WASHINGTON'S HIGH SCHOOL ASSESSMENT SYSTEM:

A REVIEW OF STUDENT PERFORMANCE ON THE WASL AND ALTERNATIVE ASSESSMENT OPTIONS

Wade Cole, Ph.D.

and

Annie Pennucci

December 2007

Washington State Institute for Public Policy

WASHINGTON'S HIGH SCHOOL ASSESSMENT SYSTEM:

A REVIEW OF STUDENT PERFORMANCE ON THE WASL AND ALTERNATIVE ASSESSMENT OPTIONS

Wade Cole, Ph.D.

and

Annie Pennucci

December 2007

WASHINGTON STATE INSTITUTE FOR PUBLIC POLICY
110 Fifth Avenue SE
P.O. Box 40999
Olympia, WA 98504-0999
Telephone: (360) 586-2677
FAX: (360) 586-2793
URL: http://www.wsipp.wa.gov

Document Number 07-12-2202

WASHINGTON STATE INSTITUTE FOR PUBLIC POLICY

Mission

The Washington 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, hires the director, and guides the development of all activities.

The Institute's mission is to carry out research, at legislative direction, on issues of importance to Washington State. The Institute conducts research activities using its own policy analysts, academic specialists from universities, and consultants. New activities grow out of requests from the Washington legislature and executive branch agencies, often directed through legislation. Institute staff work closely with legislators, as well as legislative, executive, and state agency staff to define and conduct research on appropriate state public policy topics.

Current assignments include projects in welfare reform, criminal justice, education, youth violence, and social services.

Board of Directors

Senator Karen Fraser	Robin Arnold-Williams, Department of Social and Health Services
Senator Jeanne Kohl-Welles	Victor Moore, Office of Financial Management
Senator Pam Roach	Sandra Archibald, University of Washington
Senator Mark Schoesler	Andrew Bodman, Western Washington University
Representative Fred Jarrett	Thomas L. "Les" Purce, The Evergreen State College
Representative Phyllis Kenney	Robert Rosenman, Washington State University
Representative Skip Priest	Ken Conte, House Office of Program Research
Representative Helen Sommers	Richard Rodger, Senate Committee Services

Staff

Roxanne Lieb, Director
Steve Aos, Associate Director

CONTENTS

List of Exhibits

Acknowledgements:

We thank staff members at the Office of Superintendent of Public Instruction for their assistance in obtaining WASL data and in learning about the implementation of the Collection of Evidence. At the Institute, we gratefully acknowledge the invaluable contributions of Dr. Robert (Barney) Barnoski throughout this study and the efforts of Laura Harmon in managing the WASL databases. We also thank the teachers and administrators who shared their thoughts on the Collection of Evidence.

EXECUTIVE SUMMARY

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 data to increase understanding of the students who did not meet the standard in one or more areas of assessment, identify the characteristics of those students, and identify possible barriers to student success or possible causes of the lack of success."

The Legislature further directed the Institute to "explore options to augment the current system of assessments to provide additional opportunities for students to demonstrate that they have met the state learning standards."[1]

This final report addresses both components of the Institute's assignment. The first section summarizes performance on the Washington Assessment of Student Learning (WASL) and analyzes the characteristics of students who have not yet met standard. The analyses in Section I include students in the class of 2008—the first class for which the WASL is a graduation requirement—and students in the class of 2009.

The second section examines issues related to the alternative assessment options available to students who have not met standard on the WASL. Section II considers the impact of alternative assessment options on overall met-standard rates, discusses the cultural appropriateness of various student assessment options, describes the initial implementation of the Collection of Evidence (COE) option, and provides an overview of "multiple measures" assessment systems.

With respect to the statistical analysis of WASL data, the Institute reached the following broad conclusions:

- As of spring 2007, 87 percent of students in the class of 2008 have satisfied the WASL graduation requirement in reading and writing. This met-standard rate drops to 64 percent when performance on the math WASL is considered.
- Nearly 76 percent of students in the class of 2009 met standard in reading and writing on their first attempt in spring 2007; however, students continue to struggle in math. Only half the students in the class of 2009 met standard on the math WASL in spring 2007.
- WASL performance continues to vary by student characteristics. Met-standard rates are lower for African American, American Indian, and Hispanic students than for Asian and White students; for low-income students compared with their higher-income peers; for English language learners (ELL) relative to English-speaking students; and for students with one or more disabilities compared with non-disabled students.

[1] SSB 6618 § 2(1), Chapter 352, Laws of 2006.

Students who do not meet standard on the WASL may demonstrate their proficiency in the state's learning standards using one of several alternative assessment options. As of November 2007, 612 Washington students had met standard by taking an alternative assessment (354 by submitting a Collection of Evidence, 180 by achieving a sufficient score on the SAT, PSAT, or ACT, and 79 using the GPA Subject-Area Cohort option). With respect to these options, the Institute concludes:

- For an alternative assessment to match the content and rigor of the WASL, it must be geared to Washington's learning standards. "Off-the-shelf" tests are generally not fully aligned with Washington's standards.

- Student achievement gaps persist across all assessment options reviewed. This consistency suggests that the gaps are likely not due to the cultural biases embedded in the assessments but may be partially attributable to differences in the learning experiences of students.

- Washington educators involved in the initial implementation of the Collection of Evidence (COE) view this option as challenging but manageable. Teachers and administrators from 21 school districts who were interviewed for this study reported plans to expand the COE in 2007–08.

- Washington currently has a multiple measures assessment system: students must fulfill multiple academic requirements to graduate. Education researchers consulted recommend that the state examine the degree to which curriculum and instruction is aligned with learning standards in order to improve student achievement across all assessment measures.

The 2006 Legislature required that this study "include suggestions for any follow-up studies that the legislature could undertake to continue to build on the information obtained in this study."[2] As a next step, we suggest that the state undertake a systematic review of curriculum and instruction, focusing on alignment with state learning standards.

[2] SSB 6618 § 2(2).

SECTION I: STATISTICAL ANALYSIS OF THE WASL

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 data."[3] The Institute was instructed to:

- Increase understanding of the students who did not meet the standard in one or more areas of assessment;
- Identify the characteristics of those students; and
- Identify possible barriers to student success on the WASL.

