would cancel out and have little expected impact on scores. Separate comparisons were conducted for raw scores and IRT ability estimates.

For the IRT analysis, all of the data from the matched samples (NTC and TC) were to be entered into a single calibration. These data made up an "incomplete data matrix" (or IDM) in which some students had only PBA task scores, some student had only EOY task scores, and some students had both. Blank scores in the IDM were ignored during calibration. Calibrating with all the data (as would be done in the operational analysis) ensures that ability estimates would be comparable between the NTC and TC conditions. Had the calibrations been conducted separately (as in as in Section 2.4.2), similarity between the distributions would have been guaranteed on account of anchoring the calibration with a particular distribution of ability (mean of 0 and standard deviation of 1).

Note that, in rare circumstances, a task was spoiled (or DNU'd) on the Spanish form of a mathematics assessment, but not on the English form. In those cases, a score was imputed for students taking the Spanish version (using their proportion correct on non-spoiled tasks) to ensure that all raw scores for a given form had the same maximum possible value.

The steps in the comparison of score distributions were as follows:

1. Use the IDM to calibrate all PBA and EOY tasks.

2. Use the item parameter estimates from Step 1 and the matched PBA data set to run an anchored calibration. The output should include IRT ability estimates and person-fit statistics (for use in Section 2.6.1) for all students in the matched PBA data set.

3. Use the item parameter estimates from Step 1 and the matched EOY data set to run an anchored calibration. The output should include IRT ability estimates and person-fit statistics (for use in Section 2.6.1) for all students in the matched EOY data set.

4. Use the item parameter estimates from Step 1 and the matched PBA+EOY data set to run an anchored calibration. The output should include IRT ability estimates and person-fit statistics (for use in Section 2.6.1) for all students in the matched PBA+EOY data set.
    a. Use the item parameter estimates from Step 1 and the PBA task scores from the matched PBA+EOY data set to run an anchored calibration. The output should include PBA IRT ability estimates for all students in the matched PBA+EOY data set (for use in calculating the correlation between PBA and EOY in Section 2.5.1).
    b. Use the item parameter estimates from Step 1 and the EOY task scores from the matched PBA+EOY data set to run an anchored calibration. The output should

include EOY IRT ability estimates for all students in the matched PBA+EOY data set (for use in calculating the correlation between PBA and EOY in Section 2.5.1).

5. Calculate the mean and standard deviation for each combination of test, device condition, and component for each test. Use total raw scores for the raw score distributions. Use the IRT ability estimates from Steps 2, 3, and 4 for the theta (scale score) distributions.

6. For each combination of test, device condition, and component, calculate an effect size to reflect the average difference in performance in standard deviation units.

$$\frac{\bar{X}_{TC} - \bar{X}_{NTC}}{s_{pooled}} \tag{10}$$

$$s_{pooled} = \sqrt{\frac{(N_{TC} - 1)s_{TC}^2 + (N_{NTC} - 1)s_{NTC}^2}{N_{TC} + N_{NTC} - 2}} \tag{11}$$

$\bar{X}_{TC}$ is the mean of test scores for the TC group.
$\bar{X}_{NTC}$ is the mean of test scores for the NTC group.
s_{TC}^2 is the variance of test scores for the TC group.
s_{NTC}^2 is the variance of test scores for the NTC group.
N_{TC} is the sample size of the TC group.
N_{NTC} is the sample size of the TC group.

2.5.3 Reliability

Due to possible differences between correlations among tasks in the NTC and TC conditions, device effects may be manifested in differences in reliability. The reliability of test scores for the study conditions was compared using stratified alpha as an estimate of internal consistency reliability (Feldt & Brennan, 1989). Stratified alpha, which is a weighted composite of coefficient alphas for subsets of homogeneous items, is appropriate for mixed-format tests with items having different numbers of score points. This analysis was done by subject (ELA or mathematics), grade (or EOC test), and assessment component (PBA, EOY, and PBA+EOY). The statistical significance of differences in reliability coefficients was determined using the *W* statistic (Feldt, 1969), and effect sizes were calculated using Δ (Liu & Weng, 2009). The specific steps for conducting the reliability analysis were:

1. For each set of tasks within a component having the same maximum number of score points (i.e., a "stratum"), calculate coefficient alpha using the equation

$$\alpha = \frac{K}{K-1} \times \left(1 - \frac{\sum_{i=1}^{K} \sigma_{Yi}^2}{\sigma_X^2}\right) \tag{12}$$

where *K* is the number of the tasks within a stratum, σ_X^2 is the variance of the observed stratum total scores, and σ_{Yi}^2 is the variance of item *i*.

Note: In certain cases, there was only one task with a certain maximum score, and it would not have been possible to treat this task as its own stratum. For this reason, all PCR scores were treated as part of a single stratum even if those tasks had different numbers of total points. Similar compromises were also necessary on other tests. For example, in grades 5 and 7 mathematics PBA, there was only one item with a maximum score of 6, so it was combined with the tasks having a maximum score of 4 for the purposes of estimating reliability.

2. Use the coefficient alphas from Step 1 as input when computing stratified alpha.

$$\propto_{strat} = 1 - \frac{\sum_j s_{X_j}^2 (1 - \propto_j)}{s_X^2} \quad (13)$$

where $s_{X_j}^2$ is the variance for stratum *j*, s_X^2 is the total variance of the test, and $\propto_j$ is coefficient alpha for stratum *j*.

3. Calculate the *W* statistic (Feldt, 1969).

$$W = \frac{1 - \alpha_{TC}}{1 - \alpha_{NTC}} \quad (14)$$

4. Calculate the effect size for the difference in alphas (Liu & Weng, 2009).

$$\Delta = \frac{(1/2)\ln(1 - \alpha_{TC}) - (1/2)\ln(1 - \alpha_{NTC})}{\sqrt{K/(2(K-1))}} \quad (15)$$

5. Identify the assessments for which *W* exceeded the value in the *F* distribution with $N_C - 1$ and $N_T - 1$ degrees of freedom corresponding to a probability of 0.95. Such assessments were flagged as having significantly different coefficient alphas between device conditions.

6. Identify the assessments for which *W* exceeded the value in the *F* distribution with $N_C - 1$ and $N_T - 1$ degrees of freedom corresponding to a probability of 0.95 and the absolute value of Δ was greater than 0.1. Such assessments were flagged as displaying device effects for reliability.

2.5.4 Validity

The two study conditions were compared in terms of validity coefficients by estimating the relationship between performance on the summative test score (i.e., PBA, EOY, or PBA+EOY) and a measure of prior achievement in the same content area. Similarity of the correlations across study conditions would be indicative of comparability. Validity coefficients were transformed to the *z*-score metric using a Fisher *z* transformation (Fisher, 1921), so that differences between the NTC and TC conditions could be evaluated for statistical significance. Cohen's *q* served as a measure of effect size.

1. Compute the sample correlations between

    - PBA scale scores and prior achievement using all examinees who took the PBA study form
    - EOY scale scores and prior achievement using all examinees who took the EOY study form
    - PBA+EOY scale scores and prior achievement using student who took both (PBA and EOY) study forms

2. Use equation (6) to transform the sample correlations to get z_{TC} and z_{NTC} for the TC and NTC conditions, respectively.

3. Use equation (7) to calculate the effect size (Cohen's *q* or $D_{TC–NTC}$) for the difference in correlations between conditions.

4. Use equation (8) to calculate the standard error of the difference between correlations.

5. Use equation (9) to calculate the *z* test statistic (z_{diff}).

6. Identify the tasks for which the absolute value of z_{diff} was greater than 1.96. Such tasks were flagged as having significantly different validity coefficients between the two device conditions.

7. Identify the tasks for which the absolute value of the z_{diff} was greater than 1.96 and the absolute value of Cohen's *q* was greater than 0.1. Such tasks were flagged as displaying a device effect for the validity coefficient between PBA and EOY.

2.5.5 Score Interpretations

In typical applications, IRT true-score equating (Kolen & Brennan, 2004) is used to estimate equivalent raw scores on two different forms of an assessment using the test characteristic curves (TCC), which indicate the relationship between IRT ability estimates and expected raw scores. For example, consider that the TCC of Test Form X indicates that an examinee with ability 0.73 is expected to earn a raw score of 14. The TCC of Test Form Y indicates that an examinee with ability 0.73 is expected to earn a raw score of 14.8. Thus, a raw score of 14 on

Form X corresponds to a raw score of 14.8 on Form Y. A table of raw-score equivalents can be generated with one row for each raw score. The general procedure is as follows:

1. Find the IRT ability corresponding to raw score *r* on Form X.
2. Find the raw score on Form Y corresponding to the IRT ability from the previous step.
3. Repeat for all *r*.

In this study, Form X and Form Y were actually the same form administered under different conditions (NTC and TC), and the table of raw-score equivalents indicated whether Forms X and Y should be treated as different forms due to a device effect. Let Form X be the NTC condition, and let Form Y be the TC condition. If there is no device effect, the equivalent raw scores on Forms X and Y should be the same (after rounding to the nearest integer). Tasks were omitted from this analysis for two reasons: the task could not be calibrated in at least one of the conditions (e.g., if the proportion correct was zero), or the task had different numbers of score points in the two conditions (i.e., if certain score categories were unobserved in one condition).

IRT true-score equating depends on the assumption that the Form X and Form Y item parameters are on the same scale, which should be true if the matching procedure was successful. Because IRT true-score equating requires two sets of item parameter estimates on different scales, the IRT parameter estimates generated in Section 2.4.2 were used for this analysis (rather than the single set of item parameter estimates generated in Section 2.5.2).

Practically significant differences were identified as those for which raw-score differences exceeded half a raw-score point (0.5). This criterion was based on the "differences that matter" criterion (Dorans & Feigenbaum, 1994; Dorans, Holland, Thayer, & Tateneni, 2003) and is intended to reflect the magnitude of effect that would cause a student's raw score to round up or down based on mode differences. The specific steps for this analysis were:

1. Conduct IRT true-score equating three times: for the PBA study form, the EOY study form, and the PBA+EOY study form.

2. For each table of IRT true-score equating results, calculate the difference between the raw-score equivalents for the TC condition and each integer raw score for the NTC condition.

3. Flag the raw-score points for which the absolute value of the difference between raw-score equivalents was greater than 0.5.

2.6 Student-Level Analysis

2.6.1 Person-Fit Statistics

It was suggested by the PARCC Technical Advisory Committee that device effects could potentially manifest themselves in person-fit statistics. Generally, a person-fit statistic indicates

whether a student's pattern of task scores is consistent with what would be expected, assuming that that the IRT model is correct. Poor person-fit commonly occurs when lower ability students perform well on some difficult tasks or when higher ability students perform poorly on some easier tasks (i.e., when observed scores deviate from expectations). The steps for this analysis were as follows:

1. Obtain the INFIT mean-squares person-fit statistics calculated in Steps 2, 3, and 4 of Section 2.5.2.

2. Calculate means, standard deviations, and an effect size for each comparison of INFIT statistics.

3. For each set of person-fit statistics, calculate the proportion of students falling into the following ranges:

Interpretation of parameter-level mean-square fit statistics:[1]	
> 2.0	Distorts or degrades the measurement system.
1.5–2.0	Unproductive for construction of measurement, but not degrading.
0.5–1.5	Productive for measurement.
< 0.5	Less productive for measurement, but not degrading. May produce misleadingly good reliabilities and separations.

[1] http://www.winsteps.com/winman/diagnosingmisfit.htm

Section 3: Results

3.1 Task-Level Analysis

PARCC assessment tasks were compared in terms of two measures of task difficulty: mean task performance and IRT difficulty estimates. Similarity in task difficulty between the tablet and non-tablet device conditions would provide evidence consistent with comparability between tablets and computers, and differences in task difficulty would reveal a potential lack of comparability.

3.1.1 Comparison of Task Means

Differences in task means were flagged as statistically significant using McNemar's test for dichotomous tasks or a paired-sample *t*-test for polytomous tasks. A task was flagged as showing evidence of a device effect if the difference between task means exceeded 0.20 standard deviations. Results from the comparison of task means are summarized in Table 3. Due to large sample sizes, the statistical tests were sensitive to small differences and resulted in several tasks being flagged on each assessment. At most, 27% of tasks were flagged for significant differences on an assessment (grade 9 ELA/L), all of which were positive in that case, suggesting that tasks were easier for tablet testers. On most assessments, the magnitudes of differences tended to be very small (e.g., .00, .01, .02), and zero tasks were actually flagged for device effects. The full results of the comparison of task means are provided in Appendix B.

Table 3
Summary of Differences Between Task Means (Matching Only on Demographics)

Course	Component	# Tasks	# Sig.	# Device Effects
Grade 5 Math	PBA	16	2	0
Grade 5 Math	EOY	36	8	0
Grade 7 Math	PBA	17	7	0
Grade 7 Math	EOY	33	5	0
Algebra 1	PBA	18	2	0
Algebra 1	EOY	35	5	0
Geometry	PBA	18	5	0
Geometry	EOY	35	4	0
Algebra 2	PBA	20	2	0
Algebra 2	EOY	32	5	0
Grade 3 ELA/L	PBA	20	1	0
Grade 3 ELA/L	EOY	12	1	0
Grade 7 ELA/L	PBA	22	7	0
Grade 7 ELA/L	EOY	22	2	0
Grade 9 ELA/L	PBA	23	8	0
Grade 9 ELA/L	EOY	22	4	0

3.1.2 Comparison of IRT Difficulty Estimates

To obtain IRT difficulty estimates, the tasks were calibrated separately for the TC and NTC matched groups. If the groups are well matched, their ability distributions should be similar, and the resulting IRT difficulty estimates should be on the same scale, which would make them directly comparable. The standard errors of estimation were used to determine whether differences should be considered statistically significant. Any tasks whose absolute difference in IRT difficulty exceeded 0.3 was flagged as showing evidence of a device effect.

Large sample sizes provided higher precision when estimating the IRT difficulties, and this resulted in a large number of tasks having statistically significant differences in difficulty. Grade 3 ELA/L had the smallest proportion of tasks flagged for significant differences (31%), and grade 9 ELA/L had the largest proportion (47%). Overall, PBA tasks were flagged for significant differences at a much higher rate than EOY tasks (59% vs. 25%).

Table 4
Summary of Differences Between IRT Difficulties (Matching Only on Demographics)

Course	Component	No. Tasks	No. Sig.	No. Device Effect
Grade 5 Math	PBA	16	15	1
Grade 5 Math	EOY	36	6	0
Grade 7 Math	PBA	17	10	0
Grade 7 Math	EOY	33	11	0
Algebra 1	PBA	18	16	4
Algebra 1	EOY	35	4	0
Geometry	PBA	18	8	4
Geometry	EOY	35	9	7
Algebra 2	PBA	20	7	2
Algebra 2	EOY	32	11	5
Grade 3 ELA/L	PBA	20	8	0
Grade 3 ELA/L	EOY	12	2	0
Grade 7 ELA/L	PBA	22	13	0
Grade 7 ELA/L	EOY	22	6	0
Grade 9 ELA/L	PBA	23	14	0
Grade 9 ELA/L	EOY	22	7	0

Despite sometimes large proportions of tasks with significant differences, zero ELA/L tasks were flagged for device effects. There were several mathematics tasks flagged for device effects, in particular on the Geometry and Algebra 2 assessments. For Geometry, 9 out of 11 device

effects suggested higher difficulty for tablet testers. (In Appendix C, notice that many *b* Diff. values are positive in Table C.4.) Only 2 out of 7 Algebra 2 device effects suggested higher difficulty for tablet testers.

3.2 Test-Level Analysis

3.2.1 Correlations Between Component Scores

The relationship between performance on the separate EOY and PBA components was examined in the student-level analysis. Specifically, the correlation between EOY and PBA was computed for the TC and NTC conditions separately. Assessments were flagged for possible device effects when the difference between the correlations was statistically significant (i.e., $|z_{diff}| > 1.96$) and practically significant (i.e., Cohen's *q* effect size > 0.1). Differences in the correlations could indicate that some aspects of testing on tablets differentially affected the way students demonstrated their mathematics or ELA/L proficiencies on the EOY and PBA components. However, there was no pattern of results across assessments to support this notion. Table 5 shows the correlation between EOY and PBA for each assessment. Although two of the effect sizes exceeded 0.10 in magnitude (Geometry and grade 9 ELA/L), none of the differences between correlations were statistically significant. The largest difference in correlations occurred for the Geometry assessment. There were, however, only 53 students in the matched PBA+EOY Geometry data set, so sampling variability in the correlations should be expected.

Table 5
Comparison of Correlations Between PBA and EOY Scores (Matching Only on Demographics)

Course	N	*r* TC	*r* NTC	*r* Diff.	z_{diff}	Cohen's *q*	Sig.	Device Effect
Grade 5 Math	226	.82	.82	.00	-0.01	0.00		
Grade 7 Math	879	.80	.80	.00	0.04	0.00		
Algebra 1	207	.69	.67	.01	0.26	0.03		
Geometry	53	.70	.62	.08	0.71	0.14		
Algebra 2	151	.73	.70	.03	0.56	0.07		
Grade 3 ELA/L	399	.77	.76	.01	0.32	0.02		
Grade 7 ELA/L	423	.81	.83	-.01	-0.51	-0.04		
Grade 9 ELA/L	172	.81	.76	.05	1.10	0.12		

3.2.2 Raw Score and Scale Score Distributions

The task-level analyses identified potential device effects on individual tasks. In the aggregate, those effects could cancel out and have negligible impacts on estimates of student proficiency, or they could accumulate and bias students' raw scores and scale scores. For this analysis, all students in the matched samples (PBA, EOY, and PBA+EOY) were gathered into a single data set, and the tasks were recalibrated (as would be done in the analysis of operational

administration data). Item parameters from this new calibration were then used to obtain IRT ability estimates for students in the matched samples. If the samples were well matched, differences between the distributions of raw scores and scale scores could be attributed to the aggregate impact of device effects. Results from the analysis of mathematics and ELA/L assessments are shown in Tables 6 and 7, respectively. Note that the effect sizes, reflecting average differences in performance between the TC and NTC conditions, were very similar for raw scores and scale scores (never differing by more than .03).

Considering the PBA and EOY assessments separately, the effect sizes for grade 5 mathematics, grade 9 mathematics, and Algebra 1 were very small and not in a consistent direction, which suggests no device effects. The effect size for Geometry PBA was somewhat larger, but still small (–0.13), suggesting lower performance for tablet testers. This finding is consistent with the analysis of IRT difficulties in Section 3.1.2, which identified nine tasks with device effects favoring non-tablet testers. For Algebra 2, the effect sizes consistently suggested higher performance for tablet testers, but their magnitudes were very small (0.05). The effect sizes for grade 3 ELA/L were very small (-0.01 and 0.04). For grade 7 ELA/L and grade 9 ELA/L, the effect sizes consistently suggested higher performance by tablet testers, but the magnitudes were small.

For some assessments, the size of the effect changed when analyzing the combined PBA+EOY assessment. Recall that PBA+EOY matching was conducted independently and that the PBA+EOY matched samples include only students who took PBA and EOY in the same condition (TC or NTC). In several states, the number of PBA tablet testers differed substantially from the number of EOY tablet testers. Thus, the PBA+EOY sample would not necessarily be representative of the separate PBA and EOY samples. The effect size for grade 9 ELA/L PBA+EOY was larger than any others (0.23), but it was based on a relatively small sample of 172 pairs of matched students, so variability should be expected. Nevertheless, this result is consistent with the analysis of task means in Section 3.1.1, which revealed that grade 9 ELA/L tasks tended to be easier for tablet testers.

Table 6

Comparison of Score Distributions Between Matched TC and NTC Testers for Mathematics Assessments (Matching Only on Demographics)

Course	Condition	Component	N	Mean Raw Score	SD Raw Score	Raw Score Effect Size	Mean Scale Score	SD Scale Score	Scale Score Effect Size
Grade 5 Math	TC	PBA	3242	9.81	6.28		-0.84	1.26	
Grade 5 Math	NTC	PBA	3242	9.99	6.37	-0.03	-0.80	1.26	-0.03
Grade 5 Math	TC	EOY	1544	20.92	8.74		-0.18	1.26	
Grade 5 Math	NTC	EOY	1544	20.48	8.65	0.05	-0.24	1.25	0.05
Grade 5 Math	TC	PBA+EOY	226	34.43	14.50		-0.15	1.11	
Grade 5 Math	NTC	PBA+EOY	226	33.24	13.46	0.08	-0.24	1.02	0.08
Grade 7 Math	TC	PBA	3382	9.34	6.75		-1.35	1.44	
Grade 7 Math	NTC	PBA	3382	9.24	6.55	0.02	-1.36	1.43	0.01
Grade 7 Math	TC	EOY	3721	12.53	7.17		-1.29	1.33	
Grade 7 Math	NTC	EOY	3721	12.54	6.73	0.00	-1.27	1.27	-0.01
Grade 7 Math	TC	PBA+EOY	879	19.19	12.57		-1.55	1.26	
Grade 7 Math	NTC	PBA+EOY	879	20.31	12.34	-0.09	-1.42	1.23	-0.10
Algebra 1	TC	PBA	1445	5.10	4.24		-2.16	1.34	
Algebra 1	NTC	PBA	1445	5.08	4.48	0.00	-2.21	1.38	0.03
Algebra 1	TC	EOY	1388	13.02	6.97		-1.86	0.94	
Algebra 1	NTC	EOY	1388	13.50	7.12	-0.07	-1.79	0.93	-0.08
Algebra 1	TC	PBA+EOY	207	18.35	11.56		-1.89	0.98	
Algebra 1	NTC	PBA+EOY	207	17.71	11.00	0.06	-1.95	0.98	0.06
Geometry	TC	PBA	688	5.32	4.72		-2.16	1.22	
Geometry	NTC	PBA	688	5.84	4.77	-0.11	-1.99	1.21	-0.13
Geometry	TC	EOY	467	14.29	8.58		-1.62	1.08	
Geometry	NTC	EOY	467	14.70	9.04	-0.05	-1.56	1.11	-0.05
Geometry	TC	PBA+EOY	53	22.00	14.95		-1.56	1.02	
Geometry	NTC	PBA+EOY	53	22.66	12.49	-0.05	-1.46	0.87	-0.11

Algebra 2	TC	PBA	772	8.27	6.74		-2.09	1.25	
Algebra 2	NTC	PBA	772	7.86	6.38	0.06	-2.15	1.23	0.05
Algebra 2	TC	EOY	691	13.15	7.33		-1.87	1.02	
Algebra 2	NTC	EOY	691	12.76	6.87	0.06	-1.92	0.97	0.05
Algebra 2	TC	PBA+EOY	151	20.34	13.10		-1.99	0.98	
Algebra 2	NTC	PBA+EOY	151	19.62	12.81	0.06	-2.04	0.97	0.05

Table 7

Comparison of Score Distributions Between Matched TC and NTC Testers for ELA/L Assessments (Matching Only on Demographics)

Course	Condition	Component	N	Mean Raw Score	SD Raw Score	Raw Score Effect Size	Mean Scale Score	SD Scale Score	Scale Score Effect Size
Grade 3 ELA/L	TC	PBA	2223	19.49	10.13		-0.30	1.06	
Grade 3 ELA/L	NTC	PBA	2223	19.51	9.94	0.00	-0.29	1.02	-0.01
Grade 3 ELA/L	TC	EOY	2572	10.59	4.64		-0.26	0.94	
Grade 3 ELA/L	NTC	EOY	2572	10.44	4.58	0.03	-0.30	0.93	0.04
Grade 3 ELA/L	TC	PBA+EOY	399	28.08	13.78		-0.41	0.89	
Grade 3 ELA/L	NTC	PBA+EOY	399	29.33	13.88	-0.09	-0.33	0.89	-0.09
Grade 7 ELA/L	TC	PBA	2714	22.71	10.55		-0.13	0.87	
Grade 7 ELA/L	NTC	PBA	2714	21.91	10.25	0.08	-0.19	0.83	0.08
Grade 7 ELA/L	TC	EOY	3031	21.19	8.57		-0.12	0.81	
Grade 7 ELA/L	NTC	EOY	3031	20.94	8.51	0.03	-0.14	0.81	0.03
Grade 7 ELA/L	TC	PBA+EOY	423	43.13	18.05		-0.16	0.76	
Grade 7 ELA/L	NTC	PBA+EOY	423	43.00	18.82	0.01	-0.16	0.80	0.00
Grade 9 ELA/L	TC	PBA	1314	25.84	10.22		-0.16	0.89	
Grade 9 ELA/L	NTC	PBA	1314	24.74	10.22	0.11	-0.26	0.90	0.11
Grade 9 ELA/L	TC	EOY	1492	17.51	8.25		-0.23	0.72	
Grade 9 ELA/L	NTC	EOY	1492	16.86	8.48	0.08	-0.29	0.76	0.08
Grade 9 ELA/L	TC	PBA+EOY	172	44.88	17.40		-0.13	0.73	
Grade 9 ELA/L	NTC	PBA+EOY	172	41.05	17.36	0.22	-0.30	0.75	0.23

3.2.3 Reliability

Reliability is a reflection of measurement precision, which translates into consistency in assessment scores across repeated test administrations. In this study, reliability coefficients were computed separately for the tablet and non-tablet conditions. Assessments were flagged for possible device effects when the difference between the reliability coefficients was statistically significant (i.e., $W > F_{.05,N_{NTC}-1,N_{TC}-1}$) and practically significant (i.e., effect size Δ > 0.2). Statistically significant differences between the TC and NTC conditions in terms of reliability coefficients were detected in two instances: grade 7 ELA/L PBA+EOY and grade 9 ELA/L EOY (Table 8). In neither case, however, would the observed difference (–0.01 or –0.02) be considered of any practical significance.

