Washington State
Institute for
Public Policy

110 Fifth Avenue Southeast, Suite 214 • PO Box 40999 • Olympia, WA 98504-0999 • (360) 586-2677 • www.wsipp.wa.gov

August 2007

TENTH-GRADE WASL IN SPRING 2006: PERFORMANCE ON MULTIPLE-CHOICE AND OPEN-ENDED QUESTIONS BY STUDENT SUBGROUPS

The 2006 Legislature directed the Washington State Institute for Public Policy (Institute) to conduct a "review and statistical analysis of Washington assessment of student learning [WASL] data" in order to "identify possible barriers to student success or possible causes of the lack of success."[1]

In a previous report, the Institute described the association between student performance on multiple-choice and open-ended questions on the 10th-grade WASL in spring 2006.[2] We found that open-ended and multiple-choice scores are strongly associated, especially for math.

This report examines the relationship between student performance on multiple-choice and open-ended questions on the 10th-grade WASL in spring 2006 by gender, race/ethnicity, language use, poverty status, and enrollment in special education.

To determine whether a particular question format poses a barrier to success for different categories of students, we analyze "raw" scores for multiple-choice and open-ended items on the reading and math assessments. Raw scores represent the number of multiple-choice questions answered correctly or the number of points awarded to open-ended questions.

- ***Multiple-choice questions*** require students to select one answer from a set of possible answers. These items are machine scored.
- ***Open-ended questions*** require students to construct a response. Open-ended responses are assessed by teams of scorers and may be awarded partial credit: short-answer questions are worth 2 points and extended-response questions are worth 4 points.[3]

SUMMARY

The WASL includes two kinds of questions: multiple-choice and open-ended items. Although it is more time-consuming and costly to score a test that includes open-ended questions, it is often argued that both open-ended and multiple-choice questions are needed, because they assess different kinds of skills.

In a previous report, the Institute found that open-ended and multiple-choice scores are highly correlated, especially for math.

This report extends that analysis by examining the relationship between student performance on multiple-choice and open-ended questions on the 10th-grade WASL in spring 2006 by gender, race/ethnicity, language use, poverty status, and enrollment in special education. In particular, we analyze the share of total points earned on multiple-choice questions for students with these characteristics at different levels of WASL performance.

Overall Results

- Most students—all but those in the lowest 25 percentile—showed little or no difference in their performance on multiple-choice and open-ended questions in either reading or math.
- Those in the bottom 25 percent did better on multiple-choice questions than they did on open-ended items, especially in math.
- Students enrolled in special education tended to earn a larger share of their points from multiple-choice questions, especially in math.

Conclusion

Most students performed equally well on both the open-ended and multiple-choice item formats.

[1] SSB 6618, Chapter 352, Laws of 2006.

[2] R. Barnoski & W. Cole. (2006). *Tenth-grade WASL in spring 2006: Open-ended and multiple-choice questions*. Olympia: Washington State Institute for Public Policy, Document No. 06-11-2206.

[3] https://www.k12.wa.us/assessment/pubdocs/ScoringtheWASL_FAQ100406.pdf

Our purpose is to determine whether some groups of students consistently performed better on multiple-choice or open-ended questions. The research on this topic is extensive, but the results are mixed.[4]

METHOD

To examine student performance by item format, we first divided each student group into *quartiles*: four roughly equal parts based on the distribution of total raw scores (i.e., the sum of multiple-choice and open-ended scores). The **lowest-performing** students earned scores below the 25th percentile for their group. The **highest-performing** students earned scores greater than the 75th percentile for their group. The middle range of scores was divided at the median (25th to 50th percentile and 50th to 75th percentile).

Then, for each performance level, we examined the percentage of total points earned from multiple-choice items. We examined whether some groups of students obtained a larger share of their overall WASL score from multiple-choice or open-ended items, controlling for total performance level.

We also considered an alternative way to examine performance: the share of points derived from multiple-choice items plotted against scale scores in reading and math.[5] This method enabled us to ascertain the relationship between performance on multiple-choice items and meeting standard on the WASL.[6]

[4] For example: M. Beller & N. Gafni. (2000). Can item format (multiple choice vs. open ended) account for gender differences in mathematics achievement? *Sex Roles* 42(1/2): 1-21; M. Pomplun & N. Sundbye. (1999). Gender differences in constructed response reading items. *Applied Measurement in Education* 12(1): 95-109; L.S. Hamilton. (2000). Detecting gender-based differential item functioning on a constructed-response science test. *Applied Measurement in Education* 12(3): 211-235; M. Garner & G. Engelhard, Jr. (1999). Gender differences in performance on multiple-choice and constructed response mathematics items. *Applied Measurement in Education* 12(1): 29-51; R. Lukhele, D. Thissen & H. Wainer. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement* 31(3): 234-250; N.S. Cole. (1997). *The ETS gender study: How females and males perform in educational settings*. Princeton, NJ: Educational Testing Service.

[5] Scale scores are derived from a statistical method called Rasch modeling, which places students on a common metric that allows scores to be equated from one year to the next.

[6] A method for examining differential test functioning in greater detail, SIBTEST, is available but requires item-level data (i.e., information on how students performed on each question). For this report, the Institute had access to subject-area raw and scale scores for multiple-choice and open-ended item formats. A consultant affiliated with the IRT Modeling Lab at the University of Illinois advised the Institute that a differential test functioning analysis would cost between $15,000 and $50,000 depending on the number of student characteristics, subject-area assessments, and years of data included in the analysis.