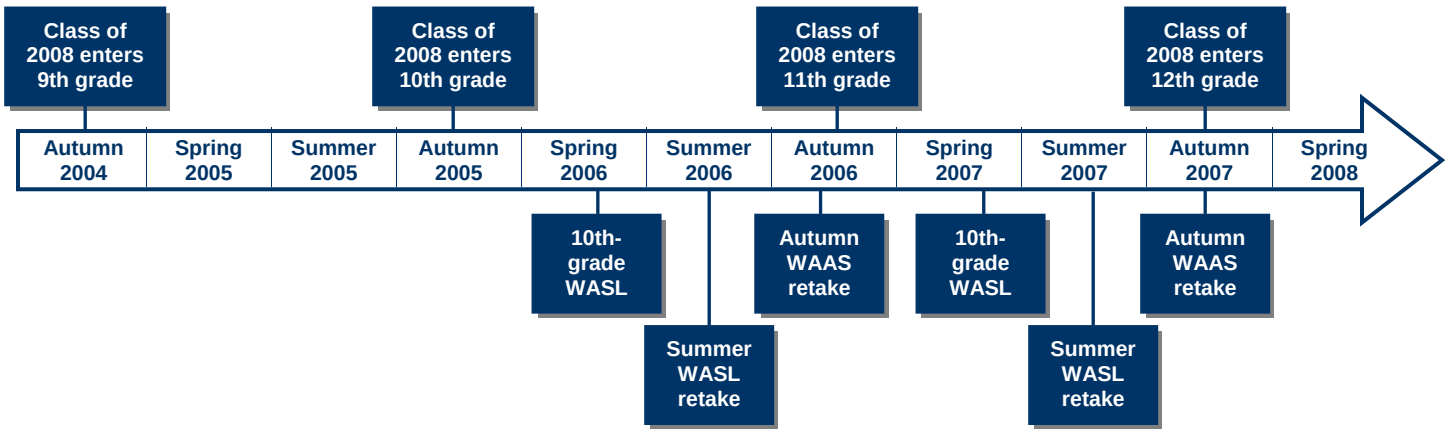
The Institute has issued several reports that examine a variety of issues related to the WASL.[4] **Section I of this report analyzes WASL performance for students in the class of 2008, the first cohort of students for whom meeting state learning standards on the WASL is a graduation requirement, and for students in the class of 2009.** Section II reviews alternative assessment options.

The report begins with an overview of Washington's high school assessment system, followed by an examination of students in the class of 2008 who have completed and met standard on the WASL as of spring 2007. We then analyze WASL completion and met-standard rates for students in the class of 2009; these students took the 10th-grade WASL for the first time in spring 2007. We close Section I by summarizing the main conclusions of the statistical analysis of the WASL.

A. OVERVIEW OF WASHINGTON'S HIGH SCHOOL ASSESSMENT SYSTEM

Exhibit 1 presents a timeline of WASL administrations for students in the class of 2008. Students in the class of 2008—those who entered 9th grade in 2004—were scheduled to take the WASL as 10th graders in spring 2006. Students who did not complete the WASL in spring 2006 have had three additional opportunities to do so: summer 2006, spring 2007, and summer 2007. At the time of this report, data for the summer 2007 WASL results were unavailable.

Exhibit 1
Administration of the WASL and WAAS for the Class of 2008



[3] SSB 6618 § 2(1). Appendix A contains the full text of the legislation.
[4] The Institute's complete series of reports on the WASL are available online at <http://www.wsipp.wa.gov/series.asp?seriesid=2>.

Instead of the WASL, students in special education may be eligible to participate in the Washington Alternate Assessment System (WAAS). At the discretion of his or her Individualized Education Program (IEP) team, a student in special education may take one of three alternate assessments:

- The **WASL-Basic** is administered to students who perform at or near grade level and who are able to take paper-and-pencil tests under routine conditions. The WASL-Basic is identical to the "regular" WASL but sets the achievement standard at Level 2 (Basic) rather than Level 3 (Proficient).[5]

- The **WAAS-Portfolio** is reserved for students with significant cognitive disabilities who are unable to take paper-and-pencil tests, even with accommodations. These students submit work samples for review.

- The **Developmentally Appropriate WASL** (WAAS-DAW) is intended for students whose performance is substantially below grade level. Students take a WASL for the grade level that most closely matches their developmental or instructional level.[6]

The WASL-Basic, WAAS-Portfolio, and WAAS-DAW may be used to satisfy high school graduation requirements, but the WAAS-Portfolio is the only alternate assessment for students in special education that may be used for the purposes of school accountability under the No Child Left Behind Act.[7]

Unlike the WASL, for which retakes are administered during the summer, WAAS retakes are administered in November. Available WAAS data on student performance cover the spring 2006, autumn 2006, and spring 2007 test administrations.

Based on changes made by the Legislature during the 2007 session,[8] students in the class of 2008 must now satisfy the following requirements to graduate:

- Meet standard on the **reading** and **writing** portions of the WASL (or approved alternative assessments);

- Complete a high school and beyond plan;[9]

- Complete a culminating project;[10] and

- Earn an additional math credit (for students who did not pass the math WASL or an approved alternative).[11]

Students in the class of 2008 who have not met standard in math must also take the math WASL an additional time.[12]

Because of this modification in the WASL graduation requirement, we report completion and met-standard rates two ways: the first includes all three WASL assessments (reading, writing, and math); the second restricts attention to the WASL subject areas needed for graduation (reading and writing).

---

[5] The WASL-Basic was known previously as the WASL-Modified and was referred to as such in the Institute's earlier reports.
[6] ESHB 2195, Chapter 19, Laws of 2004. As of December 19, 2006, the WAAS-DAW is no longer available as an assessment for 10th graders but may be taken by 11th and 12th graders. See <http://www.k12.wa.us/assessment/altassess.aspx>.
[7] Office of Superintendent of Public Instruction. (2007). *Accommodations guidelines for the students served by special education programs*. Available online at <www.k12.wa.us/SpecialEd/publications.aspx>.
[8] ESSB 6023, Chapter 354, Laws of 2007.
[9] WAC 180-051-061. For more information visit <http://www.sbe.wa.gov/gradreq/hsbeyond.htm>.
[10] Ibid. For more information visit <http://www.sbe.wa.gov/faqs/faq_pages/culminating_project_faq.htm>.
[11] This requirement is in addition to state and local course credit requirements.
[12] Students in the classes of 2009 through 2012 who do not pass the math WASL must earn two additional math credits and retake the math WASL annually. Students who graduate using this pathway do not receive a Certificate of Academic Achievement. Students will be required to meet standard on the math WASL starting in 2013. Also in 2013, the science WASL will become a graduation requirement.