Table 8

Comparison of Reliability Coefficients for TC and NTC Testers (Matching Only on Demographics)

Course	Component	α TC	α NTC	α Diff	*W*	Δ	Sig.	Device Effect
Grade 5 Math	PBA	.85	.86	.00	1.02	0.01		
Grade 5 Math	EOY	.90	.90	.00	0.98	-0.01		
Grade 5 Math	PBA+EOY	.93	.92	.01	0.91	-0.07		
Grade 7 Math	PBA	.86	.86	.01	0.95	-0.03		
Grade 7 Math	EOY	.89	.88	.01	0.88	-0.09		
Grade 7 Math	PBA+EOY	.93	.93	.01	0.90	-0.08		
Algebra 1	PBA	.79	.81	-.02	1.09	0.06		
Algebra 1	EOY	.84	.84	.00	1.03	0.02		
Algebra 1	PBA+EOY	.91	.90	.01	0.90	-0.07		
Geometry	PBA	.81	.79	.02	0.93	-0.05		
Geometry	EOY	.88	.90	-.01	1.14	0.09		
Geometry	PBA+EOY	.93	.90	.03	0.66	-0.29		
Algebra 2	PBA	.82	.81	.01	0.93	-0.05		
Algebra 2	EOY	.85	.83	.02	0.88	-0.09		
Algebra 2	PBA+EOY	.91	.91	.00	0.97	-0.02		
Grade 3 ELA/L	PBA	.90	.89	.01	0.92	-0.06		
Grade 3 ELA/L	EOY	.74	.73	.01	0.97	-0.02		
Grade 3 ELA/L	PBA+EOY	.92	.91	.01	0.94	-0.04		
Grade 7 ELA/L	PBA	.91	.90	.01	0.94	-0.04		
Grade 7 ELA/L	EOY	.85	.84	.00	0.99	-0.01		
Grade 7 ELA/L	PBA+EOY	.93	.94	-.01	1.19	0.12	*	*
Grade 9 ELA/L	PBA	.90	.90	.00	0.96	-0.03		
Grade 9 ELA/L	EOY	.82	.83	-.02	1.09	0.06	*	
Grade 9 ELA/L	PBA+EOY	.92	.93	-.01	1.15	0.10		

3.2.4 Validity

Valid interpretations and uses of test scores are commonly supported with evidence of "convergent" validity. That is, if scores from the test of interest correlate in expected ways with scores from other measures of the same construct, this provides evidence that scores from the test of interest can be interpreted as intended. For example, if PARCC grade 7 mathematics scores correlate with mathematics scores from the previous state achievement testing program, this would support the notion that the PARCC scores can be interpreted as indicators of mathematics proficiency. In this analysis, prior achievement on a state assessment served as the criterion measure (available only for certain states). EOY, PBA, and EOY+PBA scale scores were correlated with the criterion measure for the TC and NTC conditions separately within each state (correlations across states would not have been possible since the criterion test differed). Assessments were flagged for possible device effects when the difference between the correlations was statistically significant (i.e., $|z_{diff}| > 1.96$) and practically significant (i.e., Cohen's q effect size > 0.1).

As shown in Tables 8 and 9, the correlations between PARCC scale scores and prior achievement were generally in the .70 to .90 range. Note that sample sizes sometimes differed slightly for the PBA and EOY correlations because prior achievement data were not available for every student. Only two of the differences between correlations in the TC and NTC conditions were statistically and practically significant: grade 7 mathematics PBA+EOY in State B and grade 7 ELA/L PBA+EOY in State B. In both cases, the sample size was quite small, so greater sampling variability should be expected. Overall, the statistical evidence indicates that validity coefficients were the same in the TC and NTC conditions, which is consistent with the comparability of testing on tablets and non-tablet devices.

Table 8

Comparison of Validity Coefficients for TC and NTC Testers (Matching Only on Demographics)

Course	Component	State	N TC	N NTC	*r* TC	*r* NTC	*r* Diff.	z_{diff}	Cohen's *q*	Sig.	Device Effect
Grade 5 Math	PBA	B	310	306	.76	.79	-.03	-0.79	-0.06		
Grade 5 Math	EOY	B	42	42	.68	.84	-.16	-1.78	-0.40		
Grade 5 Math	PBA+EOY	B	3	3	.72	.83	-.11	-0.29			
Grade 7 Math	PBA	A	504	545	.81	.80	.01	0.31	0.02		
Grade 7 Math	PBA	B	262	247	.78	.75	.03	0.85	0.08		
Grade 7 Math	EOY	A	546	582	.82	.81	.00	0.17	0.01		
Grade 7 Math	EOY	B	195	192	.80	.78	.02	0.58	0.06		
Grade 7 Math	PBA+EOY	A	91	94	.86	.88	-.02	-0.63	-0.09		
Grade 7 Math	PBA+EOY	B	35	35	.93	.79	.13	2.19	0.55	*	*
Algebra 1	PBA	A	56	55	.81	.67	.14	1.65	0.32		
Algebra 1	PBA	B	122	119	.74	.75	-.01	-0.23	-0.03		
Algebra 1	EOY	A	117	118	.72	.73	-.01	-0.13	-0.02		
Algebra 1	EOY	B	107	108	.70	.80	-.10	-1.74	-0.24		
Algebra 1	PBA+EOY	A	4	4	.69	-.73	1.41	1.25	1.76		
Algebra 1	PBA+EOY	B	15	16	.85	.91	-.05	-0.58	-0.23		
Geometry	PBA	A	10	10	.89	.39	.50	1.86	0.99		
Algebra 2	PBA	B	34	28	.41	.68	-.27	-1.46	-0.39		
Algebra 2	EOY	B	17	15	.61	.66	-.05	-0.22	-0.09		
Algebra 2	PBA+EOY	B	5	4	.92	.87	.05	0.23	0.28		

Table 9

Comparison of Validity Coefficients for TC and NTC Testers (Matching Only on Demographics)

Course	Component	State	N TC	N NTC	*r* TC	*r* NTC	*r* Diff.	z_{diff}	Cohen's *q*	Sig.	Device Effect
Grade 7 ELA/L	PBA	A	423	447	.77	.75	.02	0.84	0.06		
Grade 7 ELA/L	PBA	B	220	211	.76	.69	.07	1.55	0.15		
Grade 7 ELA/L	EOY	A	471	514	.74	.73	.01	0.30	0.02		
Grade 7 ELA/L	EOY	B	181	175	.76	.67	.09	1.75	0.19		
Grade 7 ELA/L	PBA+EOY	A	59	65	.68	.77	-.09	-1.05	-0.19		
Grade 7 ELA/L	PBA+EOY	B	26	25	.85	.41	.43	2.72	0.81	*	*
Grade 9 ELA/L	PBA	A	33	26	.49	.40	.10	0.44	0.12		
Grade 9 ELA/L	EOY	A	46	45	.67	.43	.24	1.61	0.35		
Grade 9 ELA/L	PBA+EOY	A	5	5	.91	.36	.56	1.17	1.17		

3.2.5 Score Interpretations

In this analysis, IRT true-score equating was used to determine whether raw scores on the tablet and non-tablet device versions of the PARCC assessments could be interpreted as indicating the same level of proficiency. For each assessment, a concordance table was generated to show raw-score equivalents on the tablet and non-tablet device versions of the PARCC assessments. Raw-score points were flagged for possible device effects when the difference between the equivalent tablet and non-tablet device raw scores exceeded 0.5 points. Results of this analysis are summarized in Table 10, and full results are provided in Appendix D.

Considering the PBA and EOY assessments separately, there was no evidence of device effects for grade 5 mathematics or grade 7 mathematics. There was minor evidence of device effects for Algebra 1 EOY, where non-tablet testers at most ability levels would have been expected to score one point lower if they had tested on tablets. Similar results were observed for Geometry and Algebra 2. For Geometry, non-tablet testers in certain ranges of ability would have been expected to score one point lower had they taken PBA and EOY on tablets. For Algebra 2 EOY, certain non-tablet testers would have been expected to score one point higher had they tested on tablets. Depending on the range of the ability scale, non-tablet testers would have been expected to score one point lower or higher had they taken Algebra 2 PBA on tablets. Given the small magnitude of device effects and the irregularity of their direction, the statistical evidence is consistent with the comparability of testing on tablets and non-tablet devices.

Results were much the same for the ELA/L assessments. There was no evidence of device effects on the separate grade 3 ELA/L PBA and EOY assessments. Non-tablet testers in certain ability ranges would have been expected to score one point higher on grade 7 ELA/L PBA had they tested on tablets, but there was no evidence of device effects for grade 7 ELA/L EOY. Evidence of minor device effects was observed on grade 9 ELA/L PBA and EOY, where non-tablet testers in certain ranges of ability would have been expected to score one point higher had they tested on tablets. As with the mathematics assessments, the apparent device effects were small in magnitude. In the case of ELA/L, most device effects suggested that testing on tablets was easier, which is unexpected given that tablets have smaller screens and possibly unfamiliar input devices (i.e., touchscreens and small keyboards).

For many assessments, there was evidence of larger device effects for the combined PBA+EOY assessment. As explained in Section 3.2.2, the PBA+EOY sampling was conducted separately, and the PBA+EOY sample is not necessarily representative of the separate PBA and EOY samples. Moreover, the sample sizes for the PBA+EOY calibrations were notably smaller, which would lead to greater estimation error in the item parameters, especially for polytomous tasks with many score categories. Errors in item parameter estimates could translate into apparent device effects in the IRT true-score equating results. The largest device effects were observed on grade 7 mathematics PBA+EOY, where students would have been expected to perform up to four points lower if testing on tablets, and on grade 9 ELA/L, where students would have been expected to score up to 5 point higher if testing on tablets.

Table 10
Summary of Score Interpretation Analysis (Matched Only on Demographics)

Course	Summary
Grade 5 Math	PBA: No device effects EOY: No device effects PBA+EOY: Effect at 49/77 raw scores, max. effect 2
Grade 7 Math	PBA: No device effects EOY: No device effects PBA+EOY: Effect at 72/83 raw scores, max. effect -4
Algebra 1	PBA: No device effects EOY: Effect at 34/56 raw scores, max. effect -1 PBA+EOY: Effect at 48/72 raw scores, max. effect 2
Geometry	PBA: Effect at 22/33 raw scores, max. effect -1 EOY: Effect at 12/56 raw scores, max. effect -1 PBA+EOY: Effect at 57/70 raw scores, max. effect -2
Algebra 2	PBA: Effect at 10/41 raw scores, max. effect ±1 EOY: Effect at 25/49 raw scores, max. effect 1 PBA+EOY: Effect at 56/83 raw scores, max. effect 1
Grade 3 ELA/L	PBA: No device effects EOY: No device effects PBA+EOY: Effect at 51/68 raw scores, max. effect -1
Grade 7 ELA/L	PBA: Effect at 33/50 raw scores, max. effect 1 EOY: No device effects PBA+EOY: No device effects
Grade 9 ELA/L	PBA: Effect at 38/52 raw scores, max. effect 1 EOY: Effect at 27/45 raw scores, max. effect 1 PBA+EOY: Effect at 89/96 raw scores, max. effect 5

3.3 Student-Level Analysis

3.3.1 Person-Fit Statistics

Device effects could be manifested in person-fit statistics if, for example, a few tasks were particularly difficult for tablet testers but easy for non-tablet device testers. Such effects on person-fit statistics would not be observed if the device effect was consistent across tasks. Results from this analysis are shown in Tables 11 and 12, which show descriptive statistics for INFIT statistics. The expected value of INFIT statistics is 1.0, and students having INFIT statistics in the range of 0.5–1.5 are considered productive for measurement. INFIT statistics were computed separately for each of the matched samples (PBA, EOY, and PBA+EOY) for both testing conditions. The focus of this analysis is the percentage of students in the 0.5–1.5 range.

As shown in Tables 11 and 12, the average INFIT statistic never deviated by more than 0.07 from the expected value of 1.0. In general, the average difference in INFIT between the TC and NTC conditions was small, with most effect sizes smaller than 0.05. A few effect sizes were relatively large. For example, students taking Geometry PBA on tablets had slightly higher average INFIT, but this was mainly a reflection of having more students in the non-tablet condition whose data fit the model unexpectedly well (i.e., with INFIT < 0.5). The difference in the percentage of students with INFIT between 0.5 and 1.5 never exceeded 3.9% (Algebra 1 PBA+EOY), and the direction of differences was irregular. Thus, the evidence is consistent with the comparability of tablet and non-tablet testing in terms of person-fit statistics.

Table 11

Distributions of INFIT Statistics and Percentages of Examinees in INFIT Categories for Mathematics (Matched Only on Demographics)

Course	Component	Condition	N	Mean	SD	Effect Size	< 0.5	0.5–1.5	1.5–2.0	> 2.0
Grade 5 Math	PBA	TC	3242	0.97	0.41		7.0	83.5	7.0	2.5
Grade 5 Math	PBA	NTC	3242	0.97	0.39	0.01	7.3	83.2	7.6	1.9
Grade 5 Math	EOY	TC	1544	1.01	0.29		0.6	93.2	5.5	0.7
Grade 5 Math	EOY	NTC	1544	1.01	0.26	0.03	0.3	95.0	4.5	0.1
Grade 5 Math	PBA+EOY	TC	226	1.09	0.33		0.0	89.8	8.0	2.2
Grade 5 Math	PBA+EOY	NTC	226	1.00	0.25	0.30	0.0	96.5	3.5	0.0
Grade 7 Math	PBA	TC	3382	0.95	0.39		7.6	84.8	5.8	1.8
Grade 7 Math	PBA	NTC	3382	0.94	0.39	0.02	8.0	84.1	6.0	1.9
Grade 7 Math	EOY	TC	3721	1.00	0.30		1.6	92.0	5.8	0.6
Grade 7 Math	EOY	NTC	3721	1.00	0.31	0.00	2.2	90.8	6.3	0.7
Grade 7 Math	PBA+EOY	TC	879	1.02	0.27		0.7	94.0	5.2	0.1
Grade 7 Math	PBA+EOY	NTC	879	1.03	0.27	-0.02	0.7	93.4	5.6	0.3
Algebra 1	PBA	TC	1445	0.95	0.40		5.5	86.4	5.7	2.3
Algebra 1	PBA	NTC	1445	0.97	0.43	-0.04	5.1	85.6	6.2	3.0
Algebra 1	EOY	TC	1388	0.99	0.27		0.7	95.4	3.4	0.5
Algebra 1	EOY	NTC	1388	0.98	0.29	0.00	1.2	94.1	4.0	0.6
Algebra 1	PBA+EOY	TC	207	1.01	0.23		0.0	96.1	3.9	0.0
Algebra 1	PBA+EOY	NTC	207	1.04	0.29	-0.13	0.0	92.3	6.8	1.0
Geometry	PBA	TC	688	0.98	0.47		6.5	84.2	5.4	3.9
Geometry	PBA	NTC	688	0.93	0.40	0.12	9.7	82.8	5.2	2.2
Geometry	EOY	TC	467	1.00	0.31		0.6	92.9	5.4	1.1
Geometry	EOY	NTC	467	0.98	0.27	0.06	0.4	96.1	2.8	0.6
Geometry	PBA+EOY	TC	53	1.06	0.31		0.0	92.5	5.7	1.9
Geometry	PBA+EOY	NTC	53	1.03	0.31	0.07	0.0	92.5	5.7	1.9

Algebra 2	PBA	TC	772	0.98	0.46		8.2	79.1	8.9	3.8
Algebra 2	PBA	NTC	772	0.99	0.52	-0.03	11.5	75.5	8.4	4.5
Algebra 2	EOY	TC	691	1.00	0.35		2.0	89.3	6.9	1.7
Algebra 2	EOY	NTC	691	0.97	0.32	0.08	2.6	91.8	4.2	1.4
Algebra 2	PBA+EOY	TC	151	1.02	0.34		2.0	91.4	5.3	1.3
Algebra 2	PBA+EOY	NTC	151	1.04	0.38	-0.05	1.3	88.1	7.9	2.6

Table 12

Distributions of INFIT Statistics and Percentages of Examinees in INFIT Categories for ELA/L (Matched Only on Demographics)

Course	Component	Condition	N	Mean	SD	Effect Size	< 0.5	0.5–1.5	1.5–2.0	> 2.0
Grade 3 ELA/L	PBA	TC	2223	1.00	0.30		1.8	93.2	4.1	1.0
Grade 3 ELA/L	PBA	NTC	2223	1.00	0.30	0.01	1.9	93.0	3.8	1.3
Grade 3 ELA/L	EOY	TC	2572	0.96	0.34		5.4	87.8	5.9	0.9
Grade 3 ELA/L	EOY	NTC	2572	0.97	0.34	-0.04	5.0	87.1	7.2	0.7
Grade 3 ELA/L	PBA+EOY	TC	399	1.00	0.22		0.0	98.0	1.5	0.5
Grade 3 ELA/L	PBA+EOY	NTC	399	1.00	0.21	0.01	0.5	97.7	1.8	0.0
Grade 7 ELA/L	PBA	TC	2714	0.99	0.26		1.0	94.7	3.9	0.3
Grade 7 ELA/L	PBA	NTC	2714	0.98	0.25	0.04	1.0	95.7	3.2	0.2
Grade 7 ELA/L	EOY	TC	3031	1.00	0.23		0.5	96.9	2.5	0.1
Grade 7 ELA/L	EOY	NTC	3031	1.00	0.23	0.02	0.3	97.3	2.4	0.0
Grade 7 ELA/L	PBA+EOY	TC	423	1.01	0.16		0.0	99.5	0.5	0.0
Grade 7 ELA/L	PBA+EOY	NTC	423	1.01	0.18	-0.02	0.0	98.1	1.9	0.0
Grade 9 ELA/L	PBA	TC	1314	1.00	0.28		1.1	94.2	4.1	0.6
Grade 9 ELA/L	PBA	NTC	1314	1.01	0.28	-0.02	1.4	93.7	4.3	0.7
Grade 9 ELA/L	EOY	TC	1492	0.99	0.21		0.1	98.3	1.5	0.0
Grade 9 ELA/L	EOY	NTC	1492	0.98	0.21	0.05	0.0	98.5	1.5	0.0
Grade 9 ELA/L	PBA+EOY	TC	172	1.01	0.17		0.0	99.4	0.6	0.0
Grade 9 ELA/L	PBA+EOY	NTC	172	1.01	0.18	0.00	0.0	100.0	0.0	0.0

3.4 Results Summary

Tables 12 and 13 summarize the results from all nine analyses for each assessment. By scanning across each row, the reader can consider the quantity and strength of the evidence for the comparability of testing on tablets and non-tablet devices for each assessment.

Table 12
Mathematics Results Summary (Matching Only on Demographics)

Course	Matched N	Prior Achievement Balance	Task Means	IRT Difficulties	PBA-EOY Correlation
Grade 5 Mathematics	PBA: 3242 EOY: 1544 PBA+EOY: 226	PBA: N=299, 0.02 difference EOY: N=42, 0.11 difference, TC of higher ability PBA+EOY: N=3, -0.60 difference, TC of lower ability	10/52 tasks significant 0/52 tasks with device effect	21/52 tasks significant 1/52 task with device effect	TC: .82 NTC: .82 Not significant, no device effect
Grade 7 Mathematics	PBA: 3382 EOY: 3721 PBA+EOY: 879	PBA: N=712, -0.03 difference EOY: N=697, -0.03 difference PBA+EOY: N=117, -0.06 difference	12/50 tasks significant 0/50 tasks with device effect	21/50 tasks significant 0/50 tasks with device effect	TC: .80 NTC: .80 Not significant, no device effect
Algebra 1	PBA: 1445 EOY: 1388 PBA+EOY: 207	PBA: N=165, -0.04 difference EOY: N=206, -0.13 difference, TC of lower ability PBA+EOY: N=16, -0.04 difference	7/53 tasks significant 0/53 tasks with device effect	20/53 tasks significant 4/53 tasks with device effect 4/4 device effect tasks with higher difficulty for TC	TC: .69 NTC: .67 Not significant, no device effect
Geometry	PBA: 688 EOY: 467 PBA+EOY: 53	PBA: N=9, 0.27 difference, TC of higher ability EOY: No data PBA+EOY: No data	9/53 tasks significant 0/53 tasks with device effect	17/53 tasks significant 11/53 tasks with device effect 9/11 device effect tasks with higher difficulty for TC	TC: .70 NTC: .62 Not significant, no device effect
Algebra 2	PBA: 772 EOY: 691 PBA+EOY: 151	PBA: N=27, 0.32 difference, TC of higher ability EOY: N=15, -0.32 difference, TC of lower ability PBA+EOY: N=4, 0.01 difference	7/52 tasks significant 0/52 tasks with device effect	18/52 tasks significant 7/52 tasks with device effect 2/7 device effect tasks with higher difficulty for TC	TC: .73 NTC: .70 Not significant, no device effect

Table 12 (cont.)
Mathematics Results Summary (Matching Only on Demographics)

Course	Raw and Scale Score Distributions	Reliability	Validity	Score Interpretations	Person Fit
Grade 5 Mathematics	PBA: -0.03, -0.03 effect sizes EOY: 0.05, 0.05 effect sizes PBA+EOY: 0.08, 0.08 effect sizes	PBA: TC .85, NTC .86 EOY: TC .90, NTC .90 PBA+EOY: TC .93, NTC .92	PBA: No sig. differences EOY: No sig. differences PBA+EOY: No sig. differences	PBA: No device effects EOY: No device effects PBA+EOY: Effect at 49/77 raw scores, max. effect 2	PBA: Difference of 0.2% EOY: Difference of -1.8% PBA+EOY: Difference of -6.6%
Grade 7 Mathematics	PBA: 0.02, 0.01 effect sizes EOY: 0.00, -0.01 effect sizes PBA+EOY: -0.09, -0.10 effect sizes	PBA: TC .86, NTC .86 EOY: TC .89, NTC .88 PBA+EOY: TC .93, NTC .93	PBA: No sig. differences EOY: No sig. differences PBA+EOY: TC higher in state B, but small sample size	PBA: No device effects EOY: No device effects PBA+EOY: Effect at 72/83 raw scores, max. effect -4	PBA: Difference of 0.7% EOY: Difference of 1.2% PBA+EOY: Difference of 0.6%
Algebra 1	PBA: 0.00, 0.03 effect sizes EOY: -0.07, -0.08 effect sizes PBA+EOY: 0.06, 0.06 effect sizes	PBA: TC .79, NTC .81 EOY: TC .84, NTC .84 PBA+EOY: TC .91, NTC .90	PBA: No sig. differences EOY: No sig. differences PBA+EOY: No sig. differences	PBA: No device effects EOY: Effect at 34/56 raw scores, max. effect -1 PBA+EOY: Effect at 48/72 raw scores, max. effect 2	PBA: Difference of 0.8% EOY: Difference of 1.3% PBA+EOY: Difference of 3.9%
Geometry	PBA: -0.11, -0.13 effect sizes, TC of lower ability EOY: -0.05, -0.05 effect sizes PBA+EOY: -0.05, -0.11 effect sizes, TC of lower ability	PBA: TC .81, NTC .79 EOY: TC .88, NTC .90 PBA+EOY: TC .93, NTC .90	PBA: No sig. differences EOY: No data PBA+EOY: No data	PBA: Effect at 22/33 raw scores, max. effect -1 EOY: Effect at 12/56 raw scores, max. effect -1 PBA+EOY: Effect at 57/70 raw scores, max. effect -2	PBA: Difference of 1.3% EOY: Difference of -3.2% PBA+EOY: Difference of 0.0%
Algebra 2	PBA: 0.06, 0.05 effect sizes EOY: 0.06, 0.05 effect sizes PBA+EOY: 0.06, 0.05 effect sizes	PBA: TC .82, NTC .81 EOY: TC .85, NTC .83 PBA+EOY: TC .91, NTC .91	PBA: No sig. differences EOY: No sig. differences PBA+EOY: No sig. differences	PBA: Effect at 10/41 raw scores, max. effect ±1 EOY: Effect at 25/49 raw scores, max. effect 1 PBA+EOY: Effect at 56/83 raw scores, max. effect 1	PBA: Difference of 3.6% EOY: Difference of -2.5% PBA+EOY: Difference of 3.3%

Table 13

ELA/L Results Summary (Matching Only on Demographics)

Course	Matched N	Prior Achievement Balance	Task Means	IRT Difficulties	PBA-EOY Correlation
Grade 3 ELA/L	PBA: 2223 EOY: 2572 PBA+EOY: 399	PBA: No data EOY: No data PBA+EOY: No data	2/32 tasks significant 0/32 tasks with device effect	10/32 tasks significant 0/32 tasks with device effect	TC: .77 NTC: .76 Not significant, no device effect
Grade 7 ELA/L	PBA: 2714 EOY: 3031 PBA+EOY: 423	PBA: N=597, 0.08 difference EOY: N=609, 0.14 difference, TC of higher ability PBA+EOY: N=80, 0.05 difference	9/44 tasks significant 0/44 tasks with device effect	19/44 tasks significant 0/44 tasks with device effect	TC: .81 NTC: .83 Not significant, no device effect
Grade 9 ELA/L	PBA: 1314 EOY: 1492 PBA+EOY: 172	PBA: N=26, 0.13 difference, TC of higher ability EOY: N=43, -0.06 difference PBA+EOY: N=5, -0.33 difference, TC of lower ability	12/45 tasks significant 0/45 tasks with device effect	21/45 tasks significant 0/45 tasks with device effect	TC: .81 NTC: .76 Not significant, no device effect

Table 13 (cont.)
ELA/L Results Summary (Matching Only on Demographics)

Course	Raw and Scale Score Distributions	Reliability	Validity	Score Interpretations	Person Fit
Grade 3 ELA/L	PBA: 0.00, -0.01 effect sizes EOY: 0.03, 0.04 effect sizes PBA+EOY: -0.09, -0.09 effect sizes	PBA: TC .90, NTC .89 EOY: TC .74, NTC .73 PBA+EOY: TC .92, NTC .91	No data	PBA: No device effects EOY: No device effects PBA+EOY: Effect at 51/68 raw scores, max. effect -1	PBA: Difference of 0.2% EOY: Difference of 0.7% PBA+EOY: Difference of 0.3%
Grade 7 ELA/L	PBA: 0.08, 0.08 effect sizes EOY: 0.03, 0.03 effect sizes PBA+EOY: 0.01, 0.00 effect sizes	PBA: TC .91, NTC .90 EOY: TC .85, NTC .84 PBA+EOY: TC .93, NTC .94, device effect	PBA: No sig. differences EOY: No sig. differences PBA+EOY: TC higher in state B, but small sample size	PBA: Effect at 33/50 raw scores, max. effect 1 EOY: No device effects PBA+EOY: No device effects	PBA: Difference of -0.9% EOY: Difference of -0.4% PBA+EOY: Difference of 1.4%
Grade 9 ELA/L	PBA: 0.11, 0.11 effect sizes, TC of higher ability EOY: 0.08, 0.08 effect sizes PBA+EOY: 0.22, 0.23 effect sizes, TC of higher ability	PBA: TC .90, NTC .90 EOY: TC .82, NTC .83, sig. PBA+EOY: TC .92, NTC .93	PBA: No sig. differences EOY: No sig. differences PBA+EOY: No sig. differences	PBA: Effect at 38/52 raw scores, max. effect 1 EOY: Effect at 27/45 raw scores, max. effect 1 PBA+EOY: Effect at 89/96 raw scores, max. effect 5	PBA: Difference of 0.5% EOY: Difference of -0.1% PBA+EOY: Difference of -0.6%

Section 4: Conclusions and Implications

4.1 Addressing Research Questions

Overall, the statistical evidence generated by this study is consistent with the comparability of PARCC testing on tablets and non-tablet devices. Results from all analyses are summarized in Tables 12 and 13 of the preceding section. The prior device comparability study (analyzing data from the 2014 field test) reached the same general conclusion, although some device effects were observed at lower grades (e.g., grade 4 mathematics and grade 4 ELA/L). No such evidence was detected in the current study.