SUMMARY OF FINDINGS

Most groups of students showed little or no difference in their performance on multiple-choice and open-ended questions.

- For all groups of students (except those enrolled in special education) above the 25th percentile in either reading or math, performance on multiple-choice and open-ended items was *nearly equivalent*. That is, the share of points earned from multiple-choice items was roughly equivalent for most students regardless of demographic characteristics.[7]

Most groups of students who scored in the bottom quartile did better on multiple-choice questions than they did on open-ended items, especially in math.

- **Reading**: The lowest-performing males, White students, and non-Asian minorities earned a slightly larger share of their points from multiple-choice questions than did the lowest-performing females and Asian students.[8]

- **Math**: Low-performing "disadvantaged" students—non-Asian minorities, students in poverty, or students with language barriers—earned a larger share of their points from multiple-choice questions than did their low-performing "non-disadvantaged" peers.

Students enrolled in special education tended to earn a larger share of their points from multiple-choice questions, especially in math.

We conclude that most students performed equally well on both the open-ended and multiple-choice item formats.

[7] We note that it is possible to find *items* that display differential functioning for student subgroups in the absence of differential *test* functioning among subgroups. That is, differences on individual test questions may not add up to differences on the test as a whole.

[8] Performance on open-ended reading items is closely related to performance on the writing WASL for some groups of students. Open-ended performance, for example, was equivalent for students who met standard in writing regardless of language barriers or poverty. These results are available upon request.

TECHNICAL APPENDIX:
PERFORMANCE ON MULTIPLE-CHOICE QUESTIONS BY STUDENT SUBGROUPS

This technical appendix presents results from our analysis of performance on multiple-choice questions by student subgroups. To demonstrate how to interpret our findings, we discuss the results for gender using two methods: the percentage of points derived from multiple-choice items by (1) raw score quartiles and (2) scale scores.

RAW SCORE QUARTILES

Exhibit 1 depicts the percentage of points from multiple-choice items by total raw score quartiles for male and female students. The dashed horizontal lines indicate that the percentage of points earned from multiple-choice items is proportional to the percentage of multiple-choice points available on the reading and math assessments.[a]

Reading

Raw reading scores vary between 0 and 52 points (26 points each from multiple-choice items and open-ended items). Twenty-five percent of males and females earned fewer than 39 and 41 points in reading, respectively, whereas the highest-performing 25 percent of males and females earned scores of 47 and 48 or higher.

Quartile Ranges for Reading by Gender

	<25th	25th–50th	50th–75th	>75th
Males	0–38	39–43	44–46	47–52
Females	0–40	41–44	45–47	48–52

Males earned a larger share of their points from multiple-choice items than did females, although the differences are negligible for 75 percent of students (i.e., those with total raw scores above the 25th percentile). Females received a slightly disproportionate share of their points from open-ended items, whereas the lowest-performing males derived a small majority of their points from multiple-choice items.

Math

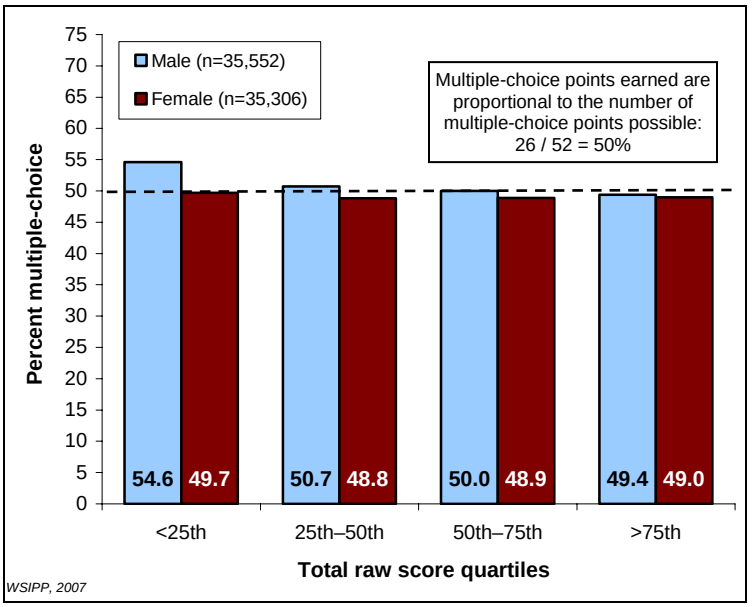
Total raw scores in math range from 1 to 65 points (27 points from multiple-choice items and 38 points from open-ended items). Twenty-five percent of males and females earned fewer than 31 points in math; another 25 percent earned at least 51 and 50 points, respectively.

Quartile Ranges for Math by Gender

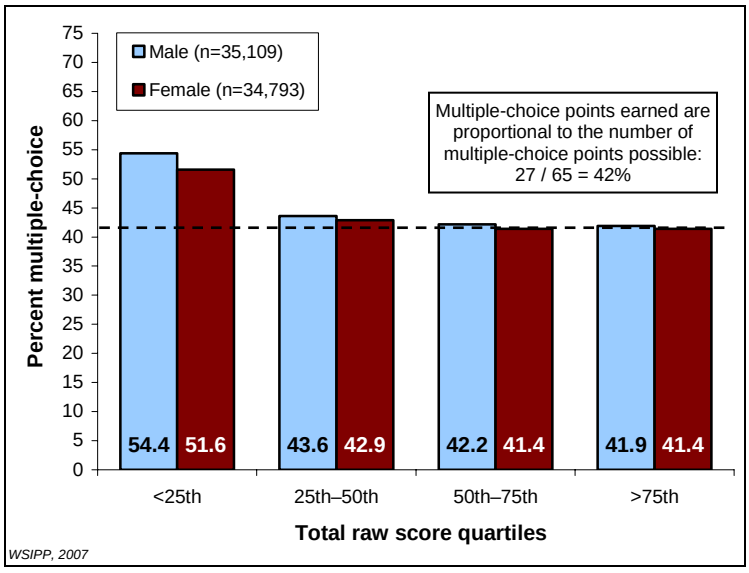
	<25th	25th–50th	50th–75th	>75th
Males	0–30	31–42	43–50	51–65
Females	0–30	31–40	41–49	50–65