B. CLASS OF 2008 RESULTS

The results of this analysis are organized around three primary questions:

- How many students **need to take** the WASL?
- How many students have **completed** the WASL?
- How many students have **met standard** on the WASL?

1) How Many Students Need to Take the WASL?

In Office of Superintendent of Public Instruction (OSPI) databases, every student in the class of 2008 is assigned a "test type" indicating which assessment—WASL, WASL-Basic, WAAS Portfolio, or WAAS-DAW—they are supposed to take for each subject. Exhibit 2 reports the number and percentage of students in the class of 2008 who were scheduled or "slated" to take the WASL and/or alternate assessments. The exhibit is based on the number of students enrolled as 11th graders in 2006–07.

Most students—93.9 percent—were slated to take the "regular" WASL rather than an alternate assessment (WAAS). An additional 4.1 percent of students were slated for alternate assessments in all three subjects (reading, writing, and math).

All other students (2 percent) were scheduled to take some combination of the WASL and the WAAS. The WAAS may be used only for content areas in which a student receives special education services.

Exhibit 2
Number of Students in the Class of 2008 Who Were Slated to Take the WASL and/or WAAS

Reading	Writing	Math	Number	Percent
Slated for the "regular" WASL				
WASL	WASL	WASL	67,653	93.9
Slated for alternate assessments (WAAS)				
WAAS	WAAS	WAAS	2,969	4.1
Slated for some combination of the WASL and WAAS				
WASL	WASL	WAAS	530	0.7
WAAS	WAAS	WASL	136	0.2
WASL	WAAS	WAAS	214	0.3
WASL	WAAS	WASL	242	0.3
WAAS	WASL	WAAS	185	0.3
WAAS	WASL	WASL	116	0.2
Total				
Either	Either	Either	72,045	100.0

WASL = Washington Assessment of Student Learning.
WAAS = Washington Alternate Assessment System.

2) How Many Students Have Completed the WASL?

Exhibit 3 shows that most students in the class of 2008 have completed (but not necessarily met standard on) the WASL. As of spring 2007, 94 percent of students in the class of 2008 who were enrolled as 11th graders in 2006–07 and who were slated to take the WASL (n=67,653) had completed all three reading, writing, and math assessments. Exhibit 4 adds students who were slated for at least one subject-area WAAS: 93 percent of all students enrolled last spring (n=72,045) completed either the WASL or WAAS as scheduled.

Exhibit 3
Percentage of Students Completing the WASL as of Spring 2007*

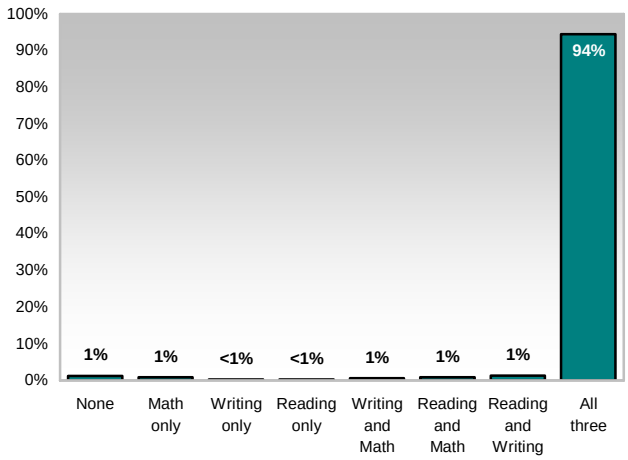
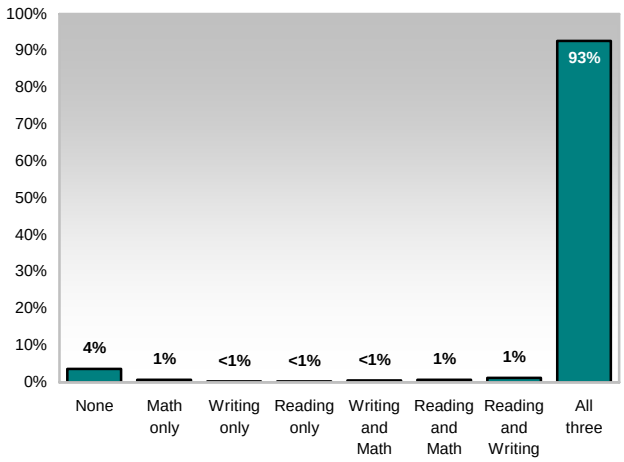


Exhibit 4
Percentage of Students Completing the WASL or WAAS as of Spring 2007†



* Based on the number of public school students in the class of 2008 who were enrolled as 11th graders in 2006–07 and who were slated to take the WASL (n=67,653).

† Based on the number of public school students in the class of 2008 who were enrolled as 11th graders in 2006–07 and who were slated to take either the WASL or the WAAS (n=72,045).

Although most students had completed the WASL by spring 2007, Exhibit 5 indicates that completion rates continue to vary by student characteristics.[13]

- **Gender:** 95 percent of females have completed the reading, writing, and math sections of the WASL compared with 93.9 percent of males.
- **Race/Ethnicity:** Completion rates are lowest among American Indians (89.8 percent), followed by African Americans (91.8 percent), Hispanics (93.9 percent), Whites (95.8 percent), and Asians (97.4 percent).
- **Low Income:** Students who are eligible for free/reduced-price meals have a lower completion rate (91.4 percent) compared with higher-income students (95.9 percent).
- **Language Barriers:** English language learners complete the WASL at a lower rate (89.7 percent) than other students (94.7 percent).
- **Disability Status:** Students with at least one documented disability have the lowest completion rate—76.1 percent—of any student characteristic considered in this report.