The first research question concerned whether student performance on individual tasks was similar when comparing tasks administered on tablets and non-tablet computing devices. The analysis of task means identified zero tasks with "device effects" as defined by the criteria used in this study. The statistical tests and criteria used for flagging device effects on the IRT difficulties were apparently more sensitive because a larger number of differences in IRT difficulty were flagged as statistically significant and some tasks were flagged for device effects. The flagged tasks in Algebra 1 and Geometry were generally more difficult for students testing on tablets, but the flagged tasks in Algebra 2 were apparently more difficult for students testing on non-tablet devices.

The next research question addressed similarity in the psychometric properties of test scores from tests administered on tablets and non-tablet computing devices. The test-level analysis revealed that correlations between the EOY and PBA components were similar in the tablet and non-tablet conditions for all eight assessments analyzed in this study. The test-level analysis also showed that scores from assessments administered on tablets and non-tablet devices were similarly reliable in general. Validity coefficients were also generally similar for tablet and non-tablet testers. The student-level analysis did not reveal notable differences in the distributions of person-fit statistics for students testing on tablets and non-tablet devices.

Finally, this study addressed the question of whether students' levels of overall test performance were similar when comparing tests administered on tablets and non-tablet devices. As would be expected if testing on tablets was comparable to testing on non-tablet devices, the raw and scale score distributions of the matched tablet and non-tablet testers were generally similar. An analysis of equivalent raw scores revealed a similar pattern. That is, IRT true-score equating indicated that students performing at a certain level on non-tablet devices would be expected to perform similarly had they tested on tablets. The differences in expected scores was one point or less in most cases, and the direction of those differences were not consistent with the notion that testing on tablets is more difficult than testing on non-tablet devices.

4.2 Possible Explanations for Observed Device Effects

Although results were generally consistent with comparability between testing on tablets and non-tablet devices, there was some evidence of device effects. The strongest evidence of

device effects was observed for the Geometry assessment, where 11 tasks were flagged for differences in IRT difficulty. The cumulative impact of these device effects was apparent in the score distributions, which showed that tablet testers scored an average of 0.13 standard deviations lower than non-tablet testers on PBA, and in the IRT true-score equating results, which revealed slightly lower expected performance for tablet testers. An examination of matching results for each state revealed a systematic pattern of lower performance on the Geometry assessment for the matched samples tablet testers. Thus, results cannot be attributed to an aberrant state.

There did not appear to be a particular type of task that was more likely to be flagged. Of the nine Geometry tasks that were more difficult for tablet testers, seven were one-point tasks, three allowed calculators, and there were a variety task types (four selected response, two fill-in-the-blank, one constructed response with an equation editor, one drag-and-drop, and one inline choice). Thus, it is difficult to draw any conclusions about what types of tasks are likely to exhibit device effects. Regardless, such effects could reflect a lack of familiarity and comfort with entering responses to mathematics tasks on tablets. For example, there are differences in the mechanics for inputting answers to drag-and-drop tasks on tablets and non-tablet devices (finger and touchscreen vs. mouse and cursor). In addition, limited screen real estate and the subsequent need to scroll more frequently could pose a challenge on certain mathematics tasks. The effect of some of these issues might be expected to decrease as familiarity with testing on tablets increases.

Any apparent device effects could also reflect unsuccessful matching between the tablet and non-tablet testers. That is, any observed difference in performance between the TC and NTC conditions could reflect device effects, differences in ability, or a combination of the two. As indicated in Appendix A (and summarized in Tables 12 and 13), there were small differences between the tablet and non-tablet matched groups in terms of prior achievement (based only on matched pairs of students who both had prior achievement scores). For example, in the grade 7 ELA/L matched samples, available data suggested higher average ability among the tablet testers. Consistent with that difference in ability, later analyses (e.g., score distributions and IRT true-score equating) indicated that tablet testers performed slightly better than non-tablet testers. In such cases, some component of apparent "device effects" is likely due to differences in ability rather than differences in the difficulty of testing on tablets versus non-tablet devices.

4.3 Implications

The major implication of this study is that there does not appear to be large or consistent differences in assessment results from tablet and non-tablet administrations. Moreover, device comparability is supported by device comparability research conducted outside of PARCC (e.g., Davis, Orr, Kong, Lin, 2014; Olsen, 2014; Davis, Kong, McBride, 2015). Comparability across devices is further supported by policies in other large-scale assessment programs (e.g., Smarter Balanced and other statewide assessments). When this study detected possible device effects, they were most frequently apparent for high school mathematics assessments. This pattern in results has implications for task development and user-interface design. Specifically, it is

important to consider that the interaction of complex (high school) tasks with complex and possibly unfamiliar interfaces may have an influence on student performance. Additional focus groups and cognitive labs with high school students would help identify construct-irrelevant aspects of the testing environment that potentially interfere with students' abilities to demonstrate the full extent of their proficiencies in mathematics.

4.4 Limitations

A limitation of this study is that it included only one form of eight of the 21 PARCC mathematics and ELA/L assessments. The conclusions drawn from this study may not generalize to other PARCC assessments. A limitation that may have affected results is that some PARCC assessment tasks were quite difficult. When average task performance is poor on both tablets and non-tablet devices, possible device effects would be hidden because most students could not answer correctly regardless of the administration mode. Moreover, the narrow range of proficiency represented in the sample (i.e., restriction of range) could have depressed reliability coefficients and correlations (low average raw scores are apparent in the score distribution tables in Tables 6 and 7).

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Davis, L.L., Orr, A., Kong, X., & Lin, C. (2014, April). *Assessing student writing on tablets.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Philadelphia, PA. Retrieved November 18, 2015 from http://researchnetwork.pearson.com/wpcontent/uploads/NCME_Assessing-Student-Writing-on-Tablets_040414.pdf

Davis, L.L., Kong, X., & McBride, M. (2015, April). *Device comparability of tablets and computers for assessment purposes.* Paper to be presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.

Dunlap, W.P., Cortina, J.M., Vaslow, J.B., & Burke, M.J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological methods, 1*(2), 170-177. doi: 10.1037/1082-989X.1.2.170

Feldt, L.S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika, 34*(3), 363-373.

Feldt, L.S., & Brennan, R.L. (1993). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 184-190). New York, NY: Macmillan.

Fisher, (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3-32.

Ho, D., Imai, K., King, G., & Stuart, E.A. (2011). Matchit: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, 42*(8), 1-28. doi: 10.18637/jss.v042.i08

Iacus, S.M., King, G., & Porro, G. (2009). cem: Software for coarsened exact matching. *Journal of Statistical Software, 30*(9), 1-27. doi: 10.18637/jss.v030.i09

Iacus, S.M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis, 20*(1), 1-24. doi: 10.1093/pan/mpr013

Keng, L., Davis, L., McBride, Y., Glaze, R., & Steedle, J. (2015). *Spring 2014 digital devices comparability research study*. Washington, DC: PARCC, Inc.

Kolen, M. J. & Brennan, R. L. (2004). *Test equating, scaling, and linking*. New York, NY: Springer.

Linacre, J. M. (2015). *Winsteps® Rasch measurement computer program.* Beaverton, Oregon: Winsteps.com

Liu, H.-Y., & Weng, L.-J. (2009). An effect size index for comparing two independent alpha coefficients. *British Journal of Mathematical and Statistical Psychology, 62*(2), 385-400.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149-174.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics*. New York, NY: Academic Press.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*(2), 153-157. doi: 10.1007/BF02295996

Miller, G. E., Rotou, O., & Twing, J. S. (2004). Evaluation of the .3 logits screening criterion in common item equating. *Journal of Applied Measurement*, *5*(2), 172-177.

Olsen, J.B. (2014, April). *Score comparability for web and iPad delivered adaptive tests.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Philadelphia, PA.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.

Appendix A: Coarsened Exact Matching Balance Tables

Table A.1

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 5 Mathematics PBA

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	85932	3253		3242	3242		308	308	
American Indian/Alaska Native	1.2	0.4	-0.8	0.4	0.4	0.0	0.0	0.0	0.0
Asian	5.4	4.2	-1.2	4.2	4.2	0.0	3.3	3.3	0.0
Black/African American	16.5	14.1	-2.4	14.1	14.1	0.0	24.4	24.4	0.0
Hispanic/Latino Ethnicity	27.6	25.9	-1.7	25.9	25.9	0.0	15.3	15.3	0.0
Hawaiian/Pacific Islander	0.3	0.0	-0.3	0.0	0.0	0.0	0.0	0.0	0.0
White	46.3	52.6	6.4	52.7	52.7	0.0	52.0	52.0	0.0
Two or More Races	1.8	1.8	0.1	1.9	1.9	0.0	5.2	5.2	0.0
Ethnicity Not Provided	0.9	0.8	0.0	0.8	0.8	0.0	0.0	0.0	0.0
Female	47.9	49.0	1.1	49.0	49.0	0.0	54.2	54.2	0.0
Economic Disadvantage	49.7	49.2	-0.5	49.2	49.2	0.0	56.2	56.2	0.0
English Learner	10.3	10.9	0.6	10.6	10.6	0.0	5.8	5.8	0.0
Student with Disability	14.2	11.5	-2.7	11.3	11.3	0.0	13.3	13.3	0.0
Prior Achievement*	-0.12	-0.27	-0.15	-0.29	-0.26	0.02	-0.29	-0.28	0.01
Grade 5	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.2

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 5 Mathematics EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	43752	1546		1544	1544		42	42	
American Indian/Alaska Native	0.9	0.3	-0.6	0.3	0.3	0.0	0.0	0.0	0.0
Asian	6.3	7.2	0.9	7.2	7.2	0.0	2.4	2.4	0.0
Black/African American	14.4	10.6	-3.8	10.6	10.6	0.0	59.5	59.5	0.0
Hispanic/Latino Ethnicity	23.3	19.2	-4.2	19.2	19.2	0.0	19.1	19.1	0.0
Hawaiian/Pacific Islander	0.2	0.1	-0.1	0.1	0.1	0.0	0.0	0.0	0.0
White	52.2	60.2	8.0	60.1	60.1	0.0	19.1	19.1	0.0
Two or More Races	1.6	1.4	-0.1	1.4	1.4	0.0	0.0	0.0	0.0
Ethnicity Not Provided	1.2	1.1	-0.1	1.1	1.1	0.0	0.0	0.0	0.0
Female	49.7	50.5	0.8	50.5	50.5	0.0	40.5	40.5	0.0
Economic Disadvantage	44.0	36.7	-7.3	36.7	36.7	0.0	69.1	69.1	0.0
English Learner	4.1	3.4	-0.7	3.3	3.3	0.0	0.0	0.0	0.0
Student with Disability	4.5	4.1	-0.4	4.1	4.1	0.0	4.8	4.8	0.0
Prior Achievement*	0.10	-0.01	-0.11	-0.12	-0.01	0.11	-0.02	-0.01	0.01
Grade 3	0.0	0.1	0.1	0.0	0.0	0.0			
Grade 5	100.0	99.9	0.0	100.0	100.0	0.0	100.0	100.0	0.0

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.3

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 5 Mathematics PBA+EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	7272	229		226	226				
American Indian/Alaska Native	0.9	0.0	-0.9	0.0	0.0	0.0			
Asian	6.0	7.9	1.8	7.1	7.1	0.0			
Black/African American	13.9	9.2	-4.7	9.3	9.3	0.0			
Hispanic/Latino Ethnicity	23.0	19.7	-3.4	19.5	19.5	0.0			
Hawaiian/Pacific Islander	0.2	0.0	-0.2	0.0	0.0	0.0			
White	53.0	61.6	8.6	62.4	62.4	0.0			
Two or More Races	1.6	1.3	-0.3	1.3	1.3	0.0			
Ethnicity Not Provided	1.4	0.4	-0.9	0.4	0.4	0.0			
Female	49.8	48.9	-0.9	48.7	48.7	0.0			
Economic Disadvantage	43.6	31.9	-11.7	31.4	31.4	0.0			
English Learner	3.5	2.2	-1.3	1.3	1.3	0.0			
Student with Disability	4.4	3.1	-1.3	3.1	3.1	0.0			
Prior Achievement*	0.01	-0.01	-0.02	-0.21	-0.82	-0.60			
Grade 5	100.0	100.0	0.0	100.0	100.0	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.4

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 7 Mathematics PBA

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	90515	3403		3382	3382		743	743	
American Indian/Alaska Native	1.1	0.5	-0.6	0.4	0.4	0.0	0.0	0.0	0.0
Asian	4.6	6.5	1.9	6.5	6.5	0.0	5.8	5.8	0.0
Black/African American	18.7	12.9	-5.8	13.0	13.0	0.0	12.9	12.9	0.0
Hispanic/Latino Ethnicity	29.0	23.3	-5.7	23.2	23.2	0.0	12.9	12.9	0.0
Hawaiian/Pacific Islander	0.3	0.1	-0.2	0.0	0.0	0.0	0.0	0.0	0.0
White	44.0	54.5	10.5	54.5	54.5	0.0	66.1	66.1	0.0
Two or More Races	1.3	0.8	-0.5	0.8	0.8	0.0	0.9	0.9	0.0
Ethnicity Not Provided	1.0	1.4	0.4	1.4	1.4	0.0	1.4	1.4	0.0
Female	47.8	47.6	-0.2	47.6	47.6	0.0	46.7	46.7	0.0
Economic Disadvantage	51.0	38.8	-12.2	38.7	38.7	0.0	19.5	19.5	0.0
English Learner	9.1	7.3	-1.8	6.9	6.9	0.0	2.4	2.4	0.0
Student with Disability	12.3	10.4	-2.0	10.1	10.1	0.0	17.0	17.0	0.0
Prior Achievement*	-0.15	-0.01	0.14	0.03	0	-0.03	0.01	0	-0.01
Grade 6	0.1	0.1	0.0	0.1	0.1	0.0			
Grade 7	99.9	99.8	-0.1	99.9	99.9	0.0	100.0	100.0	0.0
Grade 8	0.0	0.1	0.1	0.0	0.0	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.5

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 7 Mathematics EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	89779	3739		3721	3721		714	714	
American Indian/Alaska Native	1.1	0.9	-0.2	0.7	0.7	0.0	0.0	0.0	0.0
Asian	4.5	6.1	1.6	6.1	6.1	0.0	4.8	4.8	0.0
Black/African American	18.8	12.7	-6.0	12.8	12.8	0.0	11.8	11.8	0.0
Hispanic/Latino Ethnicity	29.3	26.4	-2.9	26.4	26.4	0.0	11.5	11.5	0.0
Hawaiian/Pacific Islander	0.3	0.1	-0.3	0.0	0.0	0.0	0.0	0.0	0.0
White	43.8	51.4	7.7	51.6	51.6	0.0	69.2	69.2	0.0
Two or More Races	1.3	0.8	-0.4	0.8	0.8	0.0	1.8	1.8	0.0
Ethnicity Not Provided	1.0	1.6	0.6	1.6	1.6	0.0	1.0	1.0	0.0
Female	47.9	49.6	1.7	49.6	49.6	0.0	52.1	52.1	0.0
Economic Disadvantage	52.3	44.1	-8.2	44.2	44.2	0.0	26.6	26.6	0.0
English Learner	9.2	7.4	-1.8	7.2	7.2	0.0	2.8	2.8	0.0
Student with Disability	12.3	9.6	-2.7	9.4	9.4	0.0	14.4	14.4	0.0
Prior Achievement*	-0.15	0.02	0.17	0.05	0.03	-0.03	0.02	0.02	0.00
Grade 6	0.1	0.0	-0.1	0.0	0.0	0.0			
Grade 7	99.9	100.0	0.1	100.0	100.0	0.0	100.0	100.0	0.0

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.6

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 7 Mathematics PBA+EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	28219	889		879	879		116	116	
American Indian/Alaska Native	1.1	0.9	-0.2	0.8	0.8	0.0	0.0	0.0	0.0
Asian	4.2	4.8	0.7	4.6	4.6	0.0	2.6	2.6	0.0
Black/African American	20.1	16.3	-3.8	16.4	16.4	0.0	9.5	9.5	0.0
Hispanic/Latino Ethnicity	33.4	29.6	-3.9	29.9	29.9	0.0	9.5	9.5	0.0
Hawaiian/Pacific Islander	0.3	0.1	-0.2	0.0	0.0	0.0	0.0	0.0	0.0
White	38.8	46.6	7.8	46.6	46.6	0.0	75.9	75.9	0.0
Two or More Races	1.2	0.6	-0.7	0.6	0.6	0.0	0.9	0.9	0.0
Ethnicity Not Provided	0.9	1.1	0.2	1.1	1.1	0.0	1.7	1.7	0.0
Female	47.2	45.9	-1.3	46.1	46.1	0.0	50.0	50.0	0.0
Economic Disadvantage	56.6	47.6	-9.1	48.0	48.0	0.0	16.4	16.4	0.0
English Learner	13.2	12.2	-1.0	11.8	11.8	0.0	3.5	3.5	0.0
Student with Disability	14.8	11.3	-3.6	10.9	10.9	0.0	26.7	26.7	0.0
Prior Achievement*	-0.02	-0.03	-0.01	-0.13	-0.19	-0.06	-0.19	-0.2	-0.01
Grade 6	0.1	0.0	-0.1	0.0	0.0	0.0			
Grade 7	99.9	100.0	0.1	100.0	100.0	0.0	100.0	100.0	0.0

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.7

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Algebra 1 PBA

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	62039	1489		1445	1445		166	166	
American Indian/Alaska Native	1.0	0.7	-0.3	0.5	0.5	0.0	0.0	0.0	0.0
Asian	5.2	7.5	2.3	7.0	7.0	0.0	2.4	2.4	0.0
Black/African American	24.1	14.4	-9.7	14.5	14.5	0.0	19.9	19.9	0.0
Hispanic/Latino Ethnicity	28.0	30.6	2.6	31.0	31.0	0.0	31.9	31.9	0.0
Hawaiian/Pacific Islander	0.3	0.0	-0.3	0.0	0.0	0.0	0.0	0.0	0.0
White	39.4	44.9	5.5	45.4	45.4	0.0	44.0	44.0	0.0
Two or More Races	1.4	1.1	-0.2	1.1	1.1	0.0	1.8	1.8	0.0
Ethnicity Not Provided	0.6	0.7	0.2	0.6	0.6	0.0	0.0	0.0	0.0
Female	47.6	46.8	-0.8	47.1	47.1	0.0	50.0	50.0	0.0
Economic Disadvantage	44.3	43.7	-0.6	44.2	44.2	0.0	42.8	42.8	0.0
English Learner	9.8	10.1	0.4	9.0	9.0	0.0	8.4	8.4	0.0
Student with Disability	10.1	10.5	0.4	10.4	10.4	0.0	4.2	4.2	0.0
Prior Achievement*	0.05	0.05	0	0.17	0.12	-0.04	0.15	0.11	-0.04
Grade 6	0.1	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0
Grade 7	3.8	0.6	-3.2	0.6	0.6	0.0	0.0	0.0	0.0
Grade 8	18.4	22.4	4.0	22.2	22.2	0.0	49.4	49.4	0.0
Grade 9	68.2	67.5	-0.8	68.8	68.8	0.0	48.8	48.8	0.0
Grade 10	8.1	8.1	-0.1	7.6	7.6	0.0	1.8	1.8	0.0
Grade 11	1.2	1.3	0.0	0.8	0.8	0.0			
Grade 12	0.2	0.1	0.0	0.0	0.0	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.8

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Algebra 1 EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	55451	1409		1388	1388		204	204	
American Indian/Alaska Native	1.1	0.4	-0.8	0.4	0.4	0.0	1.0	1.0	0.0
Asian	5.8	9.5	3.7	9.3	9.3	0.0	2.5	2.5	0.0
Black/African American	20.9	12.8	-8.1	12.8	12.8	0.0	15.2	15.2	0.0
Hispanic/Latino Ethnicity	30.2	27.5	-2.7	27.5	27.5	0.0	22.1	22.1	0.0
Hawaiian/Pacific Islander	0.4	0.0	-0.4	0.0	0.0	0.0	0.0	0.0	0.0
White	39.5	47.9	8.4	48.1	48.1	0.0	58.3	58.3	0.0
Two or More Races	1.6	1.0	-0.6	0.9	0.9	0.0	1.0	1.0	0.0
Ethnicity Not Provided	0.6	1.0	0.4	1.0	1.0	0.0	0.0	0.0	0.0
Female	47.8	47.6	-0.3	47.5	47.5	0.0	51.5	51.5	0.0
Economic Disadvantage	50.1	42.1	-8.0	41.9	41.9	0.0	34.3	34.3	0.0
English Learner	10.9	10.4	-0.5	9.4	9.4	0.0	7.4	7.4	0.0
Student with Disability	11.1	10.6	-0.5	10.5	10.5	0.0	6.4	6.4	0.0
Prior Achievement*	0.06	0.14	0.08	0.35	0.21	-0.13	0.18	0.16	-0.02
Grade 6	0.1	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0
Grade 8	4.2	0.6	-3.7	0.6	0.6	0.0	0.0	0.0	0.0
Grade 7	21.0	23.6	2.7	23.9	23.9	0.0	52.9	52.9	0.0
Grade 9	66.7	67.6	0.9	68.4	68.4	0.0	43.6	43.6	0.0
Grade 10	6.7	6.3	-0.5	6.0	6.0	0.0	3.4	3.4	0.0
Grade 11	1.2	1.9	0.7	1.1	1.1	0.0			
Grade 12	0.1	0.1	-0.1	0.1	0.1	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.9

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Algebra 1 PBA+EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	14938	218		207	207		16	16	
American Indian/Alaska Native	0.3	0.5	0.2	0.0	0.0	0.0	0.0	0.0	0.0
Asian	6.0	8.7	2.7	7.3	7.3	0.0	0.0	0.0	0.0
Black/African American	24.8	17.4	-7.4	18.4	18.4	0.0	6.3	6.3	0.0
Hispanic/Latino Ethnicity	32.5	30.3	-2.2	30.4	30.4	0.0	37.5	37.5	0.0
Hawaiian/Pacific Islander	0.3	0.0	-0.3	0.0	0.0	0.0	0.0	0.0	0.0
White	33.8	40.8	7.1	41.6	41.6	0.0	56.3	56.3	0.0
Two or More Races	1.6	1.4	-0.2	1.5	1.5	0.0	0.0	0.0	0.0
Ethnicity Not Provided	0.8	0.9	0.1	1.0	1.0	0.0	0.0	0.0	0.0
Female	47.3	45.4	-1.9	45.4	45.4	0.0	81.3	81.3	0.0
Economic Disadvantage	54.9	45.0	-9.9	44.4	44.4	0.0	43.8	43.8	0.0
English Learner	14.7	12.4	-2.3	9.2	9.2	0.0	25.0	25.0	0.0
Student with Disability	10.6	14.7	4.1	15.0	15.0	0.0	6.3	6.3	0.0
Prior Achievement*	0.04	0.00	-0.04	0.09	0.05	-0.04	0.07	0.12	0.05
Grade 6	0.2	0.0	-0.2	0.0	0.0	0.0	0.0	0.0	0.0
Grade 7	5.2	1.4	-3.8	1.5	1.5	0.0	0.0	0.0	0.0
Grade 8	19.1	20.6	1.6	20.8	20.8	0.0	50.0	50.0	0.0
Grade 9	67.1	65.6	-1.5	67.2	67.2	0.0	50.0	50.0	0.0
Grade 10	7.1	10.6	3.5	9.7	9.7	0.0	0.0	0.0	0.0
Grade 11	1.4	1.8	0.5	1.0	1.0	0.0			
Grade 12	0.1	0.0	-0.1	0.0	0.0	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.10

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Geometry PBA

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	24825	697		688	688				
American Indian/Alaska Native	1.8	0.1	-1.6	0.2	0.2	0.0			
Asian	5.3	7.2	1.9	6.7	6.7	0.0			
Black/African American	14.4	15.5	1.1	15.6	15.6	0.0			
Hispanic/Latino Ethnicity	29.0	21.4	-7.7	21.7	21.7	0.0			
Hawaiian/Pacific Islander	0.9	0.1	-0.7	0.2	0.2	0.0			
White	47.4	53.8	6.4	54.4	54.4	0.0			
Two or More Races	1.1	1.9	0.8	1.5	1.5	0.0			
Ethnicity Not Provided	0.2	0.0	-0.2	0.0	0.0	0.0			
Female	48.0	49.1	1.0	48.8	48.8	0.0			
Economic Disadvantage	44.3	37.9	-6.4	37.7	37.7	0.0			
English Learner	9.6	5.0	-4.6	4.8	4.8	0.0			
Student with Disability	13.1	13.5	0.4	13.2	13.2	0.0			
Prior Achievement*	0.42	0.48	0.06	0.42	0.69	0.27			
Grade 8	3.0	0.7	-2.3	0.7	0.7	0.0			
Grade 9	24.0	27.6	3.6	27.6	27.6	0.0			
Grade 10	62.7	61.7	-1.0	62.2	62.2	0.0			
Grade 11	9.8	9.9	0.1	9.3	9.3	0.0			
Grade 12	0.4	0.1	-0.3	0.2	0.2	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.11

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Geometry EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	15870	470		467	467				
American Indian/Alaska Native	2.1	0.6	-1.5	0.2	0.2	0.0			
Asian	6.9	11.1	4.2	11.1	11.1	0.0			
Black/African American	11.4	12.1	0.7	12.0	12.0	0.0			
Hispanic/Latino Ethnicity	27.0	20.2	-6.7	20.3	20.3	0.0			
Hawaiian/Pacific Islander	0.8	0.0	-0.8	0.0	0.0	0.0			
White	50.5	54.7	4.2	55.0	55.0	0.0			
Two or More Races	1.3	1.3	-0.1	1.3	1.3	0.0			
Ethnicity Not Provided	0.1	0.0	-0.1	0.0	0.0	0.0			
Female	47.9	49.6	1.7	49.5	49.5	0.0			
Economic Disadvantage	39.5	31.3	-8.3	31.3	31.3	0.0			
English Learner	4.2	4.0	-0.1	3.9	3.9	0.0			
Student with Disability	8.7	7.9	-0.9	7.7	7.7	0.0			
Prior Achievement*									
Grade 7	0.0	0.0	0.0	0.0	0.0	0.0			
Grade 8	3.9	1.7	-2.2	1.7	1.7	0.0			
Grade 9	27.3	33.6	6.3	33.6	33.6	0.0			
Grade 10	59.3	54.9	-4.4	54.8	54.8	0.0			
Grade 11	9.2	9.8	0.6	9.9	9.9	0.0			
Grade 12	0.3	0.0	-0.3	0.0	0.0	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.12