Exhibit 1
Most Males and Females Earn About the Same Share of Their Points From Multiple-Choice Items

A. Reading



B. Math



As with performance in reading, the lowest-performing males earned a slightly higher percentage of their points from multiple-choice items than did similarly performing females. Unlike the pattern for reading, however, both low-performing males and females earned a disproportionate share of their points—54.4 and 51.6 percent, respectively—from multiple-choice items. Nevertheless, 75 percent of males and females performed equally well on multiple-choice and open-ended items.

[a] In reading, both multiple-choice and open-ended scores range from 0 to 26, so that scores are distributed proportionally when 50 percent of points are earned from multiple-choice questions (26 / 52 = 50%). In math, multiple-choice scores range from 1 to 27 and open-ended scores vary between 0 and 38; therefore, scores are proportional when 42 percent of points are earned from multiple-choice questions (27 / 65 = 42%).

SCALE SCORES

Exhibit 2 depicts the percentage of points from multiple-choice items plotted against scale scores for males and females. The graphs convey two pieces of information: (1) the percentage of male and female students' total raw scores derived from multiple-choice items, and (2) the cumulative distribution of male and female students across scale scores.

To facilitate interpretation, the charts include two benchmarks. First, points falling on the dashed *horizontal* lines indicate that total raw scores are distributed proportionally between multiple-choice and open-ended items (i.e., 50 percent for reading and 42 percent for math). Second, the dashed *vertical* lines represent the cut points for meeting standard in reading and math (i.e., scale score = 400). Points falling to the right of the vertical lines represent students who met standard.

Reading

The points at which the S-shaped trend lines for the cumulative percentage of male and females intersect with the dashed vertical reference line gives the percentage of students who did not meet standard in reading: 11.8 and 16.8 percent, respectively, for females and males. Among these students, males earned a slightly higher percentage of their total points from multiple-choice items than did females. This gap in performance on multiple-choice items diminished and ultimately disappeared for most students who met standard in reading. Put differently, most males and females who met standard in reading performed equally well on multiple-choice and open-ended items.

Note the presence of a few extreme outliers—points on the graph that fall considerably above or below the general trends for performance on multiple-choice questions. These data points are based on a very small number of students and are therefore particularly sensitive to the slightest variation in performance by item format.

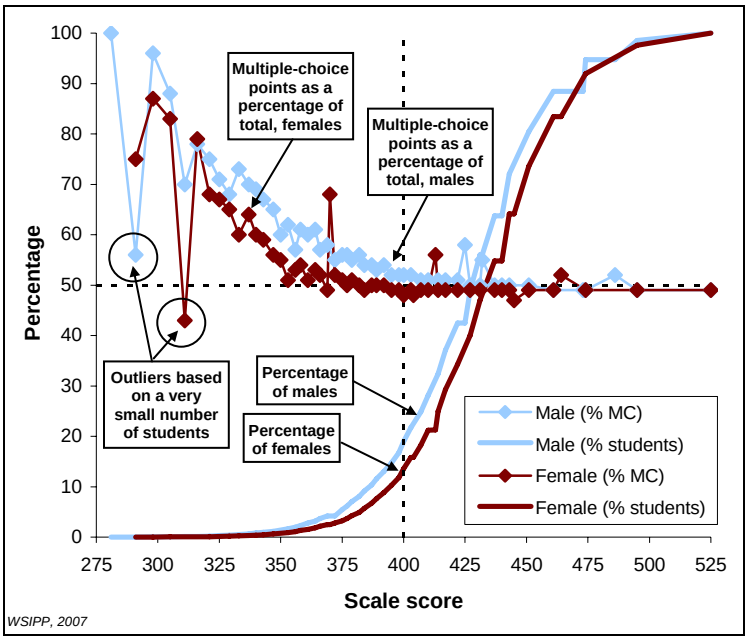
Math

In math, the proportion of points earned from multiple-choice questions was nearly identical for male and female students across the entire distribution of scale scores. That is, performance on multiple-choice and open-ended items was similar for males and females regardless of whether they met standard in math. Moreover, among students who met standard in math—roughly half of males and females—total raw scores were distributed proportionally between multiple-choice and open-ended items.

The remainder of this appendix presents the results of our analysis of performance by item format for the following subgroups: students in and not in poverty; students with and without English-language barriers; students enrolled and not enrolled in special education; and students from different racial/ethnic backgrounds.