[13] For comparison with rates of completing the spring 2006 WASL administration see R. Barnoski & W. Cole. (2007). *Who has and has not yet completed the 10th-grade WASL?* Olympia: Washington State Institute for Public Policy, Document No. 07-02-2203.

Exhibit 5
WASL Completion Rates by Student Characteristics
As of Spring 2007

Student Characteristics	Completion Rates	
	Reading, Writing, and Math	Reading and Writing
All Students	94.5	95.4
Female	95.0	95.9
Male	93.9	94.8
Asian	97.4	97.9
African American	91.8	92.9
American Indian	89.8	91.0
Hispanic	93.9	95.1
White	95.8	96.5
Low Income	91.4	92.9
Not Low Income	95.9	96.6
ELL	89.7	91.1
Not ELL	94.7	95.6
Disability	76.1	76.6
No Disability	95.8	96.9

Exhibit 5 also reports the percentage of students who have completed the reading and writing portions of the WASL, irrespective of the math assessment. Completion rates increase slightly when math is not considered but are characterized by the same discrepancies as reading and writing.

3) How Many Students Have Met Standard on the WASL?

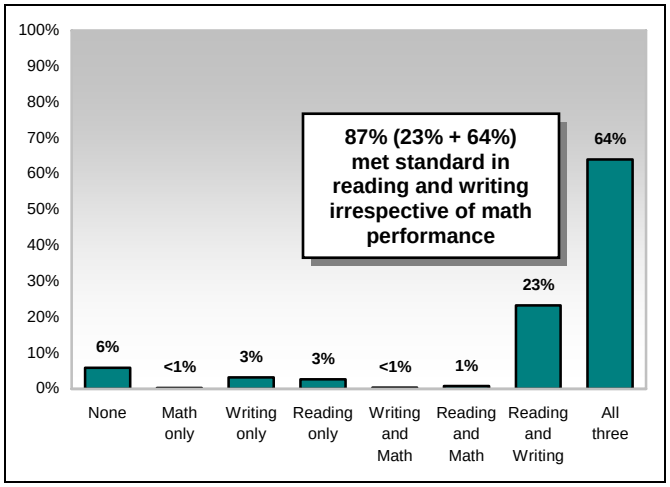
Exhibit 6 shows that 64 percent of students in the class of 2008 who were enrolled as 11th graders in 2006–07 and who were slated to take the WASL (n=67,653) had met standard in reading, writing, and math. An additional 23 percent of students had met standard in reading and writing, but not math. Taken together, these results indicate that **87 percent of students in the class of 2008 satisfied the WASL graduation requirement for reading and writing** as of spring 2007. Approximately 8 percent of students had met standard in either reading or writing but not both, and 6 percent of students had not met standard in any subject.

Met-standard rates decline somewhat when all students—those who are slated to take the WAAS in addition to those slated for the WASL—are included (n=72,045). This decline occurs because students who take the WAAS have lower met-standard rates than students who take the WASL. As illustrated in Exhibit 7, 61 percent of students who were slated for either the WASL or the WAAS met standard in reading, writing, and math; an additional 23 percent met standard in reading and writing but not math.

Using this method of calculating met-standard rates, **84 percent of students have satisfied the WASL or WAAS graduation requirement for reading and writing**. Again, 8 percent met standard in either reading or writing, but not both.

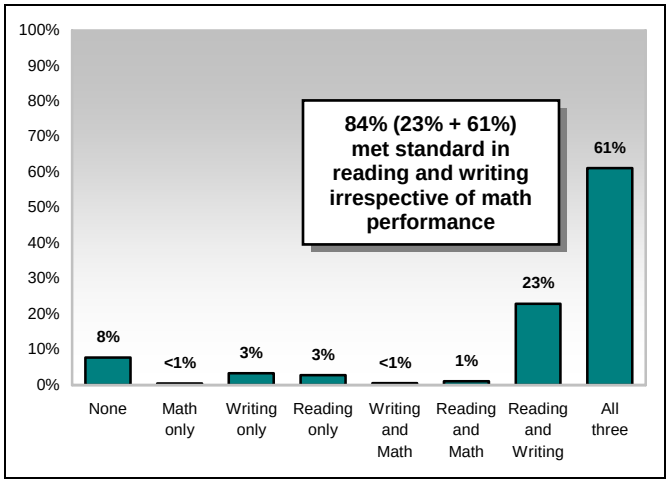
These results may overestimate the percentage of students who are eligible to graduate because they do not include information about whether other high school graduation requirements—the culminating project, a high school and beyond plan, and credit requirements—have been met. Conversely, the results slightly underestimate the percentage of students who have satisfied the WASL graduation requirement because they do not include students who met standard using an approved alternative assessment. (Student performance on alternative assessments is considered in Section II of this report.) The results also do not account for students who dropped or transferred out of Washington State's public schools prior to the 2006–07 school year or were held back to repeat a grade.

Exhibit 6
Percentage of Students Meeting Standard on the WASL
As of Spring 2007



Note: Based on the number of students in the class of 2008 who were enrolled as 11th graders in 2006–07 and who were slated to take the WASL (n=67,653).

Exhibit 7
Percentage of Students Meeting Standard on the WASL or WAAS
As of Spring 2007



Note: Based on the number of students in the class of 2008 who were enrolled as 11th graders in 2006–07 and who were slated to take either the WASL or the WAAS (n=72,045).

Among students who have not met standard on the WASL, what percentage had level 1 (Below Basic) and level 2 (Basic) scores? Exhibit 8 demonstrates that the majority of students who have not met standard in reading, writing, or math scored in the level 2 range—77 percent, 78 percent, and 60 percent, respectively. These students are closer to meeting standard than level 1 students. For comparison, we also report the number of students with level 3 (Proficient) or level 4 (Advanced) scores.

Exhibit 8
Performance Level by WASL Subject

Subject	Not Met Standard			Met Standard		
	Total	Level 1	Level 2	Total	Level 3	Level 4
Reading	4,613 (100%)	1,038 (23%)	3,575 (77%)	61,059 (100%)	18,185 (30%)	42,874 (70%)
Writing	4,414 (100%)	976 (22%)	3,438 (78%)	61,651 (100%)	30,343 (49%)	31,308 (51%)
Math	21,835 (100%)	8,636 (40%)	13,199 (60%)	44,278 (100%)	30,356 (69%)	13,922 (31%)

Exhibit 9 reports met-standard rates by student characteristics. We calculate met-standard rates as the percentage of (1) all students who were slated to take the WASL, and (2) students who completed the WASL. As with the analysis of completion rates, we report met-standard rates for reading, writing, and math as well as for reading and writing.