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Geometry PBA+EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	1878	56		53	53				
American Indian/Alaska Native	2.1	0.0	-2.1	0.0	0.0	0.0			
Asian	5.8	12.5	6.7	11.3	11.3	0.0			
Black/African American	9.9	8.9	-1.0	9.4	9.4	0.0			
Hispanic/Latino Ethnicity	29.3	16.1	-13.2	15.1	15.1	0.0			
Hawaiian/Pacific Islander	1.0	0.0	-1.0	0.0	0.0	0.0			
White	50.5	62.5	12.0	64.2	64.2	0.0			
Two or More Races	1.3	0.0	-1.3	0.0	0.0	0.0			
Ethnicity Not Provided	0.1	0.0	-0.1	0.0	0.0	0.0			
Female	48.2	44.6	-3.6	45.3	45.3	0.0			
Economic Disadvantage	40.2	28.6	-11.6	24.5	24.5	0.0			
English Learner	4.5	5.4	0.9	1.9	1.9	0.0			
Student with Disability	9.3	8.9	-0.3	9.4	9.4	0.0			
Prior Achievement*									
Grade 5	0.1	0.0	-0.1	0.0	0.0	0.0			
Grade 8	4.7	1.8	-3.0	1.9	1.9	0.0			
Grade 9	26.9	26.8	-0.2	28.3	28.3	0.0			
Grade 10	58.2	57.1	-1.0	58.5	58.5	0.0			
Grade 11	9.9	14.3	4.4	11.3	11.3	0.0			
Grade 12	0.2	0.0	-0.2	0.0	0.0	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.13

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Algebra 2 PBA

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	33941	782		772	772				
American Indian/Alaska Native	1.2	0.4	-0.8	0.3	0.3	0.0			
Asian	7.9	10.5	2.6	10.5	10.5	0.0			
Black/African American	15.9	11.9	-4.1	11.8	11.8	0.0			
Hispanic/Latino Ethnicity	22.5	20.1	-2.4	20.3	20.3	0.0			
Hawaiian/Pacific Islander	0.5	0.3	-0.3	0.1	0.1	0.0			
White	49.7	55.0	5.3	55.2	55.2	0.0			
Two or More Races	1.7	1.4	-0.3	1.4	1.4	0.0			
Ethnicity Not Provided	0.6	0.5	-0.1	0.4	0.4	0.0			
Female	49.7	50.5	0.8	50.4	50.4	0.0			
Economic Disadvantage	35.1	27.6	-7.4	26.9	26.9	0.0			
English Learner	4.2	1.8	-2.5	1.6	1.6	0.0			
Student with Disability	6.8	7.3	0.5	6.9	6.9	0.0			
Prior Achievement*	0.72	0.46	-0.26	0.13	0.45	0.32			
Grade 8	1.0	0.3	-0.7	0.3	0.3	0.0			
Grade 9	10.0	5.1	-4.8	5.2	5.2	0.0			
Grade 10	36.3	39.9	3.7	40.3	40.3	0.0			
Grade 11	49.0	54.0	5.0	53.9	53.9	0.0			
Grade 12	3.8	0.8	-3.1	0.4	0.4	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.14

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Algebra 2 EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	32858	695		691	691				
American Indian/Alaska Native	1.3	0.4	-0.9	0.4	0.4	0.0			
Asian	7.9	8.5	0.6	8.5	8.5	0.0			
Black/African American	16.5	11.5	-5.0	11.6	11.6	0.0			
Hispanic/Latino Ethnicity	22.4	18.4	-4.0	18.5	18.5	0.0			
Hawaiian/Pacific Islander	0.5	0.0	-0.5	0.0	0.0	0.0			
White	49.2	60.0	10.8	59.9	59.9	0.0			
Two or More Races	1.6	0.9	-0.8	0.7	0.7	0.0			
Ethnicity Not Provided	0.5	0.3	-0.2	0.3	0.3	0.0			
Female	49.9	49.1	-0.8	49.1	49.1	0.0			
Economic Disadvantage	34.9	26.2	-8.7	26.2	26.2	0.0			
English Learner	4.3	1.3	-3.0	1.2	1.2	0.0			
Student with Disability	6.9	7.9	1.0	7.7	7.7	0.0			
Prior Achievement*	0.76	0.31	-0.45	0.63	0.31	-0.32			
Grade 8	0.9	0.0	-0.9	0.0	0.0	0.0			
Grade 9	10.2	5.8	-4.4	5.6	5.6	0.0			
Grade 10	36.1	39.9	3.8	39.9	39.9	0.0			
Grade 11	48.6	53.4	4.7	53.4	53.4	0.0			
Grade 12	4.2	1.0	-3.2	1.0	1.0	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.15

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Algebra 2 PBA+EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	8165	156		151	151				
American Indian/Alaska Native	1.0	0.0	-1.0	0.0	0.0	0.0			
Asian	9.1	8.3	-0.8	8.0	8.0	0.0			
Black/African American	16.2	14.1	-2.1	13.9	13.9	0.0			
Hispanic/Latino Ethnicity	23.8	15.4	-8.4	15.2	15.2	0.0			
Hawaiian/Pacific Islander	0.5	0.0	-0.5	0.0	0.0	0.0			
White	47.2	60.9	13.8	62.3	62.3	0.0			
Two or More Races	1.8	0.6	-1.2	0.0	0.0	0.0			
Ethnicity Not Provided	0.4	0.6	0.3	0.7	0.7	0.0			
Female	49.1	53.9	4.8	53.6	53.6	0.0			
Economic Disadvantage	35.0	21.2	-13.8	21.2	21.2	0.0			
English Learner	6.8	2.6	-4.2	2.0	2.0	0.0			
Student with Disability	7.9	9.6	1.7	9.9	9.9	0.0			
Prior Achievement*	0.21	0.01	-0.2	0.39	0.41	0.01			
Grade 8	1.3	0.0	-1.3	0.0	0.0	0.0			
Grade 9	12.2	3.2	-9.0	2.7	2.7	0.0			
Grade 10	36.1	37.2	1.1	37.1	37.1	0.0			
Grade 11	46.3	59.0	12.7	59.6	59.6	0.0			
Grade 12	4.1	0.6	-3.5	0.7	0.7	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.16

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 3 ELA/L PBA

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	57221	2231		2223	2223				
American Indian/Alaska Native	1.0	0.5	-0.5	0.5	0.5	0.0			
Asian	5.8	6.8	1.0	6.8	6.8	0.0			
Black/African American	14.3	9.2	-5.0	9.2	9.2	0.0			
Hispanic/Latino Ethnicity	25.0	19.8	-5.2	19.8	19.8	0.0			
Hawaiian/Pacific Islander	0.2	0.2	-0.1	0.2	0.2	0.0			
White	50.6	61.1	10.4	61.2	61.2	0.0			
Two or More Races	1.9	1.2	-0.7	1.1	1.1	0.0			
Ethnicity Not Provided	1.2	1.3	0.1	1.3	1.3	0.0			
Female	48.3	49.1	0.8	49.1	49.1	0.0			
Economic Disadvantage	44.8	36.0	-8.8	36.0	36.0	0.0			
English Learner	12.7	11.2	-1.5	11.1	11.1	0.0			
Student with Disability	11.5	9.4	-2.1	9.0	9.0	0.0			
Prior Achievement*									
Grade 3	100.0	100.0	0.0	100.0	100.0	0.0			

* There were no prior achievement data for grade 3 ELA/L.

Table A.17

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 3 ELA/L EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	66865	2577		2572	2572				
American Indian/Alaska Native	1.1	0.2	-0.9	0.2	0.2	0.0			
Asian	5.8	6.1	0.3	6.1	6.1	0.0			
Black/African American	14.1	9.4	-4.7	9.3	9.3	0.0			
Hispanic/Latino Ethnicity	25.4	20.4	-5.0	20.4	20.4	0.0			
Hawaiian/Pacific Islander	0.3	0.2	-0.1	0.2	0.2	0.0			
White	50.4	61.2	10.8	61.2	61.2	0.0			
Two or More Races	1.9	1.0	-1.0	0.9	0.9	0.0			
Ethnicity Not Provided	1.1	1.6	0.6	1.6	1.6	0.0			
Female	48.5	46.6	-1.8	46.7	46.7	0.0			
Economic Disadvantage	47.0	38.6	-8.4	38.5	38.5	0.0			
English Learner	12.9	11.5	-1.5	11.4	11.4	0.0			
Student with Disability	11.1	10.5	-0.6	10.4	10.4	0.0			
Prior Achievement*									
Grade 3	100.0	100.0	0.0	100.0	100.0	0.0			

* There were no prior achievement data for grade 3 ELA/L.

Table A.18

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 3 ELA/L PBA+EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	11513	406		399	399				
American Indian/Alaska Native	1.0	0.7	-0.2	0.8	0.8	0.0			
Asian	5.8	8.4	2.5	7.5	7.5	0.0			
Black/African American	14.2	8.9	-5.4	8.8	8.8	0.0			
Hispanic/Latino Ethnicity	25.1	20.4	-4.7	20.6	20.6	0.0			
Hawaiian/Pacific Islander	0.2	0.0	-0.2	0.0	0.0	0.0			
White	50.8	59.1	8.3	60.2	60.2	0.0			
Two or More Races	1.9	1.5	-0.4	1.3	1.3	0.0			
Ethnicity Not Provided	0.9	1.0	0.1	1.0	1.0	0.0			
Female	47.0	49.0	2.1	49.6	49.6	0.0			
Economic Disadvantage	47.9	37.2	-10.7	37.6	37.6	0.0			
English Learner	12.5	12.1	-0.4	11.8	11.8	0.0			
Student with Disability	15.4	13.8	-1.6	13.3	13.3	0.0			
Prior Achievement*									
Grade 3	100.0	100.0	0.0	100.0	100.0	0.0			

* There were no prior achievement data for grade 3 ELA/L.

Table A.19

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 7 ELA/L PBA

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	65377	2722		2714	2714		622	622	
American Indian/Alaska Native	1.1	0.3	-0.7	0.2	0.2	0.0	0.0	0.0	0.0
Asian	5.5	7.6	2.1	7.6	7.6	0.0	8.4	8.4	0.0
Black/African American	16.7	11.0	-5.7	10.9	10.9	0.0	16.1	16.1	0.0
Hispanic/Latino Ethnicity	25.5	20.5	-5.0	20.5	20.5	0.0	14.2	14.2	0.0
Hawaiian/Pacific Islander	0.3	0.1	-0.2	0.1	0.1	0.0	0.0	0.0	0.0
White	48.6	58.0	9.4	58.1	58.1	0.0	59.5	59.5	0.0
Two or More Races	1.4	1.0	-0.4	1.0	1.0	0.0	0.5	0.5	0.0
Ethnicity Not Provided	1.0	1.5	0.5	1.5	1.5	0.0	1.5	1.5	0.0
Female	48.1	49.3	1.2	49.3	49.3	0.0	49.8	49.8	0.0
Economic Disadvantage	45.8	33.0	-12.8	33.0	33.0	0.0	23.6	23.6	0.0
English Learner	6.0	4.6	-1.4	4.5	4.5	0.0	2.6	2.6	0.0
Student with Disability	10.8	8.7	-2.1	8.6	8.6	0.0	11.3	11.3	0.0
Prior Achievement*	-0.06	0.06	0.12	-0.03	0.05	0.08	0.09	0.08	-0.01
Grade 7	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.20

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 7 ELA/L EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	69840	3039		3031	3031		635	635	
American Indian/Alaska Native	1.0	0.5	-0.5	0.4	0.4	0.0	0.0	0.0	0.0
Asian	5.2	7.3	2.1	7.4	7.4	0.0	7.1	7.1	0.0
Black/African American	17.5	10.2	-7.3	10.2	10.2	0.0	12.9	12.9	0.0
Hispanic/Latino Ethnicity	25.5	21.4	-4.2	21.4	21.4	0.0	13.2	13.2	0.0
Hawaiian/Pacific Islander	0.3	0.1	-0.2	0.1	0.1	0.0	0.0	0.0	0.0
White	48.0	57.8	9.8	57.8	57.8	0.0	63.8	63.8	0.0
Two or More Races	1.4	0.8	-0.6	0.8	0.8	0.0	1.4	1.4	0.0
Ethnicity Not Provided	1.0	1.9	0.9	1.9	1.9	0.0	1.6	1.6	0.0
Female	48.5	47.5	-1.0	47.5	47.5	0.0	48.0	48.0	0.0
Economic Disadvantage	47.9	37.8	-10.1	37.7	37.7	0.0	29.0	29.0	0.0
English Learner	5.9	5.0	-0.9	4.9	4.9	0.0	4.4	4.4	0.0
Student with Disability	10.9	9.5	-1.4	9.3	9.3	0.0	14.7	14.7	0.0
Prior Achievement*	-0.07	0.11	0.18	0.01	0.14	0.14	0.11	0.12	0.01
Grade 7	100.0	100.0	0.0	100.0	100.0		100.0	100.0	0.0

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.21

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 7 ELA/L PBA+EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	12164	427		423	423		80	80	
American Indian/Alaska Native	1.1	0.0	-1.1	0.0	0.0	0.0	0.0	0.0	0.0
Asian	5.7	5.6	0.0	5.2	5.2	0.0	3.8	3.8	0.0
Black/African American	16.8	9.4	-7.4	9.0	9.0	0.0	13.8	13.8	0.0
Hispanic/Latino Ethnicity	26.1	21.3	-4.8	21.5	21.5	0.0	12.5	12.5	0.0
Hawaiian/Pacific Islander	0.3	0.0	-0.3	0.0	0.0	0.0	0.0	0.0	0.0
White	47.7	60.9	13.2	61.5	61.5	0.0	68.8	68.8	0.0
Two or More Races	1.5	0.9	-0.5	1.0	1.0	0.0	1.3	1.3	0.0
Ethnicity Not Provided	0.9	1.9	1.0	1.9	1.9	0.0	0.0	0.0	0.0
Female	47.2	48.7	1.5	48.9	48.9	0.0	48.8	48.8	0.0
Economic Disadvantage	47.8	33.0	-14.8	33.1	33.1	0.0	16.3	16.3	0.0
English Learner	6.3	4.5	-1.9	4.0	4.0	0.0	2.5	2.5	0.0
Student with Disability	14.9	10.3	-4.6	9.7	9.7	0.0	16.3	16.3	0.0
Prior Achievement*	-0.02	-0.01	0.01	-0.11	-0.06	0.05	0.02	0.04	0.02
Grade 7	100.0	100.0	0.0	100.0	100.0	0.0	100.0	100.0	0.0

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.22

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 9 ELA/L PBA

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	29404	1319		1314	1314				
American Indian/Alaska Native	1.4	0.3	-1.1	0.3	0.3	0.0			
Asian	5.3	11.7	6.4	11.6	11.6	0.0			
Black/African American	13.9	9.0	-4.9	9.0	9.0	0.0			
Hispanic/Latino Ethnicity	29.4	22.4	-7.0	22.5	22.5	0.0			
Hawaiian/Pacific Islander	0.6	0.1	-0.5	0.0	0.0	0.0			
White	47.4	54.5	7.1	54.6	54.6	0.0			
Two or More Races	1.1	0.9	-0.2	0.9	0.9	0.0			
Ethnicity Not Provided	0.9	1.1	0.3	1.1	1.1	0.0			
Female	48.8	48.8	-0.1	48.7	48.7	0.0			
Economic Disadvantage	43.8	30.9	-12.9	30.9	30.9	0.0			
English Learner	6.7	3.2	-3.5	3.0	3.0	0.0			
Student with Disability	7.6	5.6	-2.0	5.5	5.5	0.0			
Prior Achievement*	-0.05	0.07	0.12	0.03	0.16	0.13			
Grade 8	0.1	0.0	-0.1	0.0	0.0	0.0			
Grade 9	98.9	99.5	0.6	99.6	99.6	0.0			
Grade 10	0.8	0.4	-0.4	0.4	0.4	0.0			
Grade 11	0.2	0.1	-0.1	0.0	0.0	0.0			
Grade 12	0.1	0.0	-0.1	0.0	0.0	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.23

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 9 ELA/L EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	33864	1501		1492	1492				
American Indian/Alaska Native	1.5	0.5	-1.1	0.3	0.3	0.0			
Asian	5.4	11.1	5.7	11.0	11.0	0.0			
Black/African American	14.4	8.5	-5.9	8.5	8.5	0.0			
Hispanic/Latino Ethnicity	29.6	21.8	-7.9	21.9	21.9	0.0			
Hawaiian/Pacific Islander	0.6	0.1	-0.4	0.1	0.1	0.0			
White	46.6	55.3	8.7	55.5	55.5	0.0			
Two or More Races	1.1	0.9	-0.2	0.9	0.9	0.0			
Ethnicity Not Provided	0.8	1.8	1.0	1.8	1.8	0.0			
Female	48.4	48.4	-0.1	48.6	48.6	0.0			
Economic Disadvantage	46.2	34.2	-12.1	34.2	34.2	0.0			
English Learner	7.2	3.9	-3.2	3.8	3.8	0.0			
Student with Disability	10.7	6.9	-3.8	6.7	6.7	0.0			
Prior Achievement*	-0.11	0.05	0.16	0.05	-0.02	-0.06			
Grade 8	0.1	0.0	-0.1	0.0	0.0	0.0			
Grade 9	98.8	99.0	0.3	99.3	99.3	0.0			
Grade 10	0.8	0.8	0.0	0.6	0.6	0.0			
Grade 11	0.2	0.1	-0.1	0.1	0.1	0.0			
Grade 12	0.1	0.1	0.0	0.0	0.0	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Table A.24

Percentages of Tablet Condition (TC) and Non-Tablet Condition (NTC) Students in Unmatched and Matched Samples for Grade 9 ELA/L PBA+EOY

	Unmatched			CEM			CEM with Prior Ach.		
	NTC	TC	Diff.	NTC	TC	Diff.	NTC	TC	Diff.
Sample Size	3466	176		172	172				
American Indian/Alaska Native	0.5	0.6	0.1	0.6	0.6	0.0			
Asian	5.5	12.5	7.0	12.2	12.2	0.0			
Black/African American	14.6	8.5	-6.1	8.7	8.7	0.0			
Hispanic/Latino Ethnicity	26.4	22.2	-4.2	21.5	21.5	0.0			
Hawaiian/Pacific Islander	0.6	0.0	-0.6	0.0	0.0	0.0			
White	50.7	53.4	2.8	54.1	54.1	0.0			
Two or More Races	1.1	0.6	-0.6	0.6	0.6	0.0			
Ethnicity Not Provided	0.6	2.3	1.7	2.3	2.3	0.0			
Female	50.8	45.5	-5.4	46.5	46.5	0.0			
Economic Disadvantage	43.3	32.4	-10.9	33.1	33.1	0.0			
English Learner	6.2	3.4	-2.8	2.3	2.3	0.0			
Student with Disability	7.8	6.3	-1.6	6.4	6.4	0.0			
Prior Achievement*	0.00	0.00	0.00	0.37	0.05	-0.33			
Grade 8	0.0	0.0	0.0	0.0	0.0	0.0			
Grade 9	98.8	99.4	0.6	100.0	100.0	0.0			
Grade 10	0.9	0.6	-0.4	0.0	0.0	0.0			
Grade 11	0.2	0.0	-0.2	0.0	0.0	0.0			

* Students were not matched on prior achievement unless otherwise indicated. The mean prior achievement scores shown here reflect only the students with prior achievement data.

Appendix B: Comparison of Task Means

Table B.1

Comparison of Task Means for Grade 5 Mathematics (Matching Only on Demographics)

Task Number	Component	p TC	p NTC	p Diff.	McNemar χ^2	*t*	Effect Size	Sig.	Device Effect
1	PBA	.48	.47	.01	0.82		0.02		
2	PBA	.08	.08	.00	0.57		0.02		
3	PBA	.32	.33	-.01		-0.64	-0.01		
4	PBA	.59	.59	.01	0.29		0.01		
5	PBA	.33	.32	.01		1.21	0.01		
6	PBA	.32	.33	-.01	0.42		-0.02		
7	PBA	.24	.25	-.01		-1.50	-0.01		
8	PBA	.09	.10	-.01		-2.02	-0.02	*	
9	PBA	.24	.28	-.03		-4.48	-0.03	*	
10	PBA	.84	.85	-.01	1.98		-0.03		
11	PBA	.82	.82	.01	0.31		0.01		
12	PBA	.52	.51	.01		1.36	0.02		
13	PBA	.27	.28	-.01		-1.20	-0.01		
14	PBA	.17	.17	.00		-0.69	0.00		
15	PBA	.03	.03	.00		0.16	0.00		
16	PBA	.11	.12	.00		-0.59	0.00		
17	EOY	.30	.31	-.01	0.10		-0.01		
18	EOY	.41	.39	.02	1.05		0.04		
19	EOY	.84	.84	.00	0.06		-0.01		
20	EOY	.28	.31	-.03		-2.02	-0.03	*	
21	EOY	.45	.46	-.01	0.19		-0.02		
22	EOY	.58	.60	-.02	0.94		-0.03		
23	EOY	.15	.10	.04		4.69	0.08	*	
24	EOY	.34	.31	.03	4.37		0.07	*	
25	EOY	.64	.60	.03	4.45		0.07	*	
26	EOY	.23	.19	.03	5.80		0.08	*	
27	EOY	.85	.83	.02	2.89		0.06		
28	EOY	.40	.38	.02		1.45	0.03		
29	EOY	.11	.10	.01	0.35		0.02		
30	EOY	.13	.12	.01	0.38		0.02		
31	EOY	.78	.77	.01	0.83		0.03		
32	EOY	.15	.16	-.01	1.24		-0.04		
33	EOY	.77	.77	.00		0.12	0.00		
34	EOY	.78	.78	-.01	0.13		-0.01		
35	EOY	.68	.67	.01	0.19		0.02		
36	EOY	.51	.49	.02	1.02		0.03		

37	EOY	.77	.78	-.01	0.24		-0.02	
38	EOY	.19	.16	.02	3.18		0.06	
39	EOY	.35	.32	.02		1.85	0.03	
40	EOY	.69	.66	.03	3.11		0.06	
41	EOY	.80	.79	.01	0.22		0.02	
42	EOY	.71	.66	.05	9.56		0.11	*
43	EOY	.48	.49	-.02		-1.03	-0.02	
44	EOY	.76	.77	-.01	0.40		-0.02	
45	EOY	.07	.09	-.01	2.18		-0.05	
46	EOY	.43	.43	.01	0.25		0.02	
47	EOY	.62	.61	.01	0.21		0.02	
48	EOY	.66	.65	.01		0.49	0.01	
49	EOY	.81	.83	-.02	1.30		-0.04	
50	EOY	.34	.30	.03		2.41	0.04	*
51	EOY	.15	.13	.02	1.66		0.04	
52	EOY	.31	.28	.03	4.04		0.07	*

Table B.2

Comparison of Task Means for Grade 7 Mathematics (Matching Only on Demographics)

Task Number	Component	p TC	p NTC	p Diff.	McNemar χ^2	*t*	Effect Size	Sig.	Device Effect
1	PBA	.27	.25	.02	3.34		0.04		
2	PBA	.59	.59	-.01	0.41		-0.02		
3	PBA	.24	.22	.02	3.96		0.05	*	
4	PBA	.25	.23	.02	3.50		0.04		
5	PBA	.63	.61	.01	1.44		0.03		
6	PBA	.66	.65	.00	0.12		0.01		
7	PBA	.46	.44	.02	2.49		0.04		
8	PBA	.05	.04	.01	2.07		0.03		
9	PBA	.51	.54	-.03		-3.02	-0.03	*	
10	PBA	.41	.43	-.02		-2.48	-0.03	*	
11	PBA	.20	.18	.02		4.13	0.02	*	
12	PBA	.08	.09	.00		-0.68	-0.01		
13	PBA	.20	.17	.02		2.74	0.02	*	
14	PBA	.37	.38	-.01		-2.63	-0.02	*	
15	PBA	.20	.19	.01		1.82	0.01		
16	PBA	.08	.09	-.01		-2.19	-0.01	*	
17	PBA	.04	.04	.00		-0.25	0.00		
18	EOY	.81	.80	.01	1.78		0.03		
19	EOY	.61	.62	-.01	0.68		-0.02		
20	EOY	.02	.02	.01	2.64		0.04		

21	EOY	.05	.05	.00	0.24		0.01	
22	EOY	.20	.21	-.01	0.59		-0.02	
23	EOY	.05	.05	.00	0.00		0.00	
24	EOY	.09	.07	.01	4.17		0.05	*
25	EOY	.13	.12	.01	1.85		0.03	
26	EOY	.61	.62	-.01	1.04		-0.02	
27	EOY	.82	.83	-.01	0.50		-0.02	
28	EOY	.09	.08	.00	0.57		0.02	
29	EOY	.62	.65	-.03	6.12		-0.06	*
30	EOY	.74	.73	.01	1.37		0.03	
31	EOY	.09	.08	.00	0.51		0.02	
32	EOY	.53	.54	-.02	1.91		-0.03	
33	EOY	.10	.09	.01	2.54		0.04	
34	EOY	.75	.74	.01	0.93		0.02	
35	EOY	.08	.07	.01		2.57	0.03	*
36	EOY	.04	.04	.00	0.06		-0.01	
37	EOY	.57	.59	-.02	3.83		-0.04	
38	EOY	.26	.26	.00		-0.31	0.00	
39	EOY	.36	.35	.01		0.89	0.01	
40	EOY	.16	.15	.01		0.96	0.01	
41	EOY	.06	.05	.01		1.84	0.02	
42	EOY	.31	.33	-.01		-1.93	-0.02	
43	EOY	.60	.61	-.01	0.59		-0.02	
44	EOY	.58	.63	-.05	23.27		-0.11	*
45	EOY	.23	.22	.01		1.92	0.02	
46	EOY	.52	.53	-.01	1.31		-0.02	
47	EOY	.08	.08	.00	0.05		0.00	
48	EOY	.13	.13	.00		-0.80	0.00	
49	EOY	.11	.09	.01		2.16	0.02	*
50	EOY	.17	.16	.00	0.09		0.01	

Table B.3

Comparison of Task Means for Algebra 1 (Matching Only on Demographics)

Task Number	Component	p TC	p NTC	p Diff.	McNemar χ^2	*t*	Effect Size	Sig.	Device Effect
1	PBA	.47	.46	.01	0.28		0.02		
2	PBA	.04	.04	.00	0.16		-0.01		
3	PBA	.57	.53	.04	4.34		0.07	*	
4	PBA	.37	.37	.00	0.04		0.01		
5	PBA	.36	.35	.01	0.61		0.03		
6	PBA	.57	.57	-.01	0.12		-0.01		