Exhibit 2
The Share of Points From Multiple-Choice Items Was Nearly Equivalent for Males and Females Who Met Standard

A. Reading



B. Math

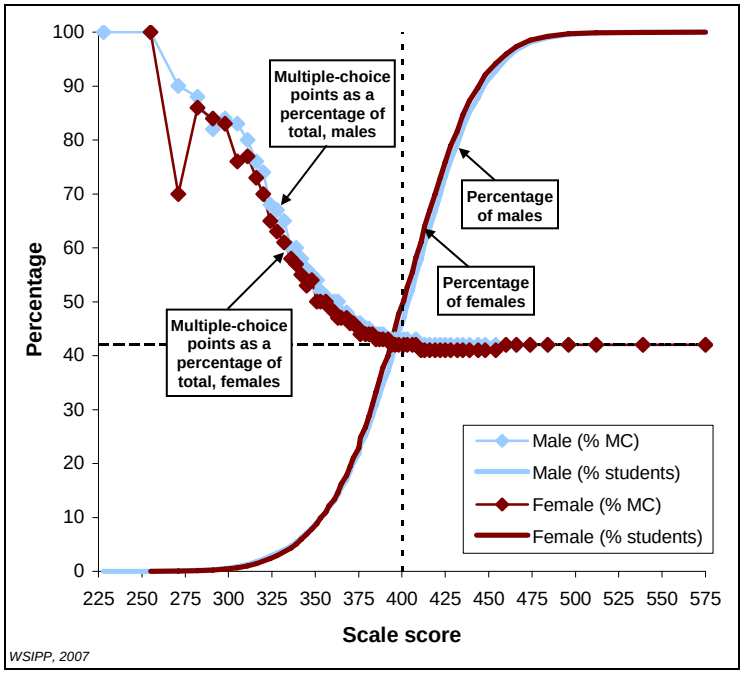
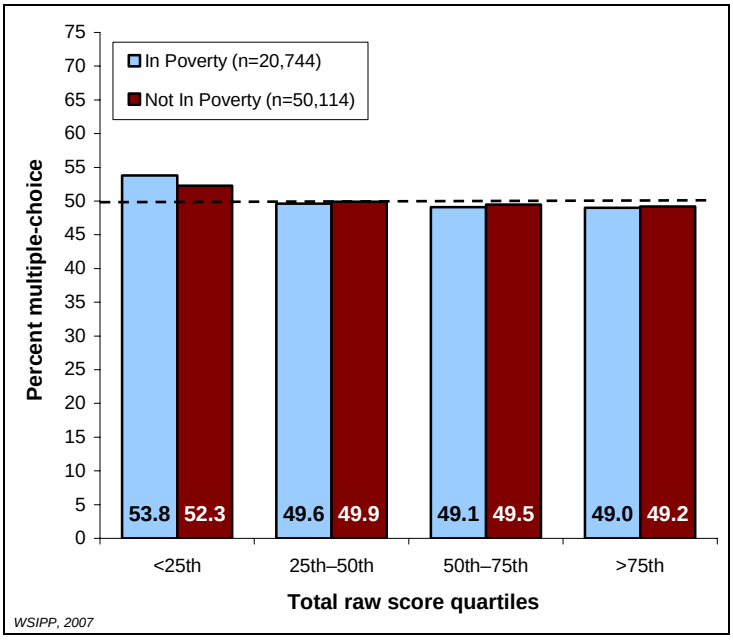


Exhibit 3
Percentage of Points From Multiple-Choice Items by Total Raw Score Quartiles

A. Reading



B. Math

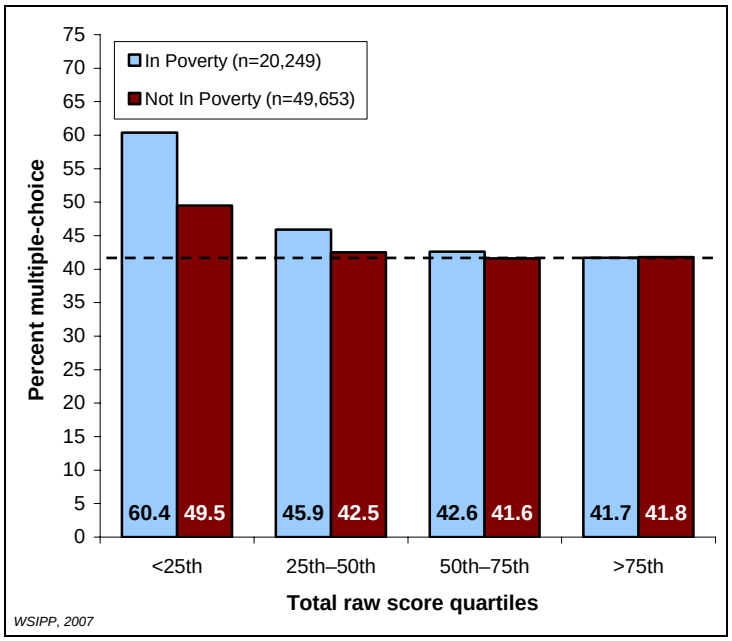
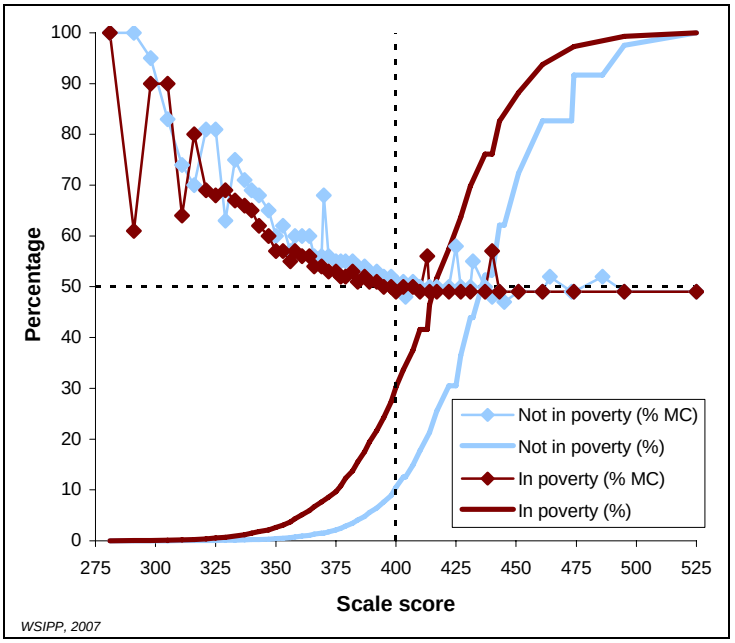