Exhibit 9
WASL Met-Standard Rates by Student Characteristics
As of Spring 2007

Student Characteristics	Met-Standard Rates			
	Percentage of Slated		Percentage of Completed	
	Reading, Writing, and Math	Reading and Writing	Reading, Writing, and Math	Reading and Writing
All Students	63.4	86.3	67.1	90.4
Female	62.0	89.3	63.4	92.1
Male	64.9	83.1	66.5	86.0
Asian	72.9	91.2	73.5	92.4
African American	35.5	75.8	37.0	79.7
American Indian	43.3	75.6	45.4	80.4
Hispanic	39.5	74.3	40.3	76.6
White	68.8	89.4	69.9	91.5
Low Income	44.5	75.8	46.1	79.6
Not Low Income	72.6	91.4	73.9	93.6
ELL	23.9	51.6	25.4	55.4
Not ELL	65.3	87.9	66.7	90.5
Disability	15.6	38.3	16.2	44.5
No Disability	66.8	90.0	68.3	92.1

As we found in our series of WASL reports,[14] performance on the WASL varies by student characteristics, often dramatically. We focus here on students who met standard in reading and writing—the WASL graduation requirement—as a percentage of students who were slated to take the WASL.

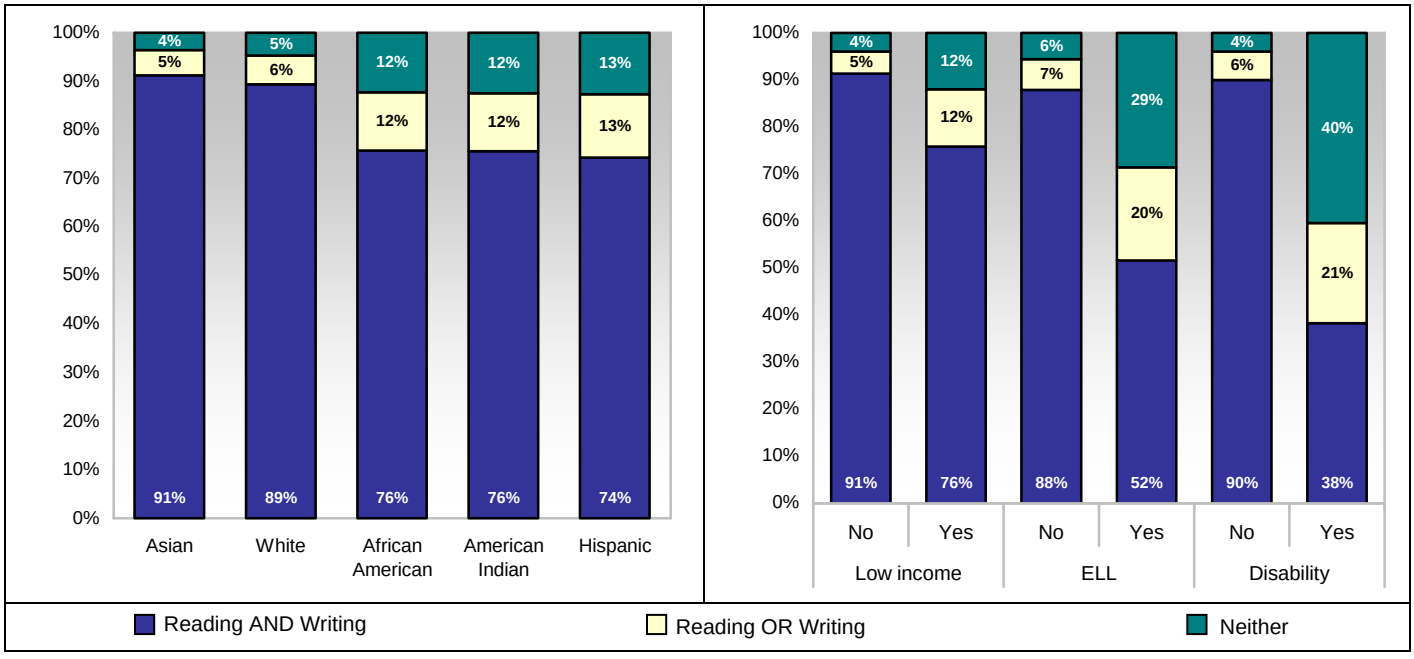
- **Gender.** 89.3 percent of females have met standard in reading and writing compared with 83.1 percent of males.
- **Race/Ethnicity.** Met-standard rates are lowest among Hispanics (74.3 percent), followed by American Indians (75.6 percent), African Americans (75.8 percent), Whites (89.4 percent), and Asians (91.2 percent).
- **Low Income.** Three-quarters of students eligible for free/reduced-price meals met standard in reading and writing compared with 91.4 percent of higher-income students.
- **Language Barriers.** Just over half of English language learners (51.6 percent) met standard in reading and writing compared with 87.9 percent of non-ELL students.
- **Disability Status.** Students with at least one documented disability have the lowest met-standard rate (38.3 percent).

Met-standard rates are slightly higher—increasing by between 1 and 6 percentage points—when calculated as a percentage of students who *completed* the WASL as opposed to students who were *slated* to take the WASL. Incorporating math performance into the calculation reduces met-standard rates substantially, by 18 to 40 percentage points.

Exhibit 10 examines student performance in reading and writing more closely by depicting the percentage of students with various demographic characteristics who have met standard in the following subjects:

- **Both** reading **and** writing,
- **Either** reading **or** writing, and
- **Neither** reading **nor** writing.

Exhibit 10
Percentage of Students Meeting Standard in Reading and Writing by Student Characteristics (as of Spring 2007)



[14] <http://www.wsipp.wa.gov/series.asp?seriesid=2>.

The percentage of students who met standard in both reading and writing has already been considered. Here we focus on students who still need to meet standard in reading, writing, or both in order to satisfy the graduation requirement. All percentages are based on the number of students who are slated to take the WASL.