7	PBA	.05	.05	.00	0.26		-0.02	
8	PBA	.05	.04	.01	0.52		0.03	
9	PBA	.69	.68	.01	0.26		0.02	
10	PBA	.07	.07	.00	0.20		0.02	
11	PBA	.03	.04	.00		-1.08	-0.01	
12	PBA	.01	.01	.00		-0.16	0.00	
13	PBA	.09	.09	.00		0.72	0.01	
14	PBA	.02	.02	.00		-1.06	-0.01	
15	PBA	.10	.10	.00		-0.31	0.00	
16	PBA	.11	.09	.01		1.88	0.02	
17	PBA	.06	.08	-.01		-2.53	-0.01	*
18	PBA	.03	.02	.00		1.12	0.01	
19	EOY	.48	.48	.00	0.00		0.00	
20	EOY	.06	.06	.00	0.06		0.01	
21	EOY	.08	.10	-.02		-2.66	-0.05	*
22	EOY	.28	.27	.01		1.15	0.02	
23	EOY	.38	.38	.00	0.02		0.00	
24	EOY	.32	.30	.02	2.00		0.05	
25	EOY	.01	.02	-.01	2.81		-0.06	
26	EOY	.06	.06	-.01	0.32		-0.02	
27	EOY	.65	.67	-.02	1.82		-0.05	
28	EOY	.23	.25	-.02	1.46		-0.04	
29	EOY	.43	.41	.02		1.04	0.02	
30	EOY	.03	.04	.00		-0.70	-0.01	
31	EOY	.35	.38	-.03		-2.98	-0.03	*
32	EOY	.43	.44	-.02	0.78		-0.03	
33	EOY	.55	.58	-.03		-1.86	-0.03	
34	EOY	.42	.45	-.03	1.92		-0.05	
35	EOY	.48	.52	-.04	5.22		-0.08	*
36	EOY	.05	.05	.00	0.08		0.01	
37	EOY	.45	.45	.01	0.10		0.01	
38	EOY	.37	.39	-.03	1.89		-0.05	
39	EOY	.18	.17	.01		0.62	0.01	
40	EOY	.09	.10	-.01		-1.64	-0.01	
41	EOY	.03	.03	.00	0.35		-0.02	
42	EOY	.03	.04	-.01	1.94		-0.05	
43	EOY	.62	.60	.02	1.19		0.04	
44	EOY	.37	.38	-.01		-0.56	-0.01	
45	EOY	.38	.39	-.01		-1.19	-0.01	
46	EOY	.16	.17	-.01		-1.29	-0.02	
47	EOY	.03	.03	-.01		-1.65	-0.03	
48	EOY	.03	.04	-.01		-2.13	-0.04	*
49	EOY	.24	.27	-.03	3.01		-0.06	

50	EOY	.04	.05	-.01	2.89		-0.06	
51	EOY	.03	.04	.00		-0.94	-0.02	
52	EOY	.01	.02	-.01	9.32		-0.12	*
53	EOY	.04	.04	.00	0.01		0.00	

Table B.4

Comparison of Task Means for Geometry (Matching Only on Demographics)

Task Number	Component	p TC	p NTC	p Diff.	McNemar χ^2	*t*	Effect Size	Sig.	Device Effect
1	PBA	.35	.41	-.06	5.98		-0.13	*	
2	PBA	.66	.67	-.01	0.18		-0.02		
3	PBA	.58	.63	-.04	2.42		-0.08		
4	PBA	.45	.53	-.08	8.38		-0.16	*	
5	PBA	.08	.09	-.01	0.26		-0.03		
6	PBA	.02	.04	-.02	3.93		-0.11	*	
7	PBA	.44	.44	-.01	0.10		-0.01		
8	PBA	.56	.58	-.01	0.33		-0.03		
9	PBA	.12	.14	-.01	0.72		-0.04		
10	PBA	.25	.24	.01	0.26		0.03		
11	PBA	.12	.14	-.03		-2.33	-0.02	*	
12	PBA	.05	.05	.00		-0.45	-0.01		
13	PBA	.02	.03	.00		-0.60	-0.01		
14	PBA	.08	.08	.00		0.00	0.00		
15	PBA	.01	.01	.00		-1.04	-0.01		
16	PBA	.08	.11	-.03		-2.48	-0.04	*	
17	PBA	.03	.03	.00		0.19	0.00		
18	PBA	.00	.00	.00		-1.07	-0.02		
19	EOY	.40	.43	-.03	1.13		-0.07		
20	EOY	.49	.47	.02	0.56		0.05		
21	EOY	.06	.08	-.03	2.88		-0.10		
22	EOY	.06	.06	.00		0.08	0.00		
23	EOY	.02	.04	-.01	1.38		-0.08		
24	EOY	.16	.21	-.04	2.64		-0.10		
25	EOY	.42	.42	.00	0.00		0.00		
26	EOY	.04	.07	-.03	6.40		-0.15	*	
27	EOY	.54	.61	-.07	5.31		-0.14	*	
28	EOY	.02	.03	-.02	2.67		-0.11		
29	EOY	.51	.59	-.08	6.75		-0.16	*	
30	EOY	.20	.17	.03		2.06	0.07	*	
31	EOY	.27	.26	.00		0.30	0.00		
32	EOY	.12	.09	.03	2.19		0.09		

33	EOY	.23	.25	-.02		-1.25	-0.04	
34	EOY	.46	.45	.01		0.42	0.01	
35	EOY	.21	.23	-.02		-1.00	-0.03	
36	EOY	.11	.11	-.01	0.11		-0.02	
37	EOY	.54	.51	.03	0.73		0.06	
38	EOY	.45	.45	.00		0.14	0.00	
39	EOY	.22	.24	-.02		-1.05	-0.02	
40	EOY	.11	.11	.01	0.11		0.02	
41	EOY	.09	.08	.01	0.55		0.05	
42	EOY	.48	.52	-.04		-1.85	-0.03	
43	EOY	.14	.10	.04	2.86		0.11	
44	EOY	.07	.07	.00		-0.29	-0.01	
45	EOY	.28	.28	.00		0.05	0.00	
46	EOY	.06	.06	.00	0.02		0.01	
47	EOY	.13	.13	.00	0.04		0.01	
48	EOY	.21	.20	.01		0.73	0.02	
49	EOY	.04	.04	-.01	0.26		-0.03	
50	EOY	.18	.18	.01		0.37	0.01	
51	EOY	.48	.50	-.02		-0.99	-0.03	
52	EOY	.30	.30	.01	0.05		0.01	
53	EOY	.47	.49	-.03	0.73		-0.05	

Table B.5

Comparison of Task Means for Algebra 2 (Matching Only on Demographics)

Task Number	Component	p TC	p NTC	p Diff.	McNemar χ^2	*t*	Effect Size	Sig.	Device Effect
1	PBA	.70	.68	.02	0.48		0.03		
2	PBA	.14	.12	.02	1.69		0.06		
3	PBA	.39	.35	.04	3.12		0.09		
4	PBA	.17	.14	.03	2.75		0.08		
5	PBA	.13	.11	.01	0.92		0.04		
6	PBA	.01	.02	.00	0.43		-0.03		
7	PBA	.16	.13	.03	3.13		0.08		
8	PBA	.72	.71	.01	0.16		0.02		
9	PBA	.23	.18	.05	5.92		0.12	*	
10	PBA	.01	.01	.00	0.00		0.00		
11	PBA	.16	.16	.00		0.37	0.00		
12	PBA	.05	.05	-.01		-0.91	-0.02		
13	PBA	.02	.03	.00		-0.29	0.00		
14	PBA	.13	.11	.02		1.37	0.02		
15	PBA	.16	.15	.01		1.10	0.01		

16	PBA	.14	.12	.02		2.02	0.02	*
17	PBA	.01	.01	.00		0.32	0.01	
18	PBA	.27	.26	.00		0.24	0.00	
19	PBA	.26	.27	-.01		-0.43	0.00	
20	PBA	.05	.04	.01		1.48	0.02	
21	EOY	.25	.23	.03	1.37		0.06	
22	EOY	.25	.25	.00	0.00		0.00	
23	EOY	.37	.34	.02	0.83		0.05	
24	EOY	.04	.03	.01	1.45		0.06	
25	EOY	.07	.08	.00	0.04		-0.01	
26	EOY	.33	.32	.01		0.77	0.02	
27	EOY	.13	.11	.02	2.35		0.08	
28	EOY	.01	.01	-.01	1.33		-0.06	
29	EOY	.71	.70	.00	0.03		0.01	
30	EOY	.38	.35	.03	1.77		0.07	
31	EOY	.49	.47	.02	0.88		0.05	
32	EOY	.20	.18	.02	0.72		0.04	
33	EOY	.69	.66	.03		1.61	0.04	
34	EOY	.15	.15	-.01	0.16		-0.02	
35	EOY	.56	.61	-.05		-2.29	-0.06	*
36	EOY	.32	.31	.01		0.80	0.02	
37	EOY	.07	.08	-.01	0.62		-0.04	
38	EOY	.15	.11	.03	3.50		0.10	
39	EOY	.33	.43	-.09	13.30		-0.19	*
40	EOY	.12	.09	.03	3.97		0.10	*
41	EOY	.22	.21	.01		0.45	0.01	
42	EOY	.54	.54	.00		-0.10	0.00	
43	EOY	.14	.10	.04	6.72		0.14	*
44	EOY	.15	.11	.04		2.83	0.07	*
45	EOY	.23	.21	.02		1.32	0.03	
46	EOY	.15	.14	.02		1.38	0.02	
47	EOY	.05	.04	.01		0.68	0.02	
48	EOY	.00	.01	-.01				
49	EOY	.01	.01	.00	0.00		0.00	
50	EOY	.13	.12	.01		0.66	0.02	
51	EOY	.12	.12	.00		-0.22	-0.01	
52	EOY	.47	.46	.01		0.59	0.02	

Table B.6

Comparison of Task Means for Grade 3 ELA/L (Matching Only on Demographics)

Task Number	Component	p TC	p NTC	p Diff.	McNemar χ^2	*t*	Effect Size	Sig.	Device Effect
1	PBA	.52	.53	-.02		-1.20	-0.02		
2	PBA	.68	.69	.00		-0.17	0.00		
3	PBA	.42	.41	.02		1.28	0.02		
4	PBA	.64	.67	-.03		-2.22	-0.03	*	
5	PBA	.50	.50	.00		0.02	0.00		
6	PBA	.32	.33	-.01		-0.95	-0.01		
7	PBA	.18	.19	-.01		-0.90	0.00		
8	PBA	.35	.35	.00		-0.09	0.00		
9	PBA	.41	.41	-.01		-0.47	-0.01		
10	PBA	.51	.52	-.01		-0.53	-0.01		
11	PBA	.56	.57	.00		-0.17	0.00		
12	PBA	.55	.53	.02		1.80	0.03		
13	PBA	.66	.67	-.01		-0.63	-0.01		
14	PBA	.22	.22	.00		-0.20	0.00		
15	PBA	.69	.67	.02		1.87	0.03		
16	PBA	.49	.47	.02		1.59	0.02		
17	PBA	.51	.50	.01		0.79	0.01		
18	PBA	.53	.52	.01		0.68	0.01		
19	PBA	.42	.44	-.02		-1.81	-0.03		
20	PBA	.34	.34	.00		-0.19	0.00		
21	EOY	.58	.57	.01		0.90	0.01		
22	EOY	.42	.41	.01		1.52	0.02		
23	EOY	.35	.34	.01		1.41	0.02		
24	EOY	.52	.52	.00		0.20	0.00		
25	EOY	.69	.69	.00		0.00	0.00		
26	EOY	.72	.73	-.01		-0.82	-0.01		
27	EOY	.39	.38	.00		0.30	0.00		
28	EOY	.14	.13	.00		0.26	0.00		
29	EOY	.48	.47	.00		0.05	0.00		
30	EOY	.29	.29	.01		0.83	0.01		
31	EOY	.45	.44	.01		0.90	0.01		
32	EOY	.26	.24	.03		2.52	0.03	*	

Table B.7

Comparison of Task Means for Grade 7 ELA/L (Matching Only on Demographics)

Task Number	Component	p TC	p NTC	p Diff.	McNemar χ^2	*t*	Effect Size	Sig.	Device Effect
1	PBA	.72	.72	.00		0.17	0.00		
2	PBA	.57	.55	.02		1.42	0.02		
3	PBA	.52	.47	.05		3.92	0.05	*	
4	PBA	.61	.60	.01		1.21	0.01		
5	PBA	.50	.49	.02		1.68	0.02		
6	PBA	.46	.44	.01		1.37	0.02		
7	PBA	.31	.29	.02		3.98	0.00	*	
8	PBA	.70	.69	.01		1.04	0.01		
9	PBA	.50	.47	.03		2.07	0.03	*	
10	PBA	.25	.25	.01		0.60	0.01		
11	PBA	.39	.38	.02		1.55	0.02		
12	PBA	.27	.27	.00		0.13	0.00		
13	PBA	.33	.33	.00		-0.23	0.00		
14	PBA	.35	.32	.02		1.97	0.03	*	
15	PBA	.45	.44	.00		0.39	0.01		
16	PBA	.42	.38	.04		5.35	0.01	*	
17	PBA	.65	.66	.00		-0.13	0.00		
18	PBA	.43	.42	.02		1.64	0.02		
19	PBA	.62	.58	.04		3.02	0.04	*	
20	PBA	.71	.73	-.02		-1.64	-0.02		
21	PBA	.36	.35	.01		1.27	0.02		
22	PBA	.35	.32	.03		3.74	0.01	*	
23	EOY	.66	.65	.00		0.24	0.00		
24	EOY	.61	.60	.01		0.74	0.01		
25	EOY	.37	.36	.02		1.50	0.02		
26	EOY	.49	.48	.01		0.97	0.01		
27	EOY	.77	.77	.00		-0.12	0.00		
28	EOY	.42	.42	.00		0.05	0.00		
29	EOY	.46	.44	.01		1.54	0.02		
30	EOY	.56	.56	.00		-0.23	0.00		
31	EOY	.71	.71	.00		0.28	0.00		
32	EOY	.37	.36	.01		0.82	0.01		
33	EOY	.37	.36	.01		1.37	0.02		
34	EOY	.48	.47	.01		0.88	0.01		
35	EOY	.42	.39	.03		2.34	0.03	*	
36	EOY	.73	.74	-.01		-1.08	-0.01		
37	EOY	.43	.42	.01		0.74	0.01		
38	EOY	.27	.25	.01		1.51	0.02		

39	EOY	.35	.34	.01	0.99	0.01	
40	EOY	.43	.44	-.01	-1.21	-0.02	
41	EOY	.46	.47	-.01	-1.14	-0.01	
42	EOY	.50	.48	.02	2.06	0.02	*
43	EOY	.31	.32	-.01	-1.14	-0.01	
44	EOY	.44	.44	.00	0.28	0.00	

Table B.7

Comparison of Task Means for Grade 7 ELA/L (Matching Only on Demographics)

Task Number	Component	p TC	p NTC	p Diff.	McNemar χ^2	*t*	Effect Size	Sig.	Device Effect
1	PBA	.72	.72	.00		0.17	0.00		
2	PBA	.57	.55	.02		1.42	0.02		
3	PBA	.52	.47	.05		3.92	0.05	*	
4	PBA	.61	.60	.01		1.21	0.01		
5	PBA	.50	.49	.02		1.68	0.02		
6	PBA	.46	.44	.01		1.37	0.02		
7	PBA	.31	.29	.02		3.98	0.00	*	
8	PBA	.70	.69	.01		1.04	0.01		
9	PBA	.50	.47	.03		2.07	0.03	*	
10	PBA	.25	.25	.01		0.60	0.01		
11	PBA	.39	.38	.02		1.55	0.02		
12	PBA	.27	.27	.00		0.13	0.00		
13	PBA	.33	.33	.00		-0.23	0.00		
14	PBA	.35	.32	.02		1.97	0.03	*	
15	PBA	.45	.44	.00		0.39	0.01		
16	PBA	.42	.38	.04		5.35	0.01	*	
17	PBA	.65	.66	.00		-0.13	0.00		
18	PBA	.43	.42	.02		1.64	0.02		
19	PBA	.62	.58	.04		3.02	0.04	*	
20	PBA	.71	.73	-.02		-1.64	-0.02		
21	PBA	.36	.35	.01		1.27	0.02		
22	PBA	.35	.32	.03		3.74	0.01	*	
23	EOY	.66	.65	.00		0.24	0.00		
24	EOY	.61	.60	.01		0.74	0.01		
25	EOY	.37	.36	.02		1.50	0.02		
26	EOY	.49	.48	.01		0.97	0.01		
27	EOY	.77	.77	.00		-0.12	0.00		
28	EOY	.42	.42	.00		0.05	0.00		
29	EOY	.46	.44	.01		1.54	0.02		
30	EOY	.56	.56	.00		-0.23	0.00		

31	EOY	.71	.71	.00	0.28	0.00	
32	EOY	.37	.36	.01	0.82	0.01	
33	EOY	.37	.36	.01	1.37	0.02	
34	EOY	.48	.47	.01	0.88	0.01	
35	EOY	.42	.39	.03	2.34	0.03	*
36	EOY	.73	.74	-.01	-1.08	-0.01	
37	EOY	.43	.42	.01	0.74	0.01	
38	EOY	.27	.25	.01	1.51	0.02	
39	EOY	.35	.34	.01	0.99	0.01	
40	EOY	.43	.44	-.01	-1.21	-0.02	
41	EOY	.46	.47	-.01	-1.14	-0.01	
42	EOY	.50	.48	.02	2.06	0.02	*
43	EOY	.31	.32	-.01	-1.14	-0.01	
44	EOY	.44	.44	.00	0.28	0.00	

Appendix C: Comparison of IRT Difficulties

Table C.1

Comparison of IRT Difficulties for Grade 5 Mathematics (Matching Only on Demographics)

Task Number	Component	*b* TC	*b* NTC	*b* Diff.	Sig.	Device Effect
1	PBA	0.07	0.12	-0.05		
2	PBA	2.92	3.01	-0.09		
3	PBA	0.85	0.80	0.05		
4	PBA	-0.50	-0.46	-0.04		
5	PBA	0.93	0.97	-0.04		
6	PBA	0.93	0.88	0.05		
7	PBA	1.43	1.35	0.09	*	
8	PBA	2.80	2.69	0.12	*	
9	PBA	1.29	1.14	0.15	*	
10	PBA	-2.07	-2.16	0.10		
11	PBA	-1.94	-1.89	-0.05		
12	PBA	-0.14	-0.07	-0.07	*	
13	PBA	1.26	1.28	-0.02		
14	PBA	1.97	2.10	-0.13	*	
15	PBA	3.00	3.12	-0.12	*	
16	PBA	2.32	2.28	0.04		
17	EOY	1.05	1.01	0.04		
18	EOY	0.44	0.52	-0.08		
19	EOY	-2.01	-2.04	0.02		
20	EOY	1.14	0.94	0.20	*	
21	EOY	0.23	0.18	0.05		
22	EOY	-0.43	-0.52	0.09		
23	EOY	1.92	2.44	-0.52	*	*
24	EOY	0.84	1.02	-0.18	*	
25	EOY	-0.73	-0.54	-0.18	*	
26	EOY	1.54	1.79	-0.25	*	
27	EOY	-2.17	-1.97	-0.20	*	
28	EOY	0.56	0.64	-0.08		
29	EOY	2.58	2.67	-0.09		
30	EOY	2.37	2.46	-0.08		
31	EOY	-1.61	-1.52	-0.10		
32	EOY	2.15	2.02	0.13		
33	EOY	-1.21	-1.21	0.00		
34	EOY	-1.57	-1.60	0.03		
35	EOY	-0.96	-0.92	-0.04		
36	EOY	-0.05	0.03	-0.08		

37	EOY	-1.54	-1.59	0.05	
38	EOY	1.83	2.03	-0.20	*
39	EOY	0.75	0.86	-0.10	*
40	EOY	-1.00	-0.85	-0.16	*
41	EOY	-1.72	-1.67	-0.05	
42	EOY	-1.12	-0.84	-0.28	*
43	EOY	0.07	-0.01	0.08	*
44	EOY	-1.45	-1.51	0.06	
45	EOY	3.09	2.88	0.21	*
46	EOY	0.32	0.35	-0.04	
47	EOY	-0.64	-0.60	-0.04	
48	EOY	-0.69	-0.66	-0.03	
49	EOY	-1.80	-1.91	0.12	
50	EOY	0.76	0.88	-0.13	*
51	EOY	2.18	2.34	-0.16	*
52	EOY	1.01	1.20	-0.18	*

Table C.2
Comparison of IRT Difficulties for Grade 7 Mathematics (Matching Only on Demographics)

Task Number	Component	*b* TC	*b* NTC	*b* Diff.	Sig.	Device Effect
1	PBA	1.29	1.39	-0.10	*	
2	PBA	-0.49	-0.54	0.06		
3	PBA	1.52	1.63	-0.11	*	
4	PBA	1.41	1.51	-0.10	*	
5	PBA	-0.72	-0.65	-0.07		
6	PBA	-0.90	-0.89	-0.01		
7	PBA	0.21	0.29	-0.08	*	
8	PBA	3.57	3.70	-0.13		
9	PBA	-0.05	-0.21	0.16	*	
10	PBA	0.46	0.33	0.14	*	
11	PBA	1.82	1.96	-0.14	*	
12	PBA	2.16	2.08	0.08	*	
13	PBA	1.42	1.51	-0.09	*	
14	PBA	0.68	0.49	0.19	*	
15	PBA	1.50	1.50	-0.01		
16	PBA	2.47	2.35	0.12	*	
17	PBA	2.72	2.66	0.06		
18	EOY	-1.84	-1.73	-0.12	*	
19	EOY	-0.63	-0.65	0.02		
20	EOY	4.65	4.81	-0.17		

21	EOY	3.59	3.55	0.04	
22	EOY	1.76	1.68	0.08	
23	EOY	3.68	3.57	0.10	
24	EOY	2.97	3.09	-0.12	
25	EOY	2.37	2.44	-0.06	
26	EOY	-0.63	-0.67	0.03	
27	EOY	-1.92	-1.94	0.02	
28	EOY	2.91	2.91	0.00	
29	EOY	-0.66	-0.78	0.12	*
30	EOY	-1.38	-1.28	-0.10	*
31	EOY	2.92	2.92	0.00	
32	EOY	-0.18	-0.24	0.06	
33	EOY	2.73	2.81	-0.08	
34	EOY	-1.41	-1.32	-0.09	*
35	EOY	2.54	2.66	-0.11	*
36	EOY	3.91	3.77	0.14	
37	EOY	-0.42	-0.51	0.09	*
38	EOY	1.31	1.30	0.02	
39	EOY	0.83	0.86	-0.03	
40	EOY	1.86	1.83	0.03	
41	EOY	3.07	3.13	-0.06	
42	EOY	1.18	1.05	0.14	*
43	EOY	-0.55	-0.57	0.02	
44	EOY	-0.44	-0.69	0.25	*
45	EOY	1.62	1.66	-0.03	
46	EOY	-0.16	-0.21	0.04	
47	EOY	3.05	2.96	0.09	
48	EOY	2.30	2.22	0.08	*
49	EOY	2.39	2.49	-0.11	*
50	EOY	2.04	2.03	0.01	

Table C.3
Comparison of IRT Difficulties for Algebra 1 (Matching Only on Demographics)

Task Number	Component	*b* TC	*b* NTC	*b* Diff.	Sig.	Device Effect
1	PBA	0.02	0.06	-0.04		
2	PBA	3.75	3.68	0.07		
3	PBA	-0.48	-0.31	-0.18	*	
4	PBA	0.56	0.57	-0.01		
5	PBA	0.61	0.67	-0.06		
6	PBA	-0.50	-0.55	0.05		

7	PBA	3.40	3.32	0.08		
8	PBA	3.45	3.59	-0.15		
9	PBA	-1.20	-1.17	-0.03		
10	PBA	3.01	3.08	-0.08		
11	PBA	3.13	3.17	-0.04		
12	PBA	3.62	3.56	0.06		
13	PBA	2.89	2.74	0.15	*	
14	PBA	3.84	3.89	-0.05		
15	PBA	2.17	2.08	0.09	*	
16	PBA	2.19	2.33	-0.14	*	
17	PBA	2.32	2.27	0.05		
18	PBA	3.23	3.26	-0.03		
19	EOY	0.07	0.07	0.00		
20	EOY	3.17	3.21	-0.05		
21	EOY	2.56	2.32	0.24	*	
22	EOY	1.21	1.41	-0.19	*	
23	EOY	0.51	0.52	-0.01		
24	EOY	0.83	0.96	-0.14	*	
25	EOY	4.85	4.32	0.54	*	*
26	EOY	3.14	3.05	0.09		
27	EOY	-0.74	-0.87	0.12	*	
28	EOY	1.36	1.24	0.11		
29	EOY	0.24	0.31	-0.07		
30	EOY	3.23	3.22	0.02		
31	EOY	0.69	0.53	0.16	*	
32	EOY	0.30	0.23	0.08		
33	EOY	-0.25	-0.36	0.11	*	
34	EOY	0.34	0.22	0.12	*	
35	EOY	0.04	-0.15	0.19	*	
36	EOY	3.30	3.35	-0.05		
37	EOY	0.20	0.23	-0.03		
38	EOY	0.59	0.47	0.13	*	
39	EOY	1.35	1.42	-0.08		
40	EOY	2.53	2.30	0.24	*	
41	EOY	4.09	3.94	0.15		
42	EOY	3.97	3.65	0.32		
43	EOY	-0.60	-0.51	-0.09		
44	EOY	0.62	0.59	0.03		
45	EOY	0.54	0.54	0.00		
46	EOY	1.93	1.77	0.16	*	
47	EOY	3.94	4.17	-0.23		
48	EOY	3.74	3.22	0.52	*	*
49	EOY	1.31	1.16	0.15	*	

50	EOY	3.68	3.34	0.34	*	*
51	EOY	3.47	3.53	-0.05		
52	EOY	5.76	4.48	1.28	*	*
53	EOY	3.57	3.59	-0.02		

Table C.4
Comparison of IRT Difficulties for Geometry (Matching Only on Demographics)