Exhibit 4
Percentage of Points From Multiple-Choice Items by Scale Scores

A. Reading



B. Math

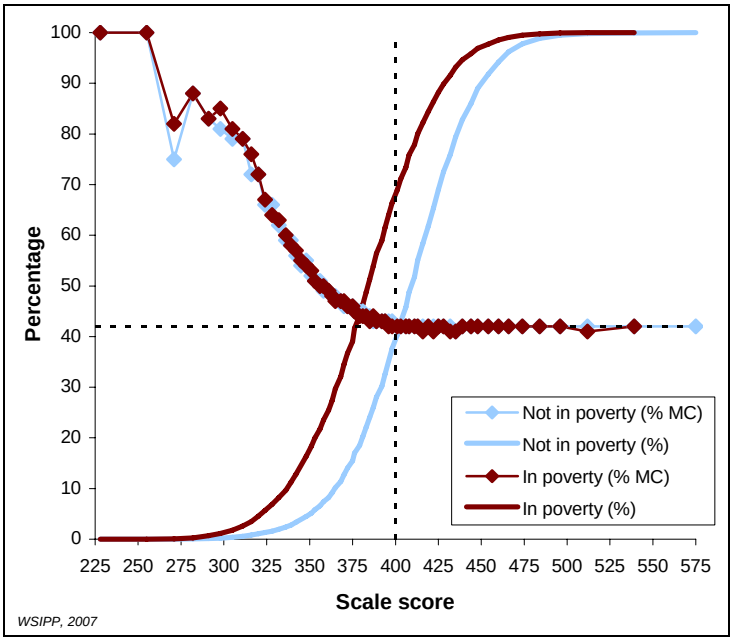
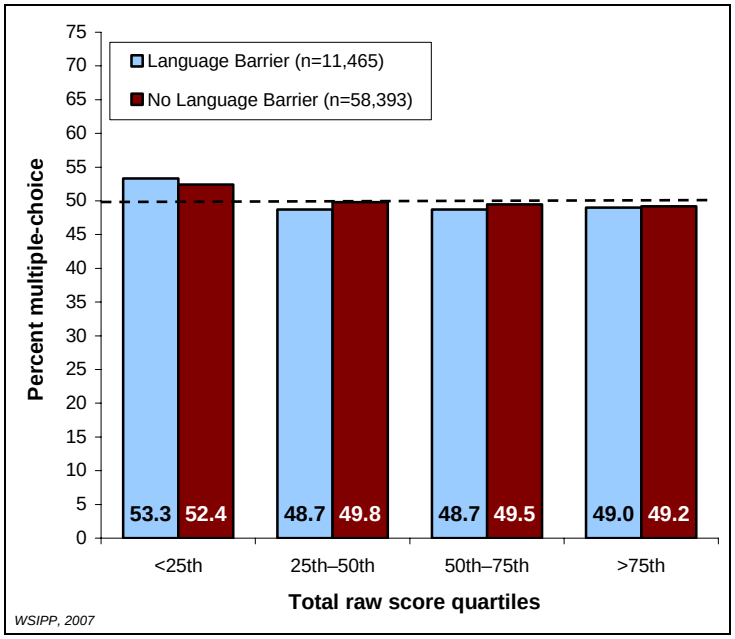


Exhibit 5
Percentage of Points From Multiple-Choice Items by Total Raw Score Quartiles

A. Reading



B. Math

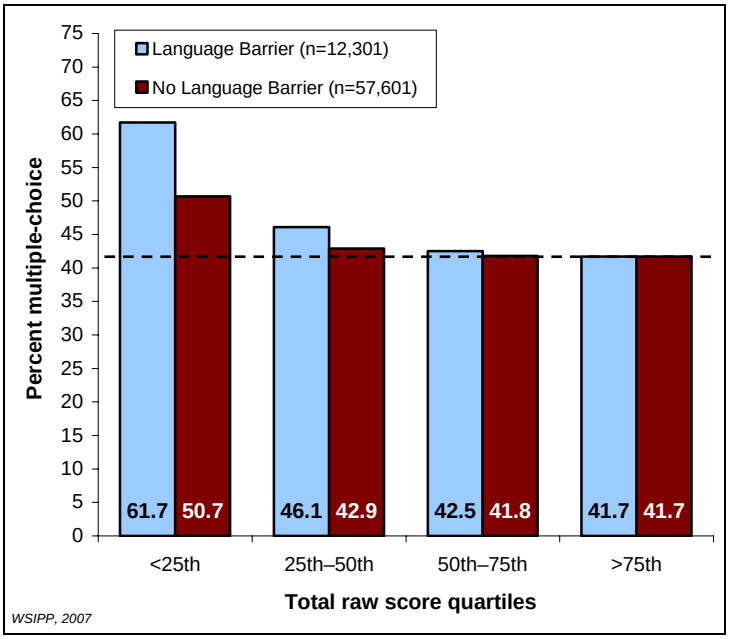
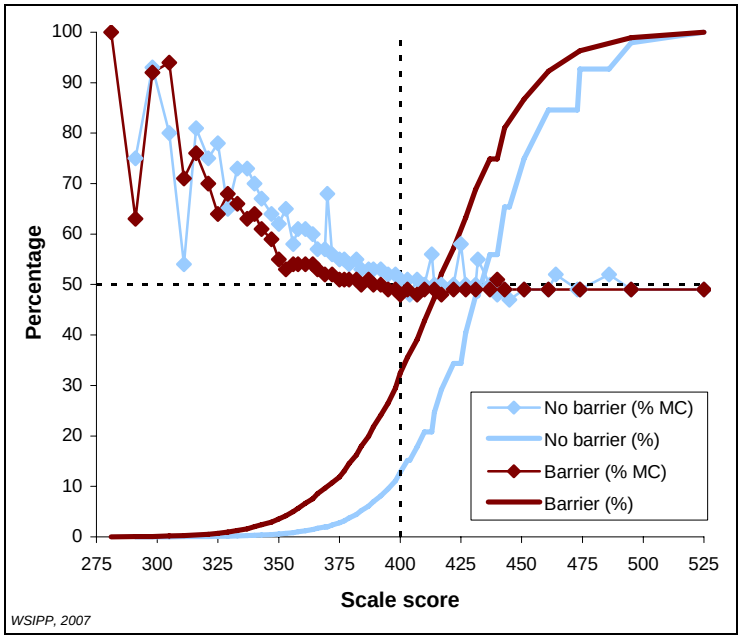