Race/Ethnicity. Between 12 and 13 percent of African American, American Indian, and Hispanic students have not yet met standard in reading and writing; another 12 to 13 percent of these students have met standard in one or the other subject, but not both. Conversely, only 5 and 6 percent of Asian and White students, respectively, have not met standard in either reading or writing but not both, while an additional 4 and 5 percent have not met standard in either subject.

Low Income. Twelve percent of students who are eligible for free/reduced-price meals have met standard in reading or writing, and an additional 12 percent have not met standard in either subject. These results compare with 5 percent of higher-income students who have met standard on one but not the other assessment and 4 percent who have met standard in neither subject.

Language Barriers. Performance in reading and writing is lowest for students who are classified as English language learners. Only half of ELL students had met standard by spring 2007, with an additional 20 percent meeting standard in reading or writing but not both. More than one-quarter of ELL students—29 percent—had not met standard in reading or writing by spring 2007.

Disabilities. Only 38 percent of students with at least one documented disability met standard in both reading and writing compared with 90 percent of their non-disabled peers. Forty percent of students with one or more disabilities had not met standard in either reading or writing as of spring 2007.

Gender (not shown in Exhibit 10). Five percent of females and 8 percent of males have not met standard in either reading or writing. An additional 6 and 9 percent of females and males, respectively, have met standard in one subject but not the other.

4) Statistical Analysis of Completion and Met-Standard Rates

To this point we have considered how student characteristics are individually associated with completion and met-standard rates. The results presented below are based on statistical analyses that estimate the relative strength of associations between different student characteristics and WASL completion or performance. From this analysis we learn the independent contribution of each student characteristic to met-standard rates.

We use a statistical technique called logistic regression. The standardized logistic regression coefficients reported in Exhibit 11 describe the association of each student characteristic with WASL completion and met-standard rates. Because the estimates are standardized, they can be compared directly: *the larger the absolute value, whether positive or negative, the stronger its association with completion or met-standard rates*. Positive values indicate that the characteristic is associated with increased completion (or met-standard) rates, whereas negative values indicate lower rates.

Exhibit 11
Relative Association of Student Characteristics
With WASL Completion and Performance
As of Spring 2007

Student Characteristics	Completion		Met Standard	
	Reading, Writing, and Math	Reading and Writing	Reading, Writing, and Math	Reading and Writing
Female	-6	-5	-22	+5
African American	0	0	-11	-4
American Indian	-2	-2	-3	-2
Asian	+9	+9	+1	+5
Hispanic	+9	+9	-6	+2
Low Income	-9	-8	-14	-15
ELL	-10	-11	-19	-25
Disability	-22	-28	-31	-37
GPA	+47	+41	+73	+56
AUC*	0.805	0.819	0.857	0.879

Standardized logistic regression coefficients (times 100).
* Area Under Receiver Operating Characteristic Curve.

By far, the characteristic having the strongest association with WASL completion and performance is grade point average (GPA). On average, higher GPAs are associated with higher completion rates and increased met-standard rates.

By comparison, the association between having a disability and WASL outcomes is about half as strong as GPA and in the opposite direction. English language learner status and eligibility for free/reduced-price meal benefits are also associated with lower completion and met-standard rates, although these associations are weaker still.

A few student characteristics relate variably to completion and performance, or else are not consistently associated with performance across subjects. For example, completion rates do not differ between African American and White students once other factors are controlled, but on average African American students have lower met-standard rates than White students. Also, females have lower met-standard rates than males in reading, writing, and math combined, but slightly higher met-standard rates when only reading and writing are considered. This finding is consistent with the Institute's previous analysis of WASL performance: compared with males, females have lower met-standard rates in math but higher met-standard rates in reading and writing.[15]

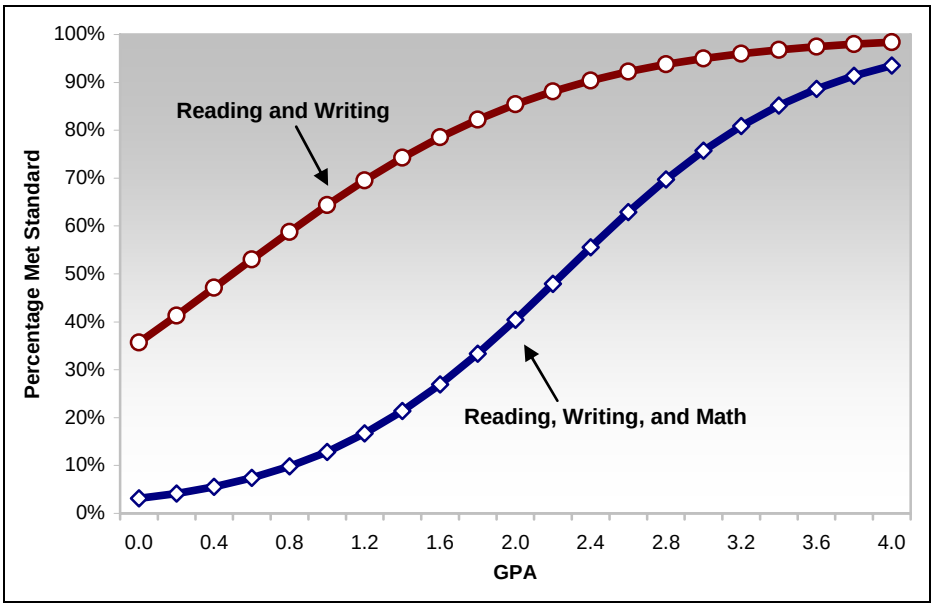
To gauge the explanatory power of our analysis—to account for how well the characteristics in our analysis account for meeting standard on the WASL—we use a statistic called the Area Under the Receiver Operating Characteristic Curve (AUC). The AUC, reported in the last row of Exhibit 11, indicates how well the characteristics in our analysis explain completion and met-standard rates. An AUC of 1.00 means that one could determine whether a student did or did not complete (or meet standard on) the WASL simply by knowing his or her combination of characteristics. Conversely, an AUC of 0.500 would indicate that the set of characteristics does not distinguish students by their completion of (or performance on) the WASL.[16]

[15] R. Barnoski & W. Cole. (2007). *Tenth-grade WASL in spring 2006: Relative strength of associations between student characteristics and met-standard rates*. Olympia: Washington State Institute for Public Policy, Document No. 07-01-2206.
[16] M.E. Rice & G.T. Harris. (2005). Comparing effect sizes in follow-up studies: ROC Area, Cohen's *d*, and *r*. *Law and Human Behavior*, *29*(5): 615-620; J.A. Swets. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*: 1285-1293.