Task Number	Component	*b* TC	*b* NTC	*b* Diff.	Sig.	Device Effect
1	PBA	0.73	0.41	0.32	*	*
2	PBA	-0.92	-0.97	0.05		
3	PBA	-0.50	-0.71	0.21	*	
4	PBA	0.19	-0.20	0.40	*	*
5	PBA	2.86	2.77	0.09		
6	PBA	4.27	3.61	0.66	*	*
7	PBA	0.25	0.23	0.02		
8	PBA	-0.39	-0.46	0.07		
9	PBA	2.32	2.19	0.13		
10	PBA	1.27	1.37	-0.10		
11	PBA	1.91	1.70	0.21	*	
12	PBA	2.75	3.03	-0.28	*	
13	PBA	3.17	3.44	-0.27		
14	PBA	2.08	2.11	-0.03		
15	PBA	3.35	3.11	0.24		
16	PBA	2.06	1.84	0.22	*	
17	PBA	2.59	2.48	0.11		
18	PBA	5.89	3.90	1.99	*	*
19	EOY	0.47	0.30	0.17		
20	EOY	-0.02	0.09	-0.11		
21	EOY	3.32	2.91	0.41		
22	EOY	2.62	2.65	-0.02		
23	EOY	4.29	3.88	0.41		
24	EOY	1.93	1.62	0.30	*	*
25	EOY	0.34	0.34	0.00		
26	EOY	3.75	3.05	0.70	*	*
27	EOY	-0.23	-0.59	0.36	*	*
28	EOY	4.63	3.95	0.68		
29	EOY	-0.09	-0.49	0.41	*	*
30	EOY	3.06	2.70	0.36	*	*
31	EOY	1.41	1.58	-0.17	*	
32	EOY	2.41	2.82	-0.40	*	*

33	EOY	1.62	1.57	0.05		
34	EOY	0.15	0.17	-0.02		
35	EOY	1.62	1.58	0.04		
36	EOY	2.52	2.48	0.04		
37	EOY	-0.21	-0.09	-0.13		
38	EOY	0.19	0.22	-0.04		
39	EOY	1.31	1.23	0.08		
40	EOY	2.44	2.56	-0.12		
41	EOY	2.71	2.95	-0.24		
42	EOY	0.05	-0.13	0.18	*	
43	EOY	2.14	2.58	-0.44	*	*
44	EOY	3.17	2.94	0.23		
45	EOY	1.26	1.14	0.12		
46	EOY	3.28	3.38	-0.10		
47	EOY	2.23	2.31	-0.08		
48	EOY	2.04	2.03	0.00		
49	EOY	3.81	3.69	0.12		
50	EOY	2.02	1.94	0.08		
51	EOY	0.07	-0.06	0.12		
52	EOY	0.98	1.02	-0.05		
53	EOY	0.11	-0.02	0.13		

Table C.5
Comparison of IRT Difficulties for Algebra 2 (Matching Only on Demographics)

Task Number	Component	*b* TC	*b* NTC	*b* Diff.	Sig.	Device Effect
1	PBA	-1.12	-1.05	-0.07		
2	PBA	2.15	2.34	-0.19		
3	PBA	0.55	0.74	-0.19	*	
4	PBA	1.96	2.18	-0.22	*	
5	PBA	2.32	2.44	-0.12		
6	PBA	4.73	4.47	0.26		
7	PBA	1.99	2.23	-0.24	*	
8	PBA	-1.26	-1.23	-0.03		
9	PBA	1.47	1.79	-0.32	*	*
10	PBA	5.05	5.04	0.01		
11	PBA	1.51	1.48	0.04		
12	PBA	3.35	2.93	0.43	*	*
13	PBA	2.79	2.54	0.25	*	
14	PBA	2.15	2.33	-0.19	*	
15	PBA	2.49	2.49	0.01		

16	PBA	2.22	2.26	-0.04		
17	PBA	4.23	3.93	0.29		
18	PBA	1.01	1.01	0.00		
19	PBA	1.26	1.22	0.04		
20	PBA	3.33	3.11	0.21	*	
21	EOY	1.27	1.43	-0.16		
22	EOY	1.27	1.27	0.00		
23	EOY	0.62	0.74	-0.12		
24	EOY	3.57	3.91	-0.34		
25	EOY	2.92	2.85	0.07		
26	EOY	0.94	0.97	-0.03		
27	EOY	2.19	2.44	-0.25		
28	EOY	5.43	4.82	0.62		
29	EOY	-1.05	-1.01	-0.04		
30	EOY	0.55	0.73	-0.18	*	
31	EOY	0.00	0.12	-0.13		
32	EOY	1.65	1.77	-0.12		
33	EOY	-0.94	-0.80	-0.14	*	
34	EOY	2.05	1.97	0.08		
35	EOY	-0.26	-0.45	0.19	*	
36	EOY	1.31	1.31	0.00		
37	EOY	3.06	2.87	0.19		
38	EOY	2.04	2.36	-0.32	*	*
39	EOY	0.80	0.34	0.46	*	*
40	EOY	2.30	2.66	-0.36	*	*
41	EOY	1.61	1.63	-0.02		
42	EOY	-0.15	-0.13	-0.02		
43	EOY	2.11	2.58	-0.47	*	*
44	EOY	2.13	2.35	-0.23	*	
45	EOY	1.47	1.82	-0.36	*	*
46	EOY	1.67	1.86	-0.20	*	
47	EOY	2.47	2.50	-0.02		
48	EOY		5.65		*	
49	EOY	5.66	5.65	0.01		
50	EOY	1.79	1.84	-0.05		
51	EOY	2.09	2.23	-0.14		
52	EOY	0.10	0.17	-0.07		

Table C.6
Comparison of IRT Difficulties for Grade 3 ELA/L (Matching Only on Demographics)

Task Number	Component	*b* TC	*b* NTC	*b* Diff.	Sig.	Device Effect
1	PBA	-0.09	-0.15	0.06	*	
2	PBA	-0.71	-0.71	0.00		
3	PBA	0.30	0.37	-0.06	*	
4	PBA	-0.54	-0.64	0.10	*	
5	PBA	-0.04	-0.04	0.00		
6	PBA	0.74	0.69	0.05		
7	PBA	2.11	2.02	0.10	*	
8	PBA	0.57	0.56	0.01		
9	PBA	0.33	0.31	0.03		
10	PBA	-0.07	-0.09	0.03		
11	PBA	-0.25	-0.26	0.01		
12	PBA	-0.22	-0.13	-0.09	*	
13	PBA	-0.62	-0.64	0.02		
14	PBA	2.02	2.07	-0.04		
15	PBA	-0.71	-0.61	-0.10	*	
16	PBA	0.01	0.09	-0.07	*	
17	PBA	-0.06	-0.01	-0.05		
18	PBA	-0.13	-0.10	-0.03		
19	PBA	0.33	0.23	0.10	*	
20	PBA	1.06	1.06	0.00		
21	EOY	-0.34	-0.30	-0.04		
22	EOY	0.37	0.44	-0.07	*	
23	EOY	1.24	1.28	-0.04		
24	EOY	-0.09	-0.08	-0.01		
25	EOY	-0.73	-0.71	-0.02		
26	EOY	-0.82	-0.84	0.02		
27	EOY	0.39	0.40	-0.01		
28	EOY	1.83	1.79	0.05		
29	EOY	0.07	0.08	-0.01		
30	EOY	1.81	1.80	0.01		
31	EOY	0.28	0.31	-0.03		
32	EOY	0.90	0.99	-0.09	*	

Table C.7

Comparison of IRT Difficulties for Grade 7 ELA/L (Matching Only on Demographics)

Task Number	Component	*b* TC	*b* NTC	*b* Diff.	Sig.	Device Effect
1	PBA	-0.74	-0.71	-0.02		
2	PBA	-0.21	-0.15	-0.06	*	
3	PBA	-0.06	0.10	-0.16	*	
4	PBA	-0.37	-0.32	-0.05		
5	PBA	-0.01	0.07	-0.08	*	
6	PBA	0.17	0.24	-0.06	*	
7	PBA	1.01	1.15	-0.14	*	
8	PBA	-0.75	-0.69	-0.06	*	
9	PBA	0.01	0.09	-0.08	*	
10	PBA	0.84	0.88	-0.03		
11	PBA	0.37	0.42	-0.05	*	
12	PBA	0.95	0.94	0.01		
13	PBA	0.55	0.53	0.01		
14	PBA	0.51	0.58	-0.07	*	
15	PBA	0.20	0.22	-0.02		
16	PBA	0.48	0.60	-0.12	*	
17	PBA	-0.48	-0.49	0.00		
18	PBA	0.34	0.39	-0.06		
19	PBA	-0.37	-0.25	-0.12	*	
20	PBA	-0.73	-0.80	0.07	*	
21	PBA	0.55	0.60	-0.05		
22	PBA	0.48	0.58	-0.10	*	
23	EOY	-0.68	-0.67	-0.01		
24	EOY	-0.36	-0.32	-0.04		
25	EOY	0.47	0.53	-0.05	*	
26	EOY	0.03	0.07	-0.04		
27	EOY	-0.91	-0.90	-0.01		
28	EOY	0.37	0.39	-0.02		
29	EOY	0.21	0.28	-0.07	*	
30	EOY	-0.20	-0.20	0.01		
31	EOY	-0.70	-0.69	-0.01		
32	EOY	0.45	0.48	-0.03		
33	EOY	0.61	0.66	-0.05		
34	EOY	0.08	0.11	-0.04		
35	EOY	0.26	0.34	-0.08	*	
36	EOY	-1.13	-1.23	0.10	*	
37	EOY	0.26	0.28	-0.03		
38	EOY	1.44	1.53	-0.09	*	

39	EOY	0.52	0.56	-0.04		
40	EOY	0.25	0.21	0.05		
41	EOY	0.29	0.23	0.06		
42	EOY	0.00	0.08	-0.08	*	
43	EOY	1.20	1.14	0.06		
44	EOY	0.29	0.31	-0.02		

Table C.8
Comparison of IRT Difficulties for Grade 9 ELA/L (Matching Only on Demographics)

Task Number	Component	*b* TC	*b* NTC	*b* Diff.	Sig.	Device Effect
1	PBA	0.14	0.10	0.04		
2	PBA	-0.41	-0.20	-0.21	*	
3	PBA	0.05	0.18	-0.13	*	
4	PBA	0.37	0.41	-0.04		
5	PBA	-0.95	-0.83	-0.12	*	
6	PBA	-0.26	-0.23	-0.04		
7	PBA	1.13	1.33	-0.21	*	
8	PBA	-0.47	-0.38	-0.10	*	
9	PBA	-0.11	0.00	-0.11	*	
10	PBA	-1.16	-1.23	0.07		
11	PBA	-1.09	-0.98	-0.11	*	
12	PBA	-1.32	-1.26	-0.06		
13	PBA	-0.15	0.00	-0.15	*	
14	PBA	-0.31	-0.19	-0.12	*	
15	PBA	0.55	0.60	-0.06		
16	PBA	0.10	0.20	-0.10	*	
17	PBA	0.79	0.87	-0.08	*	
18	PBA	1.35	1.40	-0.05		
19	PBA	0.36	0.40	-0.04		
20	PBA	0.76	0.90	-0.15	*	
21	PBA	-0.40	-0.27	-0.12	*	
22	PBA	-0.29	-0.23	-0.06		
23	PBA	0.36	0.50	-0.13	*	
24	EOY	-0.46	-0.37	-0.09	*	
25	EOY	0.85	0.93	-0.07		
26	EOY	-0.17	-0.11	-0.06		
27	EOY	0.06	0.06	0.00		
28	EOY	1.30	1.38	-0.08		
29	EOY	0.66	0.71	-0.05		
30	EOY	0.37	0.38	-0.01		

31	EOY	1.69	1.79	-0.10	
32	EOY	0.50	0.44	0.06	
33	EOY	0.30	0.36	-0.06	
34	EOY	1.77	1.75	0.02	
35	EOY	-0.19	-0.10	-0.09	*
36	EOY	0.44	0.59	-0.15	*
37	EOY	0.88	0.91	-0.03	
38	EOY	-0.23	-0.09	-0.14	*
39	EOY	0.63	0.63	-0.01	
40	EOY	-0.46	-0.37	-0.09	*
41	EOY	0.33	0.41	-0.08	*
42	EOY	0.02	0.02	-0.01	
43	EOY	0.36	0.47	-0.11	*
44	EOY	1.00	1.15	-0.15	*
45	EOY	-0.24	-0.25	0.01	

Appendix D: Raw-Score Equivalents

Table D.1

Raw-Score Equivalents for Grade 5 Mathematics PBA (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-3.09	0.99	1	-0.01	0	
2	-2.22	1.98	2	-0.02	0	
3	-1.67	2.96	3	-0.04	0	
4	-1.26	3.93	4	-0.07	0	
5	-0.92	4.90	5	-0.10	0	
6	-0.64	5.86	6	-0.14	0	
7	-0.39	6.81	7	-0.19	0	
8	-0.18	7.77	8	-0.23	0	
9	0.02	8.73	9	-0.27	0	
10	0.20	9.70	10	-0.30	0	
11	0.38	10.68	11	-0.32	0	
12	0.54	11.66	12	-0.34	0	
13	0.70	12.67	13	-0.33	0	
14	0.85	13.67	14	-0.33	0	
15	1.00	14.69	15	-0.31	0	
16	1.14	15.71	16	-0.29	0	
17	1.28	16.75	17	-0.25	0	
18	1.43	17.79	18	-0.21	0	
19	1.57	18.82	19	-0.18	0	
20	1.71	19.87	20	-0.13	0	
21	1.84	20.91	21	-0.09	0	
22	1.98	21.95	22	-0.05	0	
23	2.12	23.00	23	0.00	0	
24	2.25	24.05	24	0.05	0	
25	2.39	25.10	25	0.10	0	
26	2.52	26.16	26	0.16	0	
27	2.66	27.22	27	0.22	0	
28	2.80	28.27	28	0.27	0	
29	2.95	29.31	29	0.31	0	
30	3.10	30.34	30	0.34	0	
31	3.27	31.36	31	0.36	0	
32	3.45	32.35	32	0.35	0	
33	3.66	33.33	33	0.33	0	

34	3.90	34.29	34	0.29	0
35	4.22	35.24	35	0.24	0
36	4.66	36.18	36	0.18	0
37	5.39	37.11	37	0.11	0
38	8.00	37.93	38	-0.07	0

Table D.2
Raw-Score Equivalents for Grade 5 Mathematics EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.02	0	0.02	0	
1	-4.19	1.04	1	0.04	0	
2	-3.44	2.07	2	0.07	0	
3	-2.98	3.10	3	0.10	0	
4	-2.64	4.13	4	0.13	0	
5	-2.36	5.16	5	0.16	0	
6	-2.13	6.19	6	0.19	0	
7	-1.92	7.22	7	0.22	0	
8	-1.73	8.25	8	0.25	0	
9	-1.56	9.27	9	0.27	0	
10	-1.40	10.30	10	0.30	0	
11	-1.25	11.32	11	0.32	0	
12	-1.11	12.34	12	0.34	0	
13	-0.97	13.37	13	0.37	0	
14	-0.84	14.39	14	0.39	0	
15	-0.71	15.40	15	0.40	0	
16	-0.58	16.42	16	0.42	0	
17	-0.45	17.44	17	0.44	0	
18	-0.32	18.45	18	0.45	0	
19	-0.20	19.46	19	0.46	0	
20	-0.07	20.46	20	0.46	0	
21	0.05	21.47	21	0.47	0	
22	0.18	22.47	22	0.47	0	
23	0.31	23.48	23	0.48	0	
24	0.44	24.48	24	0.48	0	
25	0.57	25.48	25	0.48	0	
26	0.70	26.48	26	0.48	0	
27	0.83	27.48	27	0.48	0	
28	0.97	28.48	28	0.48	0	

29	1.11	29.47	29	0.47	0
30	1.25	30.48	30	0.48	0
31	1.40	31.48	31	0.48	0
32	1.55	32.47	32	0.47	0
33	1.71	33.48	33	0.48	0
34	1.87	34.48	34	0.48	0
35	2.05	35.47	35	0.47	0
36	2.24	36.46	36	0.46	0
37	2.44	37.44	37	0.44	0
38	2.66	38.41	38	0.41	0
39	2.91	39.36	39	0.36	0
40	3.20	40.30	40	0.30	0
41	3.55	41.23	41	0.23	0
42	4.02	42.15	42	0.15	0
43	4.77	43.07	43	0.07	0
44	8.00	43.96	44	-0.04	0

Table D.3
Raw-Score Equivalents for Grade 5 Mathematics PBA+EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.04	0	0.04	0	
1	-4.77	0.97	1	-0.03	0	
2	-4.02	1.94	2	-0.06	0	
3	-3.56	2.93	3	-0.07	0	
4	-3.22	3.93	4	-0.07	0	
5	-2.94	4.94	5	-0.06	0	
6	-2.71	5.95	6	-0.05	0	
7	-2.51	6.97	7	-0.03	0	
8	-2.33	7.99	8	-0.01	0	
9	-2.16	9.03	9	0.03	0	
10	-2.01	10.05	10	0.05	0	
11	-1.88	11.09	11	0.09	0	
12	-1.75	12.12	12	0.12	0	
13	-1.63	13.16	13	0.16	0	
14	-1.51	14.20	14	0.20	0	
15	-1.40	15.23	15	0.23	0	
16	-1.30	16.27	16	0.27	0	
17	-1.20	17.31	17	0.31	0	

18	-1.10	18.35	18	0.35	0	
19	-1.01	19.40	19	0.40	0	
20	-0.92	20.44	20	0.44	0	
21	-0.83	21.48	21	0.48	0	
22	-0.75	22.53	23	0.53	1	*
23	-0.66	23.58	24	0.58	1	*
24	-0.58	24.63	25	0.63	1	*
25	-0.50	25.69	26	0.69	1	*
26	-0.42	26.74	27	0.74	1	*
27	-0.34	27.80	28	0.80	1	*
28	-0.26	28.86	29	0.86	1	*
29	-0.18	29.92	30	0.92	1	*
30	-0.11	30.98	31	0.98	1	*
31	-0.03	32.04	32	1.04	1	*
32	0.05	33.10	33	1.10	1	*
33	0.12	34.17	34	1.17	1	*
34	0.20	35.23	35	1.23	1	*
35	0.27	36.29	36	1.29	1	*
36	0.35	37.35	37	1.35	1	*
37	0.42	38.41	38	1.41	1	*
38	0.50	39.47	39	1.47	1	*
39	0.57	40.52	41	1.52	2	*
40	0.65	41.57	42	1.57	2	*
41	0.72	42.63	43	1.63	2	*
42	0.80	43.69	44	1.69	2	*
43	0.87	44.73	45	1.73	2	*
44	0.95	45.79	46	1.79	2	*
45	1.02	46.84	47	1.84	2	*
46	1.10	47.87	48	1.87	2	*
47	1.18	48.93	49	1.93	2	*
48	1.25	49.97	50	1.97	2	*
49	1.33	51.01	51	2.01	2	*
50	1.41	52.05	52	2.05	2	*
51	1.50	53.11	53	2.11	2	*
52	1.58	54.15	54	2.15	2	*
53	1.66	55.19	55	2.19	2	*
54	1.75	56.24	56	2.24	2	*
55	1.84	57.27	57	2.27	2	*
56	1.93	58.31	58	2.31	2	*
57	2.02	59.32	59	2.32	2	*
58	2.11	60.32	60	2.32	2	*
59	2.20	61.31	61	2.31	2	*
60	2.30	62.28	62	2.28	2	*

61	2.40	63.21	63	2.21	2	*
62	2.50	64.12	64	2.12	2	*
63	2.60	65.00	65	2.00	2	*
64	2.71	65.85	66	1.85	2	*
65	2.82	66.68	67	1.68	2	*
66	2.94	67.48	67	1.48	1	*
67	3.06	68.26	68	1.26	1	*
68	3.19	69.04	69	1.04	1	*
69	3.34	69.81	70	0.81	1	*
70	3.50	70.59	71	0.59	1	*
71	3.70	71.39	71	0.39	0	
72	3.93	72.22	72	0.22	0	
73	4.24	73.09	73	0.09	0	
74	4.66	74.00	74	0.00	0	
75	5.38	74.97	75	-0.03	0	
76	8.00	75.92	76	-0.08	0	

Note: 1 task omitted because different score categories were observed in the TC and NTC conditions or because of calibration failure (e.g., due to zero proportion correct) in the TC or NTC condition.

Table D.4

Raw-Score Equivalents for Grade 7 Mathematics PBA (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-2.94	0.95	1	-0.05	0	
2	-2.05	1.94	2	-0.06	0	
3	-1.49	2.93	3	-0.07	0	
4	-1.07	3.93	4	-0.07	0	
5	-0.72	4.94	5	-0.06	0	
6	-0.43	5.94	6	-0.06	0	
7	-0.16	6.96	7	-0.04	0	
8	0.07	7.97	8	-0.03	0	
9	0.28	8.99	9	-0.01	0	
10	0.47	10.01	10	0.01	0	
11	0.64	11.02	11	0.02	0	
12	0.80	12.04	12	0.04	0	
13	0.94	13.04	13	0.04	0	
14	1.07	14.03	14	0.03	0	
15	1.19	15.01	15	0.01	0	

16	1.31	15.98	16	-0.02	0
17	1.42	16.95	17	-0.05	0
18	1.52	17.91	18	-0.09	0
19	1.62	18.87	19	-0.13	0
20	1.72	19.83	20	-0.17	0
21	1.82	20.80	21	-0.20	0
22	1.92	21.76	22	-0.24	0
23	2.02	22.73	23	-0.27	0
24	2.11	23.72	24	-0.28	0
25	2.21	24.71	25	-0.29	0
26	2.31	25.70	26	-0.30	0
27	2.41	26.72	27	-0.28	0
28	2.52	27.73	28	-0.27	0
29	2.64	28.76	29	-0.24	0
30	2.76	29.79	30	-0.21	0
31	2.90	30.83	31	-0.17	0
32	3.05	31.86	32	-0.14	0
33	3.23	32.90	33	-0.10	0
34	3.45	33.93	34	-0.07	0
35	3.73	34.95	35	-0.05	0
36	4.14	35.98	36	-0.02	0
37	4.85	36.99	37	-0.01	0
38	8.00	37.96	38	-0.04	0

Table D.5
Raw-Score Equivalents for Grade 7 Mathematics EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-3.57	1.00	1	0.00	0	
2	-2.79	2.00	2	0.00	0	
3	-2.29	2.98	3	-0.02	0	
4	-1.91	3.97	4	-0.03	0	
5	-1.59	4.95	5	-0.05	0	
6	-1.30	5.93	6	-0.07	0	
7	-1.05	6.92	7	-0.08	0	
8	-0.81	7.91	8	-0.09	0	
9	-0.59	8.90	9	-0.10	0	
10	-0.37	9.89	10	-0.11	0	

11	-0.17	10.88	11	-0.12	0
12	0.03	11.88	12	-0.12	0
13	0.23	12.88	13	-0.12	0
14	0.41	13.88	14	-0.12	0
15	0.59	14.88	15	-0.12	0
16	0.76	15.88	16	-0.12	0
17	0.93	16.88	17	-0.12	0
18	1.09	17.88	18	-0.12	0
19	1.24	18.89	19	-0.11	0
20	1.39	19.89	20	-0.11	0
21	1.53	20.89	21	-0.11	0
22	1.67	21.90	22	-0.10	0
23	1.81	22.91	23	-0.09	0
24	1.94	23.92	24	-0.08	0
25	2.07	24.93	25	-0.07	0
26	2.20	25.95	26	-0.05	0
27	2.32	26.96	27	-0.04	0
28	2.45	27.97	28	-0.03	0
29	2.58	28.98	29	-0.02	0
30	2.71	29.99	30	-0.01	0
31	2.84	31.00	31	0.00	0
32	2.97	32.00	32	0.00	0
33	3.12	33.01	33	0.01	0
34	3.27	34.01	34	0.01	0
35	3.42	35.01	35	0.01	0
36	3.59	36.01	36	0.01	0
37	3.78	37.01	37	0.01	0
38	3.99	38.00	38	0.00	0
39	4.23	39.00	39	0.00	0
40	4.51	40.00	40	0.00	0
41	4.86	41.01	41	0.01	0
42	5.33	42.01	42	0.01	0
43	6.10	43.01	43	0.01	0
44	8.00	43.84	44	-0.16	0

Table D.6

Raw-Score Equivalents for Grade 7 Mathematics PBA+EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-3.82	0.92	1	-0.08	0	
2	-3.06	1.84	2	-0.16	0	
3	-2.60	2.77	3	-0.23	0	
4	-2.26	3.71	4	-0.29	0	
5	-1.98	4.64	5	-0.36	0	
6	-1.74	5.59	6	-0.41	0	
7	-1.53	6.52	7	-0.48	0	
8	-1.34	7.46	7	-0.54	-1	*
9	-1.17	8.41	8	-0.59	-1	*
10	-1.01	9.35	9	-0.65	-1	*
11	-0.86	10.30	10	-0.70	-1	*
12	-0.72	11.24	11	-0.76	-1	*
13	-0.58	12.18	12	-0.82	-1	*
14	-0.46	13.12	13	-0.88	-1	*
15	-0.33	14.07	14	-0.93	-1	*
16	-0.22	15.02	15	-0.98	-1	*
17	-0.10	15.96	16	-1.04	-1	*
18	0.01	16.91	17	-1.09	-1	*
19	0.11	17.86	18	-1.14	-1	*
20	0.21	18.80	19	-1.20	-1	*
21	0.31	19.75	20	-1.25	-1	*
22	0.41	20.70	21	-1.30	-1	*
23	0.50	21.64	22	-1.36	-1	*
24	0.59	22.58	23	-1.42	-1	*
25	0.68	23.52	24	-1.48	-1	*
26	0.76	24.46	24	-1.54	-2	*
27	0.84	25.39	25	-1.61	-2	*
28	0.92	26.31	26	-1.69	-2	*
29	0.99	27.25	27	-1.75	-2	*
30	1.06	28.17	28	-1.83	-2	*
31	1.13	29.08	29	-1.92	-2	*
32	1.20	30.00	30	-2.00	-2	*
33	1.27	30.91	31	-2.09	-2	*
34	1.34	31.83	32	-2.17	-2	*
35	1.40	32.74	33	-2.26	-2	*

36	1.46	33.65	34	-2.35	-2	*
37	1.52	34.55	35	-2.45	-2	*
38	1.59	35.46	35	-2.54	-3	*
39	1.65	36.36	36	-2.64	-3	*
40	1.70	37.25	37	-2.75	-3	*
41	1.76	38.15	38	-2.85	-3	*
42	1.82	39.07	39	-2.93	-3	*
43	1.88	39.97	40	-3.03	-3	*
44	1.93	40.87	41	-3.13	-3	*
45	1.99	41.78	42	-3.22	-3	*
46	2.04	42.69	43	-3.31	-3	*
47	2.10	43.61	44	-3.39	-3	*
48	2.16	44.54	45	-3.46	-3	*
49	2.21	45.46	45	-3.54	-4	*
50	2.27	46.40	46	-3.60	-4	*
51	2.32	47.37	47	-3.63	-4	*
52	2.38	48.32	48	-3.68	-4	*
53	2.44	49.31	49	-3.69	-4	*
54	2.50	50.30	50	-3.70	-4	*
55	2.55	51.29	51	-3.71	-4	*
56	2.61	52.31	52	-3.69	-4	*
57	2.68	53.35	53	-3.65	-4	*
58	2.74	54.40	54	-3.60	-4	*
59	2.80	55.47	55	-3.53	-4	*
60	2.87	56.55	57	-3.45	-3	*
61	2.93	57.64	58	-3.36	-3	*
62	3.00	58.75	59	-3.25	-3	*
63	3.07	59.86	60	-3.14	-3	*
64	3.15	61.00	61	-3.00	-3	*
65	3.22	62.14	62	-2.86	-3	*
66	3.30	63.29	63	-2.71	-3	*
67	3.39	64.46	64	-2.54	-3	*
68	3.48	65.61	66	-2.39	-2	*
69	3.57	66.78	67	-2.22	-2	*
70	3.67	67.95	68	-2.05	-2	*
71	3.78	69.13	69	-1.87	-2	*
72	3.90	70.29	70	-1.71	-2	*
73	4.02	71.47	71	-1.53	-2	*
74	4.16	72.63	73	-1.37	-1	*
75	4.32	73.80	74	-1.20	-1	*
76	4.50	74.96	75	-1.04	-1	*
77	4.72	76.13	76	-0.87	-1	*
78	4.97	77.29	77	-0.71	-1	*

79	5.30	78.45	78	-0.55	-1	*
80	5.75	79.62	80	-0.38	0	
81	6.48	80.80	81	-0.20	0	
82	8.00	81.72	82	-0.28	0	

Table D.7

Raw-Score Equivalents for Algebra 1 PBA (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.00	0	0.00	0	
1	-2.01	1.04	1	0.04	0	
2	-1.11	2.08	2	0.08	0	
3	-0.51	3.10	3	0.10	0	
4	-0.04	4.11	4	0.11	0	
5	0.33	5.11	5	0.11	0	
6	0.63	6.09	6	0.09	0	
7	0.87	7.05	7	0.05	0	
8	1.07	8.00	8	0.00	0	
9	1.25	8.94	9	-0.06	0	
10	1.40	9.88	10	-0.12	0	
11	1.54	10.81	11	-0.19	0	
12	1.67	11.75	12	-0.25	0	
13	1.79	12.69	13	-0.31	0	
14	1.91	13.64	14	-0.36	0	
15	2.02	14.59	15	-0.41	0	
16	2.13	15.56	16	-0.44	0	
17	2.25	16.53	17	-0.47	0	
18	2.36	17.53	18	-0.47	0	
19	2.47	18.53	19	-0.47	0	
20	2.58	19.56	20	-0.44	0	
21	2.70	20.60	21	-0.40	0	
22	2.81	21.65	22	-0.35	0	
23	2.93	22.70	23	-0.30	0	
24	3.05	23.76	24	-0.24	0	
25	3.16	24.83	25	-0.17	0	
26	3.28	25.89	26	-0.11	0	
27	3.41	26.96	27	-0.04	0	
28	3.53	28.03	28	0.03	0	
29	3.66	29.09	29	0.09	0	

30	3.80	30.15	30	0.15	0	
31	3.94	31.20	31	0.20	0	
32	4.11	32.22	32	0.22	0	
33	4.29	33.24	33	0.24	0	
34	4.52	34.23	34	0.23	0	
35	4.82	35.22	35	0.22	0	
36	5.24	36.18	36	0.18	0	
37	5.98	37.12	37	0.12	0	
38	8.00	37.88	38	-0.12	0	

Note: 1 task omitted because different score categories were observed in the TC and NTC conditions or because of calibration failure (e.g., due to zero proportion correct) in the TC or NTC condition.