Exhibit 6
Percentage of Points From Multiple-Choice Items by Scale Scores

A. Reading



B. Math

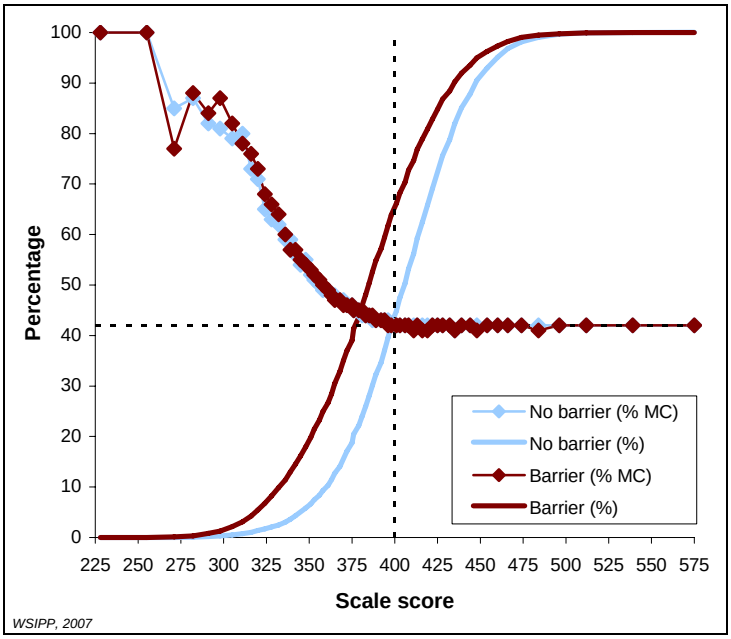
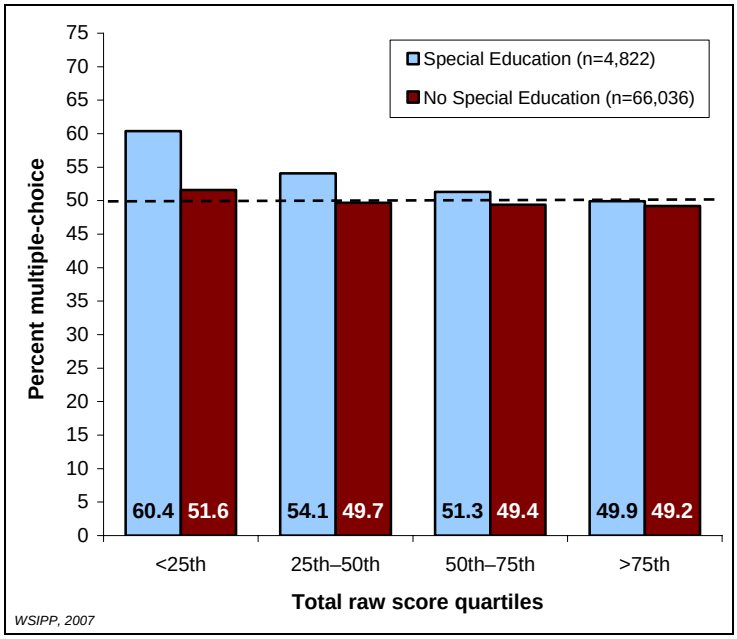


Exhibit 7
Percentage of Points From Multiple-Choice Items by Total Raw Score Quartiles