The AUCs for WASL completion and performance are in the 0.800s, indicating that gender, race/ethnicity, eligibility for free/reduced-price meals, English language learner status, disability status, and GPA are, in combination, strongly associated with WASL outcomes. The AUCs also show that these characteristics are more strongly associated with met-standard rates than with completion rates, and that associations are slightly stronger for reading and writing than when math is also considered.

To illustrate how strongly GPA is associated with WASL performance, Exhibit 12 uses the results of the statistical analysis to calculate the predicted probability that students met standard on the WASL given their GPA, holding other characteristics constant at their averages across all students in the class of 2008. Assuming that a student has "average" characteristics, how is GPA related to WASL performance?

Exhibit 12
Assessing the Relationship Between GPA and WASL Performance



Note: Estimates adjusted for gender, race/ethnicity, poverty, English language learner status, and disability status.

The trends are clear: WASL performance increases as GPA increases. Nearly all students with a GPA of 4.0 meet standard on the WASL. Students with a GPA of 2.74—the average GPA for students in the class of 2008—have a predicted met standard rate of 93.3 percent for reading and writing and 67.7 percent for reading, writing, and math.

This analysis suggests that students who did not meet standard on the WASL also struggle in the classroom.

C. CLASS OF 2009 RESULTS

This section examines WASL completion and performance for students in the class of 2009 who were scheduled to take the 10th-grade WASL for the first time in spring 2007. The results include only the spring 2007 administration of the WASL; they do not include results from the summer 2007 retake because these data were not yet available when this report was written. We present information on (1) the number of students in the class of 2009 slated for different assessments, (2) completion rates, and (3) met-standard rates.

1) Number of "Slated" Students

Exhibit 13 shows that most students—95.2 percent—in the class of 2009 were scheduled to take the "regular" WASL for all three subjects: reading, writing, and math. An additional 2.7 percent of students were slated to take some combination of the WAAS-Portfolio or WASL-Basic.[17] The remaining 2.1 percent of students were slated to take some combination of the WASL and alternate assessments.

Exhibit 13
Number of Students Who Were Slated to Take the WASL or WAAS
Class of 2009

Reading	Writing	Math	Number	Percent
Slated for the "regular" WASL				
WASL	WASL	WASL	71,567	95.2
Slated for alternate assessments (WAAS)				
WAAS	WAAS	WAAS	2,053	2.7
Slated for some combination of the WASL and WAAS				
WAAS	WAAS	WASL	465	0.6
WASL	WASL	WAAS	340	0.5
WASL	WAAS	WASL	238	0.3
WASL	WAAS	WAAS	232	0.3
WAAS	WASL	WAAS	159	0.2
WAAS	WASL	WASL	143	0.2
Total				
Either	Either	Either	75,197	100.0

WASL = Washington Assessment of Student Learning.
WAAS = Washington Alternate Assessment System.

2) Completion Rates

Most students in the class of 2009 who were slated to take the WASL in spring 2007 did so. As illustrated in Exhibit 14, 86 percent of students completed the reading, writing, and math portions of the WASL as scheduled, and only 3 percent did not complete any of the assessments. Likewise, 86 percent of all students in the class of 2009 completed either a WASL or WAAS in reading, writing, and math (see Exhibit 15).

The percentage of students in the class of 2009 who did not complete the WASL as scheduled in spring 2007—approximately 14 percent—is comparable to the percentage of students in the class of 2008 who did not complete the WASL during the spring of their 10th-grade year.[18]

[17] Due to requirements established under the federal No Child Left Behind Act, the Developmentally Appropriate WASL is no longer administered to 10th graders. Eligible students in the 11th and 12th grades may continue to take the WAAS-DAW to satisfy Washington State's graduation requirements.
[18] Barnoski, 2007, Document No. 07-02-2203.

Exhibit 14
Percentage of Students Completing the WASL in Spring 2007*
Class of 2009

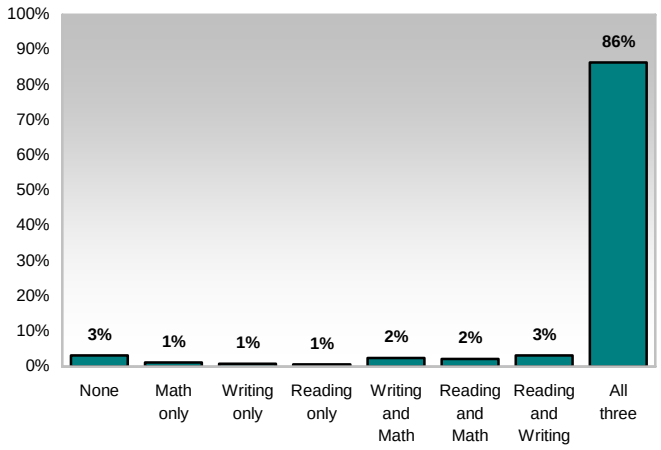
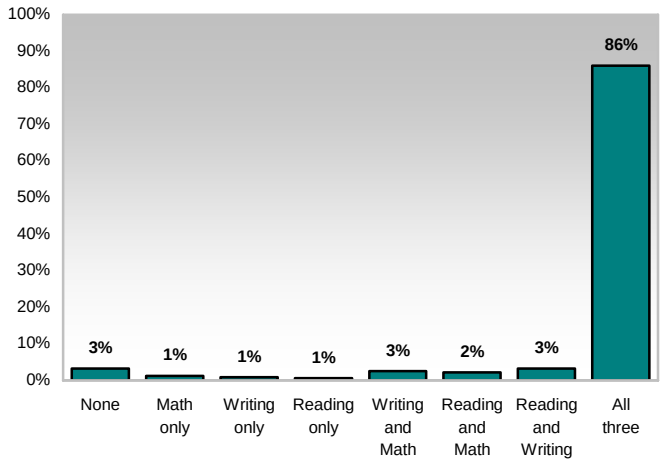


Exhibit 15
Percentage of Students Completing the WASL or WAAS in Spring 2007†
Class of 2009



* Based on the number of public school students in the class of 2009 who were enrolled as 10th graders in 2006–07 and who were slated to take the WASL (n=71,567).