Table D.8

Raw-Score Equivalents for Algebra 1 EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-3.10	0.93	1	-0.07	0	
2	-2.37	1.86	2	-0.14	0	
3	-1.92	2.80	3	-0.20	0	
4	-1.59	3.75	4	-0.25	0	
5	-1.33	4.70	5	-0.30	0	
6	-1.10	5.66	6	-0.34	0	
7	-0.90	6.63	7	-0.37	0	
8	-0.72	7.60	8	-0.40	0	
9	-0.56	8.57	9	-0.43	0	
10	-0.40	9.55	10	-0.45	0	
11	-0.26	10.54	11	-0.46	0	
12	-0.12	11.53	12	-0.47	0	
13	0.02	12.52	13	-0.48	0	
14	0.14	13.52	14	-0.48	0	
15	0.27	14.50	15	-0.50	0	
16	0.39	15.50	15	-0.50	-1	*
17	0.51	16.50	17	-0.50	0	
18	0.63	17.49	17	-0.51	-1	*
19	0.74	18.49	18	-0.51	-1	*
20	0.86	19.48	19	-0.52	-1	*
21	0.97	20.47	20	-0.53	-1	*
22	1.08	21.46	21	-0.54	-1	*

23	1.20	22.45	22	-0.55	-1	*
24	1.31	23.43	23	-0.57	-1	*
25	1.42	24.40	24	-0.60	-1	*
26	1.53	25.38	25	-0.62	-1	*
27	1.64	26.36	26	-0.64	-1	*
28	1.75	27.33	27	-0.67	-1	*
29	1.86	28.29	28	-0.71	-1	*
30	1.97	29.25	29	-0.75	-1	*
31	2.08	30.21	30	-0.79	-1	*
32	2.19	31.18	31	-0.82	-1	*
33	2.30	32.13	32	-0.87	-1	*
34	2.41	33.08	33	-0.92	-1	*
35	2.52	34.04	34	-0.96	-1	*
36	2.64	35.01	35	-0.99	-1	*
37	2.75	35.96	36	-1.04	-1	*
38	2.86	36.93	37	-1.07	-1	*
39	2.97	37.90	38	-1.10	-1	*
40	3.09	38.88	39	-1.12	-1	*
41	3.21	39.87	40	-1.13	-1	*
42	3.34	40.87	41	-1.13	-1	*
43	3.47	41.88	42	-1.12	-1	*
44	3.60	42.91	43	-1.09	-1	*
45	3.75	43.94	44	-1.06	-1	*
46	3.91	44.98	45	-1.02	-1	*
47	4.08	46.04	46	-0.96	-1	*
48	4.27	47.12	47	-0.88	-1	*
49	4.48	48.20	48	-0.80	-1	*
50	4.72	49.29	49	-0.71	-1	*
51	5.01	50.39	50	-0.61	-1	*
52	5.36	51.50	52	-0.50	0	
53	5.84	52.63	53	-0.37	0	
54	6.61	53.78	54	-0.22	0	
55	8.00	54.67	55	-0.33	0	

Table D.9

Raw-Score Equivalents for Algebra 1 PBA+EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-3.23	1.14	1	0.14	0	
2	-2.51	2.25	2	0.25	0	
3	-2.07	3.34	3	0.34	0	
4	-1.76	4.40	4	0.40	0	
5	-1.50	5.45	5	0.45	0	
6	-1.29	6.49	6	0.49	0	
7	-1.10	7.51	8	0.51	1	*
8	-0.94	8.52	9	0.52	1	*
9	-0.79	9.52	10	0.52	1	*
10	-0.65	10.53	11	0.53	1	*
11	-0.51	11.52	12	0.52	1	*
12	-0.39	12.52	13	0.52	1	*
13	-0.27	13.52	14	0.52	1	*
14	-0.16	14.52	15	0.52	1	*
15	-0.05	15.51	16	0.51	1	*
16	0.05	16.52	17	0.52	1	*
17	0.16	17.52	18	0.52	1	*
18	0.26	18.54	19	0.54	1	*
19	0.35	19.55	20	0.55	1	*
20	0.45	20.58	21	0.58	1	*
21	0.54	21.60	22	0.60	1	*
22	0.63	22.64	23	0.64	1	*
23	0.72	23.68	24	0.68	1	*
24	0.81	24.74	25	0.74	1	*
25	0.90	25.79	26	0.79	1	*
26	0.99	26.85	27	0.85	1	*
27	1.07	27.92	28	0.92	1	*
28	1.15	28.98	29	0.98	1	*
29	1.24	30.05	30	1.05	1	*
30	1.32	31.12	31	1.12	1	*
31	1.40	32.19	32	1.19	1	*
32	1.48	33.26	33	1.26	1	*
33	1.56	34.33	34	1.33	1	*
34	1.64	35.39	35	1.39	1	*
35	1.72	36.44	36	1.44	1	*

36	1.80	37.49	37	1.49	1	*
37	1.88	38.52	39	1.52	2	*
38	1.96	39.54	40	1.54	2	*
39	2.04	40.55	41	1.55	2	*
40	2.12	41.56	42	1.56	2	*
41	2.20	42.54	43	1.54	2	*
42	2.28	43.50	44	1.50	2	*
43	2.36	44.46	44	1.46	1	*
44	2.44	45.41	45	1.41	1	*
45	2.52	46.35	46	1.35	1	*
46	2.61	47.28	47	1.28	1	*
47	2.69	48.20	48	1.20	1	*
48	2.78	49.11	49	1.11	1	*
49	2.87	50.02	50	1.02	1	*
50	2.95	50.93	51	0.93	1	*
51	3.05	51.83	52	0.83	1	*
52	3.14	52.74	53	0.74	1	*
53	3.23	53.65	54	0.65	1	*
54	3.33	54.55	55	0.55	1	*
55	3.43	55.46	55	0.46	0	
56	3.54	56.37	56	0.37	0	
57	3.65	57.29	57	0.29	0	
58	3.76	58.22	58	0.22	0	
59	3.88	59.15	59	0.15	0	
60	4.01	60.09	60	0.09	0	
61	4.14	61.03	61	0.03	0	
62	4.29	61.99	62	-0.01	0	
63	4.45	62.95	63	-0.05	0	
64	4.62	63.93	64	-0.07	0	
65	4.82	64.92	65	-0.08	0	
66	5.05	65.93	66	-0.07	0	
67	5.33	66.95	67	-0.05	0	
68	5.67	67.97	68	-0.03	0	
69	6.14	69.01	69	0.01	0	
70	6.91	70.02	70	0.02	0	
71	8.00	70.66	71	-0.34	0	

Note: 6 tasks omitted because different score categories were observed in the TC and NTC conditions or because of calibration failure (e.g., due to zero proportion correct) in the TC or NTC condition.

Table D.10

Raw-Score Equivalents for Geometry PBA (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.00	0	0.00	0	
1	-2.13	0.87	1	-0.13	0	
2	-1.27	1.77	2	-0.23	0	
3	-0.68	2.68	3	-0.32	0	
4	-0.22	3.60	4	-0.40	0	
5	0.16	4.52	5	-0.48	0	
6	0.46	5.44	5	-0.56	-1	*
7	0.71	6.37	6	-0.63	-1	*
8	0.92	7.33	7	-0.67	-1	*
9	1.09	8.30	8	-0.70	-1	*
10	1.24	9.29	9	-0.71	-1	*
11	1.38	10.29	10	-0.71	-1	*
12	1.50	11.30	11	-0.70	-1	*
13	1.61	12.31	12	-0.69	-1	*
14	1.71	13.33	13	-0.67	-1	*
15	1.80	14.34	14	-0.66	-1	*
16	1.90	15.35	15	-0.65	-1	*
17	1.99	16.35	16	-0.65	-1	*
18	2.08	17.34	17	-0.66	-1	*
19	2.17	18.33	18	-0.67	-1	*
20	2.26	19.31	19	-0.69	-1	*
21	2.35	20.29	20	-0.71	-1	*
22	2.45	21.28	21	-0.72	-1	*
23	2.56	22.26	22	-0.74	-1	*
24	2.67	23.28	23	-0.72	-1	*
25	2.81	24.30	24	-0.70	-1	*
26	2.97	25.36	25	-0.64	-1	*
27	3.16	26.46	26	-0.54	-1	*
28	3.42	27.58	28	-0.42	0	
29	3.76	28.73	29	-0.27	0	
30	4.26	29.88	30	-0.12	0	
31	5.10	30.97	31	-0.03	0	
32	8.00	31.94	32	-0.06	0	

Note: 3 tasks omitted because different score categories were observed in the TC and NTC conditions or because of calibration failure (e.g., due to zero proportion correct) in the TC or NTC condition.

Table D.11

Raw-Score Equivalents for Geometry EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-3.11	0.95	1	-0.05	0	
2	-2.39	1.91	2	-0.09	0	
3	-1.95	2.88	3	-0.12	0	
4	-1.62	3.85	4	-0.15	0	
5	-1.37	4.82	5	-0.18	0	
6	-1.15	5.79	6	-0.21	0	
7	-0.96	6.76	7	-0.24	0	
8	-0.79	7.73	8	-0.27	0	
9	-0.63	8.70	9	-0.30	0	
10	-0.49	9.67	10	-0.33	0	
11	-0.35	10.65	11	-0.35	0	
12	-0.22	11.62	12	-0.38	0	
13	-0.10	12.60	13	-0.40	0	
14	0.02	13.58	14	-0.42	0	
15	0.14	14.56	15	-0.44	0	
16	0.25	15.55	16	-0.45	0	
17	0.36	16.55	17	-0.45	0	
18	0.47	17.55	18	-0.45	0	
19	0.58	18.55	19	-0.45	0	
20	0.69	19.56	20	-0.44	0	
21	0.80	20.57	21	-0.43	0	
22	0.90	21.58	22	-0.42	0	
23	1.01	22.59	23	-0.41	0	
24	1.12	23.60	24	-0.40	0	
25	1.22	24.61	25	-0.39	0	
26	1.32	25.62	26	-0.38	0	
27	1.43	26.63	27	-0.37	0	
28	1.53	27.64	28	-0.36	0	
29	1.64	28.65	29	-0.35	0	
30	1.74	29.65	30	-0.35	0	
31	1.85	30.66	31	-0.34	0	
32	1.95	31.66	32	-0.34	0	
33	2.06	32.66	33	-0.34	0	
34	2.16	33.65	34	-0.35	0	

35	2.27	34.65	35	-0.35	0	
36	2.38	35.64	36	-0.36	0	
37	2.49	36.63	37	-0.37	0	
38	2.61	37.60	38	-0.40	0	
39	2.72	38.58	39	-0.42	0	
40	2.84	39.55	40	-0.45	0	
41	2.97	40.53	41	-0.47	0	
42	3.09	41.50	41	-0.50	-1	*
43	3.23	42.46	42	-0.54	-1	*
44	3.37	43.42	43	-0.58	-1	*
45	3.52	44.39	44	-0.61	-1	*
46	3.67	45.35	45	-0.65	-1	*
47	3.84	46.32	46	-0.68	-1	*
48	4.03	47.29	47	-0.71	-1	*
49	4.24	48.28	48	-0.72	-1	*
50	4.48	49.28	49	-0.72	-1	*
51	4.76	50.30	50	-0.70	-1	*
52	5.10	51.35	51	-0.65	-1	*
53	5.56	52.44	52	-0.56	-1	*
54	6.31	53.63	54	-0.37	0	
55	8.00	54.71	55	-0.29	0	

Table D.12

Raw-Score Equivalents for Geometry PBA+EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-3.68	0.85	1	-0.15	0	
2	-2.94	1.73	2	-0.27	0	
3	-2.49	2.62	3	-0.38	0	
4	-2.16	3.52	4	-0.48	0	
5	-1.89	4.44	4	-0.56	-1	*
6	-1.67	5.36	5	-0.64	-1	*
7	-1.47	6.30	6	-0.70	-1	*
8	-1.29	7.25	7	-0.75	-1	*
9	-1.13	8.20	8	-0.80	-1	*
10	-0.99	9.16	9	-0.84	-1	*
11	-0.85	10.13	10	-0.87	-1	*
12	-0.72	11.10	11	-0.90	-1	*

13	-0.60	12.08	12	-0.92	-1	*
14	-0.49	13.05	13	-0.95	-1	*
15	-0.38	14.03	14	-0.97	-1	*
16	-0.27	15.02	15	-0.98	-1	*
17	-0.17	16.00	16	-1.00	-1	*
18	-0.08	16.97	17	-1.03	-1	*
19	0.02	17.94	18	-1.06	-1	*
20	0.10	18.90	19	-1.10	-1	*
21	0.19	19.86	20	-1.14	-1	*
22	0.27	20.82	21	-1.18	-1	*
23	0.35	21.76	22	-1.24	-1	*
24	0.43	22.70	23	-1.30	-1	*
25	0.50	23.63	24	-1.37	-1	*
26	0.57	24.56	25	-1.44	-1	*
27	0.64	25.50	26	-1.50	-1	*
28	0.71	26.44	26	-1.56	-2	*
29	0.78	27.37	27	-1.63	-2	*
30	0.85	28.32	28	-1.68	-2	*
31	0.91	29.27	29	-1.73	-2	*
32	0.98	30.22	30	-1.78	-2	*
33	1.04	31.18	31	-1.82	-2	*
34	1.11	32.14	32	-1.86	-2	*
35	1.17	33.10	33	-1.90	-2	*
36	1.23	34.08	34	-1.92	-2	*
37	1.30	35.07	35	-1.93	-2	*
38	1.36	36.04	36	-1.96	-2	*
39	1.43	37.04	37	-1.96	-2	*
40	1.49	38.04	38	-1.96	-2	*
41	1.56	39.04	39	-1.96	-2	*
42	1.63	40.05	40	-1.95	-2	*
43	1.70	41.08	41	-1.92	-2	*
44	1.77	42.10	42	-1.90	-2	*
45	1.85	43.13	43	-1.87	-2	*
46	1.92	44.17	44	-1.83	-2	*
47	2.00	45.22	45	-1.78	-2	*
48	2.09	46.27	46	-1.73	-2	*
49	2.17	47.33	47	-1.67	-2	*
50	2.26	48.40	48	-1.60	-2	*
51	2.36	49.47	49	-1.53	-2	*
52	2.46	50.55	51	-1.45	-1	*
53	2.56	51.63	52	-1.37	-1	*
54	2.67	52.73	53	-1.27	-1	*
55	2.78	53.82	54	-1.18	-1	*

56	2.90	54.91	55	-1.09	-1	*
57	3.03	56.02	56	-0.98	-1	*
58	3.17	57.12	57	-0.88	-1	*
59	3.32	58.22	58	-0.78	-1	*
60	3.48	59.32	59	-0.68	-1	*
61	3.66	60.42	60	-0.58	-1	*
62	3.85	61.50	62	-0.50	0	
63	4.06	62.58	63	-0.42	0	
64	4.31	63.65	64	-0.35	0	
65	4.60	64.72	65	-0.28	0	
66	4.95	65.79	66	-0.21	0	
67	5.42	66.85	67	-0.15	0	
68	6.18	67.92	68	-0.08	0	
69	8.00	68.81	69	-0.19	0	

Note: 7 tasks omitted because different score categories were observed in the TC and NTC conditions or because of calibration failure (e.g., due to zero proportion correct) in the TC or NTC condition.

Table D.13

Raw-Score Equivalents for Algebra 2 PBA (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.00	0	0.00	0	
1	-2.12	1.05	1	0.05	0	
2	-1.28	2.07	2	0.07	0	
3	-0.77	3.06	3	0.06	0	
4	-0.41	4.05	4	0.05	0	
5	-0.13	5.05	5	0.05	0	
6	0.10	6.07	6	0.07	0	
7	0.30	7.10	7	0.10	0	
8	0.48	8.13	8	0.13	0	
9	0.64	9.18	9	0.18	0	
10	0.79	10.22	10	0.22	0	
11	0.93	11.27	11	0.27	0	
12	1.06	12.32	12	0.32	0	
13	1.19	13.37	13	0.37	0	
14	1.31	14.42	14	0.42	0	
15	1.43	15.47	15	0.47	0	
16	1.55	16.51	17	0.51	1	*
17	1.67	17.53	18	0.53	1	*

18	1.78	18.54	19	0.54	1	*
19	1.90	19.54	20	0.54	1	*
20	2.00	20.51	21	0.51	1	*
21	2.10	21.45	21	0.45	0	
22	2.20	22.38	22	0.38	0	
23	2.30	23.27	23	0.27	0	
24	2.39	24.16	24	0.16	0	
25	2.48	25.04	25	0.04	0	
26	2.58	25.92	26	-0.08	0	
27	2.68	26.82	27	-0.18	0	
28	2.79	27.73	28	-0.27	0	
29	2.90	28.66	29	-0.34	0	
30	3.04	29.60	30	-0.40	0	
31	3.18	30.56	31	-0.44	0	
32	3.34	31.52	32	-0.48	0	
33	3.51	32.48	32	-0.52	-1	*
34	3.71	33.46	33	-0.54	-1	*
35	3.94	34.45	34	-0.55	-1	*
36	4.22	35.46	35	-0.54	-1	*
37	4.59	36.50	36	-0.50	-1	*
38	5.10	37.57	38	-0.43	0	
39	5.94	38.72	39	-0.28	0	
40	8.00	39.80	40	-0.20	0	

Note: 3 tasks omitted because different score categories were observed in the TC and NTC conditions or because of calibration failure (e.g., due to zero proportion correct) in the TC or NTC condition.

Table D.14
Raw-Score Equivalents for Algebra 2 EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-2.94	1.03	1	0.03	0	
2	-2.20	2.07	2	0.07	0	
3	-1.76	3.10	3	0.10	0	
4	-1.43	4.14	4	0.14	0	
5	-1.17	5.17	5	0.17	0	
6	-0.96	6.20	6	0.20	0	
7	-0.77	7.23	7	0.23	0	
8	-0.60	8.25	8	0.25	0	

9	-0.45	9.27	9	0.27	0	
10	-0.30	10.29	10	0.29	0	
11	-0.16	11.32	11	0.32	0	
12	-0.02	12.35	12	0.35	0	
13	0.11	13.38	13	0.38	0	
14	0.24	14.40	14	0.40	0	
15	0.37	15.44	15	0.44	0	
16	0.50	16.49	16	0.49	0	
17	0.63	17.53	18	0.53	1	*
18	0.76	18.58	19	0.58	1	*
19	0.89	19.63	20	0.63	1	*
20	1.01	20.68	21	0.68	1	*
21	1.14	21.74	22	0.74	1	*
22	1.26	22.81	23	0.81	1	*
23	1.38	23.86	24	0.86	1	*
24	1.49	24.92	25	0.92	1	*
25	1.61	25.97	26	0.97	1	*
26	1.72	27.02	27	1.02	1	*
27	1.83	28.06	28	1.06	1	*
28	1.94	29.10	29	1.10	1	*
29	2.05	30.11	30	1.11	1	*
30	2.16	31.12	31	1.12	1	*
31	2.27	32.11	32	1.11	1	*
32	2.38	33.10	33	1.10	1	*
33	2.50	34.06	34	1.06	1	*
34	2.62	35.03	35	1.03	1	*
35	2.74	35.98	36	0.98	1	*
36	2.87	36.92	37	0.92	1	*
37	3.01	37.85	38	0.85	1	*
38	3.16	38.77	39	0.77	1	*
39	3.32	39.69	40	0.69	1	*
40	3.50	40.60	41	0.60	1	*
41	3.70	41.51	42	0.51	1	*
42	3.93	42.41	42	0.41	0	
43	4.20	43.31	43	0.31	0	
44	4.52	44.20	44	0.20	0	
45	4.93	45.09	45	0.09	0	
46	5.47	46.00	46	0.00	0	
47	6.34	46.96	47	-0.04	0	
48	8.00	47.76	48	-0.24	0	

Note: 1 task omitted because different score categories were observed in the TC and NTC conditions or because of calibration failure (e.g., due to zero proportion correct) in the TC or NTC condition.

Table D.15

Raw-Score Equivalents for Algebra 2 PBA+EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-3.20	1.09	1	0.09	0	
2	-2.47	2.18	2	0.18	0	
3	-2.03	3.27	3	0.27	0	
4	-1.71	4.36	4	0.36	0	
5	-1.46	5.44	5	0.44	0	
6	-1.25	6.51	7	0.51	1	*
7	-1.07	7.57	8	0.57	1	*
8	-0.92	8.62	9	0.62	1	*
9	-0.78	9.67	10	0.67	1	*
10	-0.65	10.70	11	0.70	1	*
11	-0.54	11.71	12	0.71	1	*
12	-0.43	12.72	13	0.72	1	*
13	-0.33	13.72	14	0.72	1	*
14	-0.24	14.72	15	0.72	1	*
15	-0.15	15.71	16	0.71	1	*
16	-0.07	16.71	17	0.71	1	*
17	0.01	17.69	18	0.69	1	*
18	0.09	18.69	19	0.69	1	*
19	0.17	19.68	20	0.68	1	*
20	0.24	20.68	21	0.68	1	*
21	0.32	21.66	22	0.66	1	*
22	0.39	22.65	23	0.65	1	*
23	0.46	23.64	24	0.64	1	*
24	0.52	24.63	25	0.63	1	*
25	0.59	25.61	26	0.61	1	*
26	0.66	26.58	27	0.58	1	*
27	0.72	27.56	28	0.56	1	*
28	0.79	28.54	29	0.54	1	*
29	0.85	29.52	30	0.52	1	*
30	0.91	30.49	30	0.49	0	
31	0.97	31.48	31	0.48	0	
32	1.03	32.46	32	0.46	0	
33	1.09	33.45	33	0.45	0	

34	1.15	34.44	34	0.44	0	
35	1.21	35.44	35	0.44	0	
36	1.27	36.44	36	0.44	0	
37	1.33	37.46	37	0.46	0	
38	1.39	38.48	38	0.48	0	
39	1.45	39.51	40	0.51	1	*
40	1.52	40.56	41	0.56	1	*
41	1.58	41.61	42	0.61	1	*
42	1.64	42.67	43	0.67	1	*
43	1.70	43.72	44	0.72	1	*
44	1.76	44.79	45	0.79	1	*
45	1.82	45.87	46	0.87	1	*
46	1.89	46.94	47	0.94	1	*
47	1.95	48.01	48	1.01	1	*
48	2.01	49.08	49	1.08	1	*
49	2.08	50.15	50	1.15	1	*
50	2.14	51.21	51	1.21	1	*
51	2.21	52.27	52	1.27	1	*
52	2.27	53.32	53	1.32	1	*
53	2.34	54.36	54	1.36	1	*
54	2.41	55.39	55	1.39	1	*
55	2.47	56.41	56	1.41	1	*
56	2.54	57.43	57	1.43	1	*
57	2.61	58.41	58	1.41	1	*
58	2.68	59.40	59	1.40	1	*
59	2.76	60.39	60	1.39	1	*
60	2.83	61.36	61	1.36	1	*
61	2.91	62.32	62	1.32	1	*
62	2.98	63.27	63	1.27	1	*
63	3.06	64.20	64	1.20	1	*
64	3.15	65.14	65	1.14	1	*
65	3.24	66.07	66	1.07	1	*
66	3.33	66.99	67	0.99	1	*
67	3.42	67.90	68	0.90	1	*
68	3.52	68.81	69	0.81	1	*
69	3.63	69.71	70	0.71	1	*
70	3.74	70.60	71	0.60	1	*
71	3.86	71.48	71	0.48	0	
72	3.99	72.37	72	0.37	0	
73	4.13	73.25	73	0.25	0	
74	4.28	74.13	74	0.13	0	
75	4.45	75.01	75	0.01	0	
76	4.64	75.90	76	-0.10	0	

77	4.86	76.80	77	-0.20	0
78	5.12	77.72	78	-0.28	0
79	5.45	78.67	79	-0.33	0
80	5.90	79.67	80	-0.33	0
81	6.64	80.76	81	-0.24	0
82	8.00	81.65	82	-0.35	0

Note: 5 tasks omitted because different score categories were observed in the TC and NTC conditions or because of calibration failure (e.g., due to zero proportion correct) in the TC or NTC condition.