A. Reading



B. Math

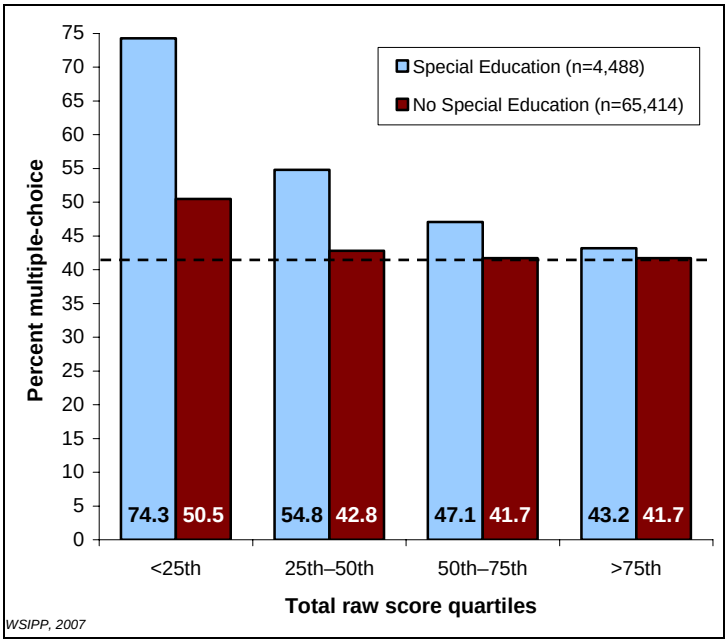
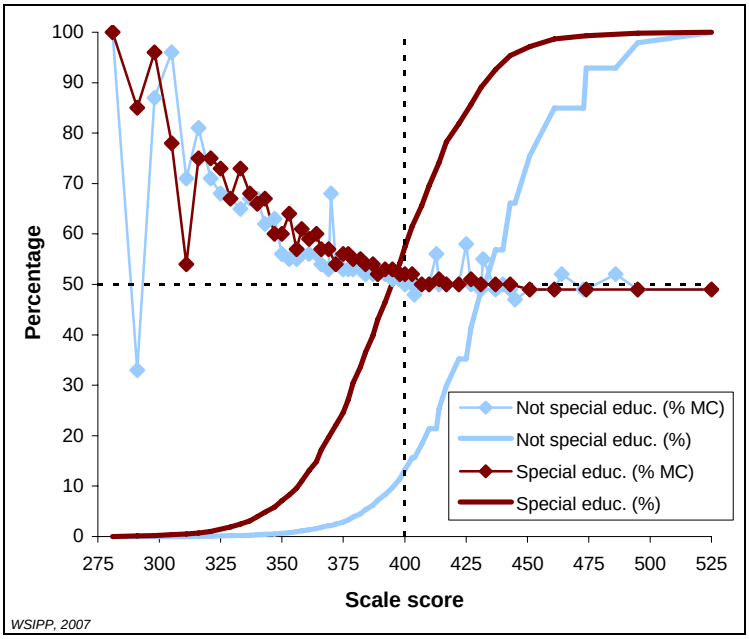


Exhibit 8
Percentage of Points From Multiple-Choice Items by Scale Scores

A. Reading



B. Math

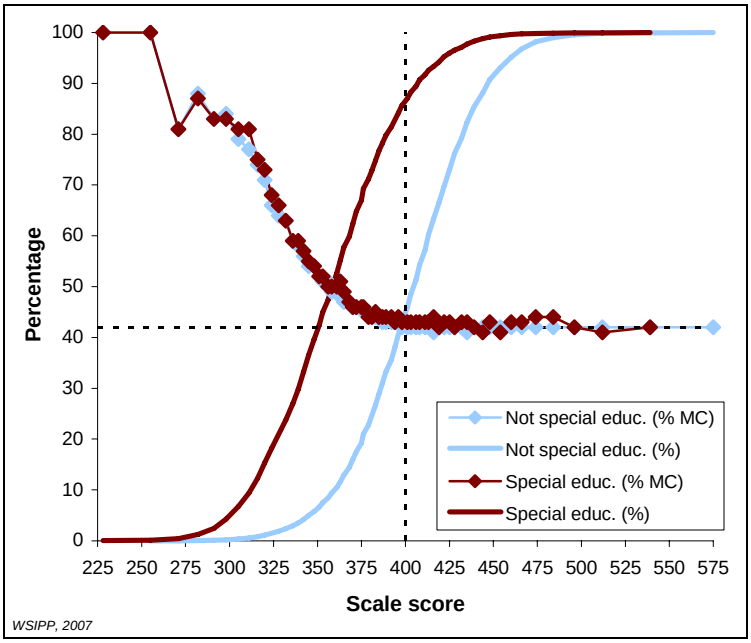
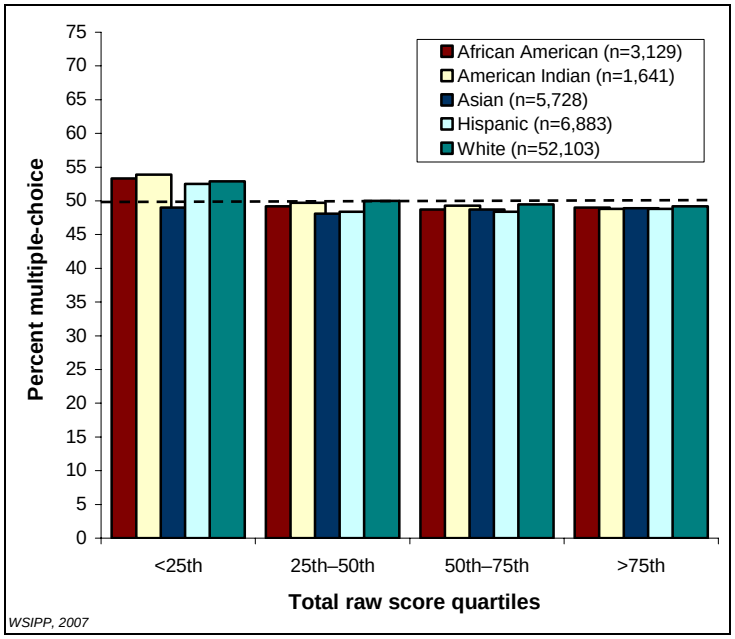