Table D.16

Raw-Score Equivalents for Grade 3 ELA/L PBA (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-2.84	0.99	1	-0.01	0	
2	-2.23	1.99	2	-0.01	0	
3	-1.88	2.98	3	-0.02	0	
4	-1.64	3.98	4	-0.02	0	
5	-1.45	4.98	5	-0.02	0	
6	-1.29	5.97	6	-0.03	0	
7	-1.15	6.97	7	-0.03	0	
8	-1.03	7.97	8	-0.03	0	
9	-0.92	8.98	9	-0.02	0	
10	-0.82	9.98	10	-0.02	0	
11	-0.72	10.97	11	-0.03	0	
12	-0.63	11.97	12	-0.03	0	
13	-0.55	12.97	13	-0.03	0	
14	-0.46	13.98	14	-0.02	0	
15	-0.38	14.97	15	-0.03	0	
16	-0.30	15.97	16	-0.03	0	
17	-0.22	16.97	17	-0.03	0	
18	-0.15	17.97	18	-0.03	0	
19	-0.07	18.97	19	-0.03	0	
20	0.01	19.96	20	-0.04	0	
21	0.09	20.97	21	-0.03	0	
22	0.17	21.97	22	-0.03	0	
23	0.25	22.96	23	-0.04	0	
24	0.33	23.95	24	-0.05	0	
25	0.41	24.95	25	-0.05	0	
26	0.50	25.95	26	-0.05	0	

27	0.59	26.95	27	-0.05	0	
28	0.69	27.95	28	-0.05	0	
29	0.79	28.94	29	-0.06	0	
30	0.90	29.94	30	-0.06	0	
31	1.02	30.94	31	-0.06	0	
32	1.14	31.95	32	-0.05	0	
33	1.28	32.95	33	-0.05	0	
34	1.44	33.95	34	-0.05	0	
35	1.61	34.96	35	-0.04	0	
36	1.81	35.96	36	-0.04	0	
37	2.04	36.97	37	-0.03	0	
38	2.31	37.97	38	-0.03	0	
39	2.63	38.98	39	-0.02	0	
40	3.03	39.98	40	-0.02	0	
41	3.57	40.99	41	-0.01	0	
42	4.42	42.01	42	0.01	0	
43	8.00	42.97	43	-0.03	0	

Table D.17
Raw-Score Equivalents for Grade 3 ELA/L EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-2.65	1.07	1	0.07	0	
2	-1.97	2.11	2	0.11	0	
3	-1.56	3.14	3	0.14	0	
4	-1.26	4.15	4	0.15	0	
5	-1.02	5.16	5	0.16	0	
6	-0.80	6.17	6	0.17	0	
7	-0.60	7.17	7	0.17	0	
8	-0.42	8.18	8	0.18	0	
9	-0.25	9.18	9	0.18	0	
10	-0.08	10.19	10	0.19	0	
11	0.09	11.18	11	0.18	0	
12	0.26	12.18	12	0.18	0	
13	0.43	13.17	13	0.17	0	
14	0.62	14.16	14	0.16	0	
15	0.81	15.15	15	0.15	0	
16	1.01	16.12	16	0.12	0	

17	1.24	17.09	17	0.09	0	
18	1.49	18.05	18	0.05	0	
19	1.79	19.01	19	0.01	0	
20	2.15	19.97	20	-0.03	0	
21	2.61	20.94	21	-0.06	0	
22	3.24	21.93	22	-0.07	0	
23	4.23	22.95	23	-0.05	0	
24	8.00	23.97	24	-0.03	0	

Table D.18

Raw-Score Equivalents for Grade 3 ELA/L PBA+EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-3.35	0.94	1	-0.06	0	
2	-2.70	1.87	2	-0.13	0	
3	-2.33	2.79	3	-0.21	0	
4	-2.07	3.70	4	-0.30	0	
5	-1.87	4.61	5	-0.39	0	
6	-1.70	5.52	6	-0.48	0	
7	-1.56	6.43	6	-0.57	-1	*
8	-1.44	7.35	7	-0.65	-1	*
9	-1.33	8.26	8	-0.74	-1	*
10	-1.24	9.18	9	-0.82	-1	*
11	-1.15	10.10	10	-0.90	-1	*
12	-1.06	11.03	11	-0.97	-1	*
13	-0.98	11.97	12	-1.03	-1	*
14	-0.91	12.90	13	-1.10	-1	*
15	-0.84	13.85	14	-1.15	-1	*
16	-0.77	14.79	15	-1.21	-1	*
17	-0.71	15.75	16	-1.25	-1	*
18	-0.64	16.70	17	-1.30	-1	*
19	-0.58	17.66	18	-1.34	-1	*
20	-0.52	18.63	19	-1.37	-1	*
21	-0.47	19.60	20	-1.40	-1	*
22	-0.41	20.58	21	-1.42	-1	*
23	-0.35	21.55	22	-1.45	-1	*
24	-0.30	22.53	23	-1.47	-1	*
25	-0.24	23.53	24	-1.47	-1	*

26	-0.19	24.51	25	-1.49	-1	*
27	-0.14	25.52	26	-1.48	-1	*
28	-0.08	26.51	27	-1.49	-1	*
29	-0.03	27.51	28	-1.49	-1	*
30	0.02	28.52	29	-1.48	-1	*
31	0.08	29.52	30	-1.48	-1	*
32	0.13	30.53	31	-1.47	-1	*
33	0.18	31.55	32	-1.45	-1	*
34	0.24	32.56	33	-1.44	-1	*
35	0.29	33.58	34	-1.42	-1	*
36	0.35	34.60	35	-1.40	-1	*
37	0.40	35.61	36	-1.39	-1	*
38	0.46	36.64	37	-1.36	-1	*
39	0.52	37.67	38	-1.33	-1	*
40	0.58	38.69	39	-1.31	-1	*
41	0.64	39.72	40	-1.28	-1	*
42	0.70	40.76	41	-1.24	-1	*
43	0.77	41.79	42	-1.21	-1	*
44	0.83	42.83	43	-1.17	-1	*
45	0.90	43.87	44	-1.13	-1	*
46	0.97	44.91	45	-1.09	-1	*
47	1.05	45.94	46	-1.06	-1	*
48	1.13	46.99	47	-1.01	-1	*
49	1.21	48.03	48	-0.97	-1	*
50	1.30	49.09	49	-0.91	-1	*
51	1.39	50.14	50	-0.86	-1	*
52	1.48	51.18	51	-0.82	-1	*
53	1.59	52.24	52	-0.76	-1	*
54	1.70	53.30	53	-0.70	-1	*
55	1.82	54.36	54	-0.64	-1	*
56	1.94	55.42	55	-0.58	-1	*
57	2.09	56.49	56	-0.51	-1	*
58	2.24	57.56	58	-0.44	0	
59	2.41	58.62	59	-0.38	0	
60	2.61	59.69	60	-0.31	0	
61	2.83	60.76	61	-0.24	0	
62	3.09	61.81	62	-0.19	0	
63	3.40	62.86	63	-0.14	0	
64	3.78	63.89	64	-0.11	0	
65	4.31	64.93	65	-0.07	0	
66	5.14	65.97	66	-0.03	0	
67	8.00	66.93	67	-0.07	0	

Table D.19

Raw-Score Equivalents for Grade 7 ELA/L PBA (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-2.89	1.05	1	0.05	0	
2	-2.28	2.10	2	0.10	0	
3	-1.93	3.15	3	0.15	0	
4	-1.69	4.20	4	0.20	0	
5	-1.50	5.26	5	0.26	0	
6	-1.34	6.32	6	0.32	0	
7	-1.20	7.38	7	0.38	0	
8	-1.08	8.44	8	0.44	0	
9	-0.98	9.49	9	0.49	0	
10	-0.88	10.55	11	0.55	1	*
11	-0.78	11.60	12	0.60	1	*
12	-0.70	12.66	13	0.66	1	*
13	-0.61	13.70	14	0.70	1	*
14	-0.53	14.75	15	0.75	1	*
15	-0.46	15.79	16	0.79	1	*
16	-0.38	16.83	17	0.83	1	*
17	-0.31	17.87	18	0.87	1	*
18	-0.24	18.91	19	0.91	1	*
19	-0.17	19.94	20	0.94	1	*
20	-0.10	20.96	21	0.96	1	*
21	-0.04	21.99	22	0.99	1	*
22	0.03	23.01	23	1.01	1	*
23	0.10	24.02	24	1.02	1	*
24	0.17	25.04	25	1.04	1	*
25	0.23	26.05	26	1.05	1	*
26	0.30	27.04	27	1.04	1	*
27	0.37	28.05	28	1.05	1	*
28	0.44	29.04	29	1.04	1	*
29	0.51	30.02	30	1.02	1	*
30	0.58	31.01	31	1.01	1	*
31	0.65	32.00	32	1.00	1	*
32	0.72	32.98	33	0.98	1	*
33	0.80	33.95	34	0.95	1	*
34	0.88	34.91	35	0.91	1	*

35	0.96	35.87	36	0.87	1	*
36	1.05	36.84	37	0.84	1	*
37	1.14	37.80	38	0.80	1	*
38	1.24	38.75	39	0.75	1	*
39	1.35	39.70	40	0.70	1	*
40	1.46	40.64	41	0.64	1	*
41	1.59	41.57	42	0.57	1	*
42	1.72	42.50	43	0.50	1	*
43	1.88	43.44	43	0.44	0	
44	2.07	44.36	44	0.36	0	
45	2.29	45.29	45	0.29	0	
46	2.58	46.21	46	0.21	0	
47	2.97	47.13	47	0.13	0	
48	3.65	48.06	48	0.06	0	
49	8.00	48.99	49	-0.01	0	

Table D.20
Raw-Score Equivalents for Grade 7 ELA/L EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-3.38	0.97	1	-0.03	0	
2	-2.67	1.97	2	-0.03	0	
3	-2.25	2.99	3	-0.01	0	
4	-1.95	4.01	4	0.01	0	
5	-1.72	5.04	5	0.04	0	
6	-1.52	6.06	6	0.06	0	
7	-1.36	7.09	7	0.09	0	
8	-1.22	8.11	8	0.11	0	
9	-1.09	9.14	9	0.14	0	
10	-0.97	10.16	10	0.16	0	
11	-0.86	11.18	11	0.18	0	
12	-0.75	12.22	12	0.22	0	
13	-0.66	13.23	13	0.23	0	
14	-0.56	14.25	14	0.25	0	
15	-0.47	15.27	15	0.27	0	
16	-0.39	16.28	16	0.28	0	
17	-0.30	17.31	17	0.31	0	
18	-0.22	18.32	18	0.32	0	

19	-0.14	19.33	19	0.33	0
20	-0.06	20.33	20	0.33	0
21	0.02	21.35	21	0.35	0
22	0.10	22.34	22	0.34	0
23	0.18	23.34	23	0.34	0
24	0.26	24.34	24	0.34	0
25	0.34	25.34	25	0.34	0
26	0.42	26.33	26	0.33	0
27	0.50	27.33	27	0.33	0
28	0.59	28.32	28	0.32	0
29	0.68	29.30	29	0.30	0
30	0.77	30.29	30	0.29	0
31	0.87	31.26	31	0.26	0
32	0.97	32.25	32	0.25	0
33	1.08	33.23	33	0.23	0
34	1.20	34.21	34	0.21	0
35	1.33	35.19	35	0.19	0
36	1.48	36.16	36	0.16	0
37	1.65	37.14	37	0.14	0
38	1.84	38.12	38	0.12	0
39	2.07	39.10	39	0.10	0
40	2.35	40.08	40	0.08	0
41	2.70	41.06	41	0.06	0
42	3.19	42.04	42	0.04	0
43	3.98	43.02	43	0.02	0
44	8.00	43.98	44	-0.02	0

Table D.21
Raw-Score Equivalents for Grade 7 ELA/L PBA+EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-3.81	0.94	1	-0.06	0	
2	-3.13	1.89	2	-0.11	0	
3	-2.74	2.86	3	-0.14	0	
4	-2.46	3.83	4	-0.17	0	
5	-2.25	4.81	5	-0.19	0	
6	-2.07	5.80	6	-0.20	0	
7	-1.93	6.79	7	-0.21	0	

8	-1.80	7.78	8	-0.22	0
9	-1.69	8.78	9	-0.22	0
10	-1.59	9.77	10	-0.23	0
11	-1.50	10.77	11	-0.23	0
12	-1.42	11.77	12	-0.23	0
13	-1.34	12.78	13	-0.22	0
14	-1.27	13.78	14	-0.22	0
15	-1.21	14.78	15	-0.22	0
16	-1.14	15.79	16	-0.21	0
17	-1.08	16.79	17	-0.21	0
18	-1.03	17.80	18	-0.20	0
19	-0.97	18.81	19	-0.19	0
20	-0.92	19.81	20	-0.19	0
21	-0.87	20.83	21	-0.17	0
22	-0.82	21.84	22	-0.16	0
23	-0.77	22.86	23	-0.14	0
24	-0.73	23.87	24	-0.13	0
25	-0.68	24.89	25	-0.11	0
26	-0.64	25.88	26	-0.12	0
27	-0.59	26.90	27	-0.10	0
28	-0.55	27.92	28	-0.08	0
29	-0.51	28.93	29	-0.07	0
30	-0.47	29.96	30	-0.04	0
31	-0.43	30.96	31	-0.04	0
32	-0.39	31.99	32	-0.01	0
33	-0.35	32.99	33	-0.01	0
34	-0.32	34.00	34	0.00	0
35	-0.28	35.02	35	0.02	0
36	-0.24	36.03	36	0.03	0
37	-0.20	37.04	37	0.04	0
38	-0.17	38.06	38	0.06	0
39	-0.13	39.06	39	0.06	0
40	-0.09	40.09	40	0.09	0
41	-0.06	41.10	41	0.10	0
42	-0.02	42.11	42	0.11	0
43	0.01	43.09	43	0.09	0
44	0.05	44.11	44	0.11	0
45	0.09	45.12	45	0.12	0
46	0.12	46.11	46	0.11	0
47	0.16	47.12	47	0.12	0
48	0.19	48.13	48	0.13	0
49	0.23	49.13	49	0.13	0
50	0.27	50.13	50	0.13	0

51	0.30	51.13	51	0.13	0
52	0.34	52.11	52	0.11	0
53	0.37	53.12	53	0.12	0
54	0.41	54.12	54	0.12	0
55	0.45	55.11	55	0.11	0
56	0.48	56.09	56	0.09	0
57	0.52	57.08	57	0.08	0
58	0.56	58.08	58	0.08	0
59	0.60	59.07	59	0.07	0
60	0.64	60.05	60	0.05	0
61	0.68	61.03	61	0.03	0
62	0.72	62.02	62	0.02	0
63	0.76	63.02	63	0.02	0
64	0.81	63.99	64	-0.01	0
65	0.85	64.99	65	-0.01	0
66	0.89	65.96	66	-0.04	0
67	0.94	66.95	67	-0.05	0
68	0.99	67.94	68	-0.06	0
69	1.03	68.91	69	-0.09	0
70	1.08	69.90	70	-0.10	0
71	1.14	70.87	71	-0.13	0
72	1.19	71.86	72	-0.14	0
73	1.24	72.83	73	-0.17	0
74	1.30	73.82	74	-0.18	0
75	1.36	74.80	75	-0.20	0
76	1.43	75.79	76	-0.21	0
77	1.49	76.77	77	-0.23	0
78	1.57	77.76	78	-0.24	0
79	1.64	78.75	79	-0.25	0
80	1.72	79.75	80	-0.25	0
81	1.81	80.74	81	-0.26	0
82	1.91	81.74	82	-0.26	0
83	2.01	82.74	83	-0.26	0
84	2.12	83.75	84	-0.25	0
85	2.25	84.77	85	-0.23	0
86	2.40	85.79	86	-0.21	0
87	2.57	86.81	87	-0.19	0
88	2.77	87.84	88	-0.16	0
89	3.01	88.87	89	-0.13	0
90	3.33	89.91	90	-0.09	0
91	3.77	90.94	91	-0.06	0
92	4.50	91.97	92	-0.03	0
93	8.00	92.97	93	-0.03	0

Table D.22

Raw-Score Equivalents for Grade 9 ELA/L PBA (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-3.34	1.12	1	0.12	0	
2	-2.70	2.23	2	0.23	0	
3	-2.34	3.34	3	0.34	0	
4	-2.08	4.44	4	0.44	0	
5	-1.88	5.54	6	0.54	1	*
6	-1.71	6.64	7	0.64	1	*
7	-1.56	7.73	8	0.73	1	*
8	-1.43	8.83	9	0.83	1	*
9	-1.32	9.91	10	0.91	1	*
10	-1.21	10.98	11	0.98	1	*
11	-1.11	12.07	12	1.07	1	*
12	-1.01	13.14	13	1.14	1	*
13	-0.92	14.20	14	1.20	1	*
14	-0.83	15.25	15	1.25	1	*
15	-0.75	16.30	16	1.30	1	*
16	-0.67	17.34	17	1.34	1	*
17	-0.59	18.39	18	1.39	1	*
18	-0.51	19.41	19	1.41	1	*
19	-0.44	20.43	20	1.43	1	*
20	-0.36	21.45	21	1.45	1	*
21	-0.29	22.46	22	1.46	1	*
22	-0.21	23.47	23	1.47	1	*
23	-0.14	24.46	24	1.46	1	*
24	-0.07	25.44	25	1.44	1	*
25	0.01	26.42	26	1.42	1	*
26	0.08	27.40	27	1.40	1	*
27	0.15	28.38	28	1.38	1	*
28	0.23	29.34	29	1.34	1	*
29	0.31	30.30	30	1.30	1	*
30	0.38	31.25	31	1.25	1	*
31	0.46	32.20	32	1.20	1	*
32	0.54	33.16	33	1.16	1	*
33	0.63	34.10	34	1.10	1	*
34	0.71	35.04	35	1.04	1	*
35	0.81	35.99	36	0.99	1	*

36	0.90	36.93	37	0.93	1	*
37	1.00	37.86	38	0.86	1	*
38	1.11	38.80	39	0.80	1	*
39	1.22	39.74	40	0.74	1	*
40	1.34	40.67	41	0.67	1	*
41	1.47	41.61	42	0.61	1	*
42	1.61	42.55	43	0.55	1	*
43	1.76	43.48	43	0.48	0	
44	1.93	44.42	44	0.42	0	
45	2.12	45.36	45	0.36	0	
46	2.34	46.30	46	0.30	0	
47	2.61	47.24	47	0.24	0	
48	2.93	48.18	48	0.18	0	
49	3.37	49.11	49	0.11	0	
50	4.10	50.05	50	0.05	0	
51	8.00	50.98	51	-0.02	0	

Table D.23
Raw-Score Equivalents for Grade 9 ELA/L EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.00	0	0.00	0	
1	-2.53	1.03	1	0.03	0	
2	-1.92	2.08	2	0.08	0	
3	-1.57	3.14	3	0.14	0	
4	-1.33	4.21	4	0.21	0	
5	-1.14	5.27	5	0.27	0	
6	-0.98	6.34	6	0.34	0	
7	-0.85	7.39	7	0.39	0	
8	-0.73	8.45	8	0.45	0	
9	-0.62	9.50	9	0.50	0	
10	-0.52	10.54	11	0.54	1	*
11	-0.42	11.59	12	0.59	1	*
12	-0.33	12.62	13	0.62	1	*
13	-0.25	13.66	14	0.66	1	*
14	-0.17	14.69	15	0.69	1	*
15	-0.09	15.71	16	0.71	1	*
16	-0.01	16.74	17	0.74	1	*
17	0.06	17.76	18	0.76	1	*

18	0.14	18.77	19	0.77	1	*
19	0.21	19.79	20	0.79	1	*
20	0.28	20.81	21	0.81	1	*
21	0.35	21.81	22	0.81	1	*
22	0.43	22.82	23	0.82	1	*
23	0.50	23.83	24	0.83	1	*
24	0.57	24.82	25	0.82	1	*
25	0.65	25.82	26	0.82	1	*
26	0.72	26.81	27	0.81	1	*
27	0.80	27.80	28	0.80	1	*
28	0.88	28.79	29	0.79	1	*
29	0.97	29.78	30	0.78	1	*
30	1.05	30.75	31	0.75	1	*
31	1.14	31.72	32	0.72	1	*
32	1.24	32.69	33	0.69	1	*
33	1.34	33.65	34	0.65	1	*
34	1.45	34.62	35	0.62	1	*
35	1.56	35.57	36	0.57	1	*
36	1.69	36.52	37	0.52	1	*
37	1.84	37.46	37	0.46	0	
38	2.00	38.41	38	0.41	0	
39	2.19	39.34	39	0.34	0	
40	2.42	40.28	40	0.28	0	
41	2.72	41.21	41	0.21	0	
42	3.13	42.14	42	0.14	0	
43	3.83	43.07	43	0.07	0	
44	8.00	43.99	44	-0.01	0	

Table D.24
Raw-Score Equivalents for Grade 9 ELA/L PBA+EOY (Matched on Demographics Only)

NTC Raw Score	θ	TC Exp. Raw Score	TC Exp. Raw Score (Rounded)	Diff.	Rounded Diff.	Device Effect
0	-8.00	0.01	0	0.01	0	
1	-3.61	1.17	1	0.17	0	
2	-2.96	2.35	2	0.35	0	
3	-2.59	3.55	4	0.55	1	*
4	-2.33	4.76	5	0.76	1	*
5	-2.13	5.96	6	0.96	1	*
6	-1.97	7.16	7	1.16	1	*

7	-1.83	8.37	8	1.37	1	*
8	-1.71	9.56	10	1.56	2	*
9	-1.61	10.75	11	1.75	2	*
10	-1.51	11.93	12	1.93	2	*
11	-1.43	13.11	13	2.11	2	*
12	-1.35	14.27	14	2.27	2	*
13	-1.27	15.45	15	2.45	2	*
14	-1.20	16.59	17	2.59	3	*
15	-1.14	17.75	18	2.75	3	*
16	-1.08	18.88	19	2.88	3	*
17	-1.02	20.01	20	3.01	3	*
18	-0.96	21.14	21	3.14	3	*
19	-0.90	22.27	22	3.27	3	*
20	-0.85	23.37	23	3.37	3	*
21	-0.80	24.48	24	3.48	3	*
22	-0.75	25.58	26	3.58	4	*
23	-0.70	26.67	27	3.67	4	*
24	-0.66	27.75	28	3.75	4	*
25	-0.61	28.85	29	3.85	4	*
26	-0.57	29.92	30	3.92	4	*
27	-0.52	31.01	31	4.01	4	*
28	-0.48	32.06	32	4.06	4	*
29	-0.44	33.12	33	4.12	4	*
30	-0.40	34.18	34	4.18	4	*
31	-0.36	35.24	35	4.24	4	*
32	-0.32	36.29	36	4.29	4	*
33	-0.28	37.32	37	4.32	4	*
34	-0.24	38.36	38	4.36	4	*
35	-0.20	39.41	39	4.41	4	*
36	-0.16	40.43	40	4.43	4	*
37	-0.13	41.46	41	4.46	4	*
38	-0.09	42.49	42	4.49	4	*
39	-0.05	43.49	43	4.49	4	*
40	-0.01	44.53	45	4.53	5	*
41	0.03	45.53	46	4.53	5	*
42	0.06	46.54	47	4.54	5	*
43	0.10	47.51	48	4.51	5	*
44	0.14	48.51	49	4.51	5	*
45	0.17	49.51	50	4.51	5	*
46	0.21	50.50	51	4.50	5	*
47	0.25	51.49	51	4.49	4	*
48	0.28	52.47	52	4.47	4	*
49	0.32	53.44	53	4.44	4	*

50	0.36	54.41	54	4.41	4	*
51	0.39	55.36	55	4.36	4	*
52	0.43	56.33	56	4.33	4	*
53	0.47	57.27	57	4.27	4	*
54	0.51	58.25	58	4.25	4	*
55	0.55	59.19	59	4.19	4	*
56	0.59	60.14	60	4.14	4	*
57	0.62	61.08	61	4.08	4	*
58	0.66	62.02	62	4.02	4	*
59	0.70	62.95	63	3.95	4	*
60	0.75	63.89	64	3.89	4	*
61	0.79	64.81	65	3.81	4	*
62	0.83	65.74	66	3.74	4	*
63	0.87	66.66	67	3.66	4	*
64	0.91	67.57	68	3.57	4	*
65	0.96	68.50	68	3.50	3	*
66	1.00	69.40	69	3.40	3	*
67	1.05	70.33	70	3.33	3	*
68	1.10	71.24	71	3.24	3	*
69	1.15	72.15	72	3.15	3	*
70	1.20	73.04	73	3.04	3	*
71	1.25	73.95	74	2.95	3	*
72	1.30	74.85	75	2.85	3	*
73	1.36	75.75	76	2.75	3	*
74	1.42	76.65	77	2.65	3	*
75	1.48	77.53	78	2.53	3	*
76	1.54	78.43	78	2.43	2	*
77	1.60	79.32	79	2.32	2	*
78	1.67	80.21	80	2.21	2	*
79	1.74	81.09	81	2.09	2	*
80	1.82	81.98	82	1.98	2	*
81	1.90	82.87	83	1.87	2	*
82	1.99	83.75	84	1.75	2	*
83	2.08	84.63	85	1.63	2	*
84	2.18	85.51	86	1.51	2	*
85	2.29	86.39	86	1.39	1	*
86	2.40	87.26	87	1.26	1	*
87	2.53	88.13	88	1.13	1	*
88	2.68	89.00	89	1.00	1	*
89	2.85	89.86	90	0.86	1	*
90	3.04	90.72	91	0.72	1	*
91	3.28	91.58	92	0.58	1	*
92	3.58	92.44	92	0.44	0	

93	4.00	93.29	93	0.29	0
94	4.71	94.15	94	0.15	0
95	8.00	94.97	95	-0.03	0