Exhibit 9
Percentage of Points From Multiple-Choice Items by Total Raw Score Quartiles

A. Reading



B. Math

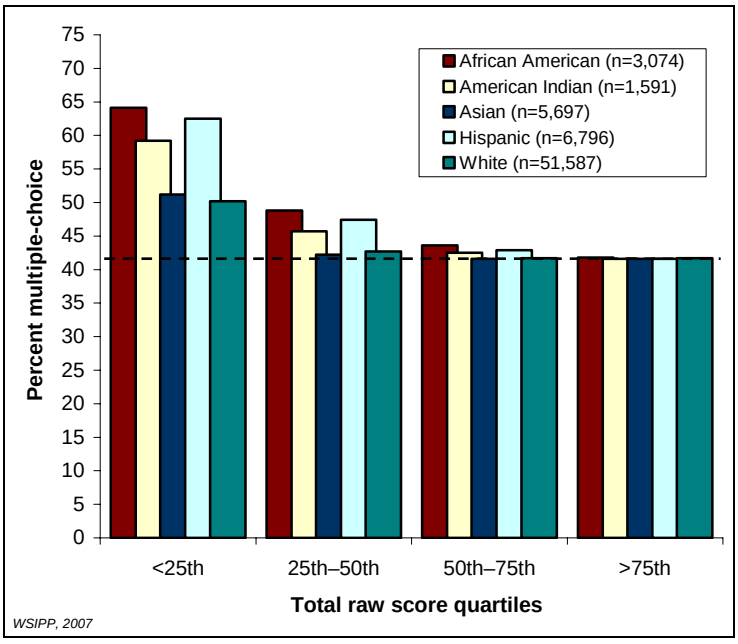
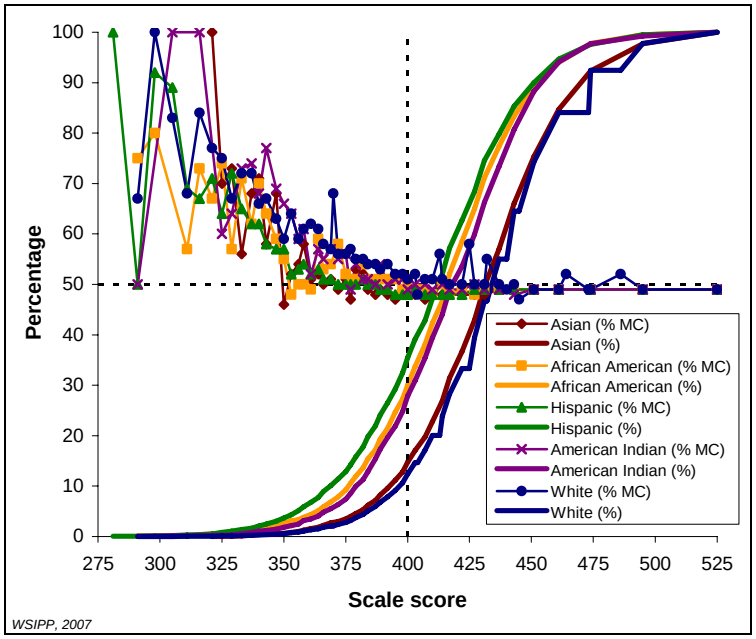
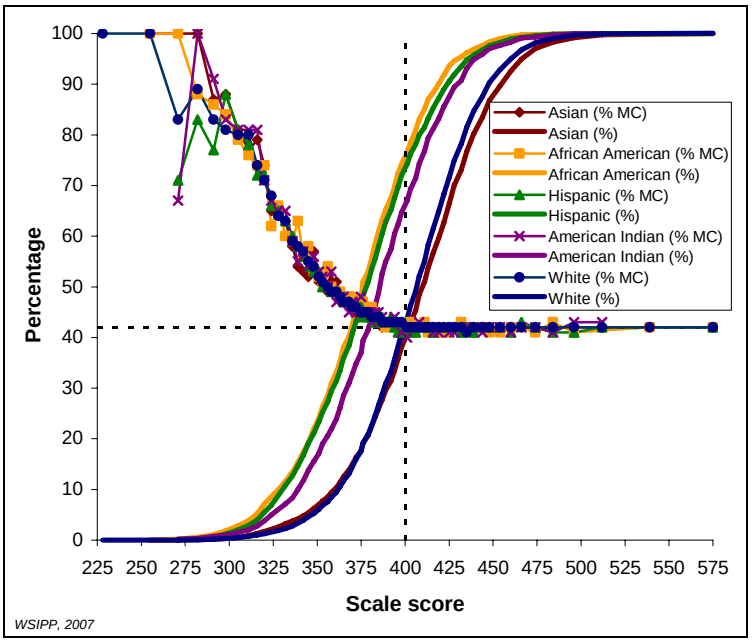


Exhibit 10
Percentage of Points From Multiple-Choice Items by Scale Scores

A. Reading



B. Math



For further information, please contact:
Robert Barnoski at barney@wsipp.wa.gov (360) 586-2744, or
Wade Cole at wcole@wsipp.wa.gov (360) 586-2791

Document No. 07-08-2202
Available for download at:
http://www.wsipp.wa.gov/pub.asp?docid=07-08-2202

Washington State Institute for Public Policy

The Washington State Legislature created the Washington State Institute for Public Policy in 1983. A Board of Directors—representing the legislature, the governor, and public universities—governs the Institute and guides the development of all activities. The Institute's mission is to carry out practical research, at legislative direction, on issues of importance to Washington State.