† Based on the number of public school students in the class of 2009 who were enrolled as 10th graders in 2006–07 and who were slated to take either the WASL or the WAAS (n=75,197).

Exhibit 16 reports completion rates by subject area and student characteristics. In every subject, completion rates are lower for "disadvantaged" students—non-Asian minorities, low-income students, English language learners, and students with disabilities—relative to their peers. Moreover, completion rates for all groups tend to be slightly lower in math than in reading and writing.

Exhibit 16
WASL Completion Rates by Student Characteristics
Class of 2009

Student Characteristics	WASL Subjects				
	Reading	Writing	Math	Reading and Writing	Reading, Writing, and Math
All Students	92.2	92.7	92.1	89.6	86.4
Female	92.7	93.3	92.6	90.1	86.6
Male	91.6	92.1	91.5	88.9	85.4
Asian	94.3	94.5	94.1	92.3	90.1
African American	88.1	89.0	88.7	84.7	80.1
American Indian	85.7	87.0	84.6	81.3	75.3
Hispanic	88.7	89.1	88.3	85.2	81.1
White	93.0	93.5	92.9	90.4	87.0
Low Income	88.0	89.0	87.7	84.3	79.4
Not Low Income	94.3	94.7	94.4	92.2	89.4
ELL	84.6	84.7	85.4	80.8	77.4
Not ELL	92.6	93.2	92.4	89.9	86.5
Disability	80.8	81.0	80.0	75.9	64.5
No Disability	93.0	93.6	92.9	90.4	87.4

Exhibit 17 charts completion rates by selected student characteristics for two combinations of subject-area assessments: reading and writing, and reading, writing, and math. Completion rates by race/ethnicity are lowest for American Indian students, followed by African American, Hispanic, White, and Asian students.

Exhibit 17
Percentage of Students Completing the WASL by Student Characteristics
Class of 2009

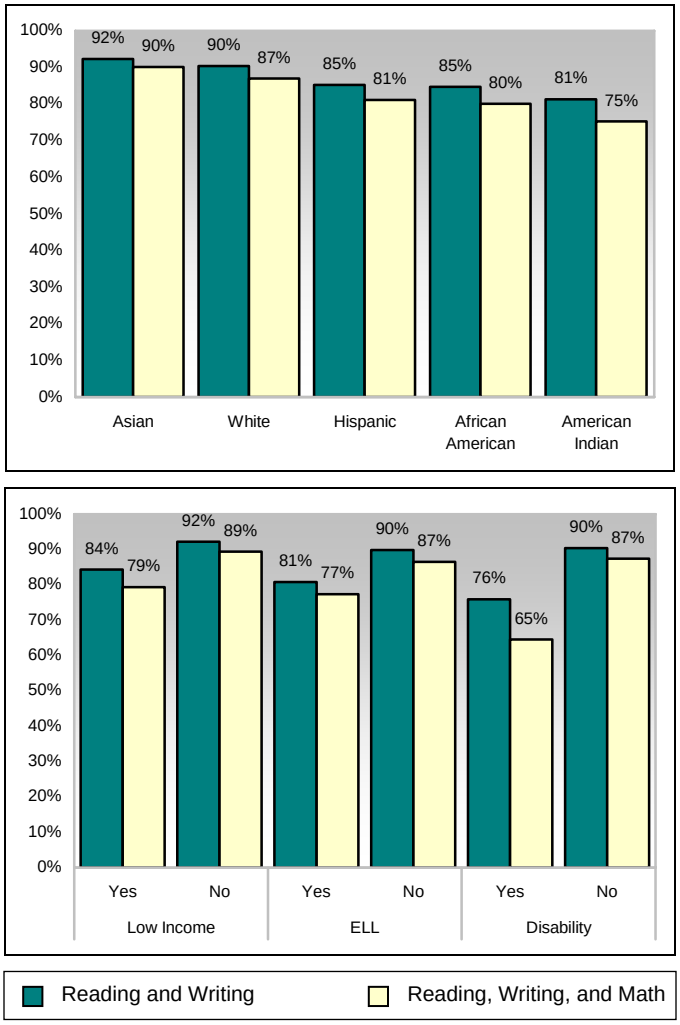


Exhibit 17 also displays completion rates by income, English language learner status, and disability status.

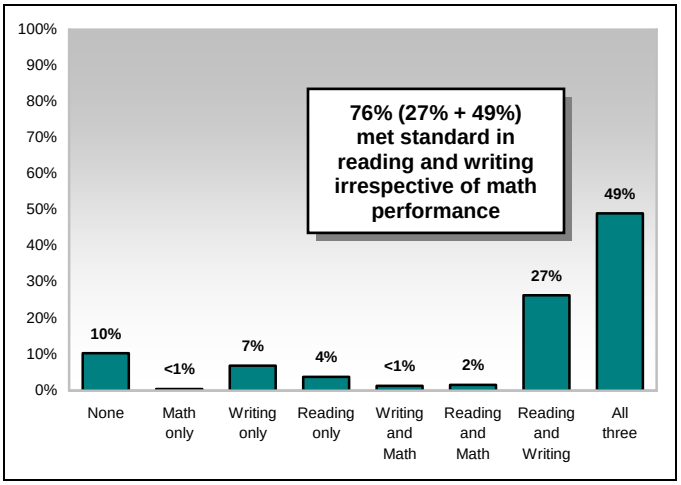
- 79 percent of students who are eligible for free/reduced-price meals completed the reading, writing, and math WASL assessments compared with 89 percent of non-eligible students.
- 77 percent of English language learners completed reading, writing, and math compared with 87 percent of students not receiving bilingual services.
- 65 percent of students with at least one documented disability completed all three subject-area assessments of the WASL, while 87 percent of their non-disabled peers did so.

Completion rates for reading and writing are slightly higher than completion rates in reading, writing, and math combined but are characterized by the same gaps.

3) Met-Standard Rates

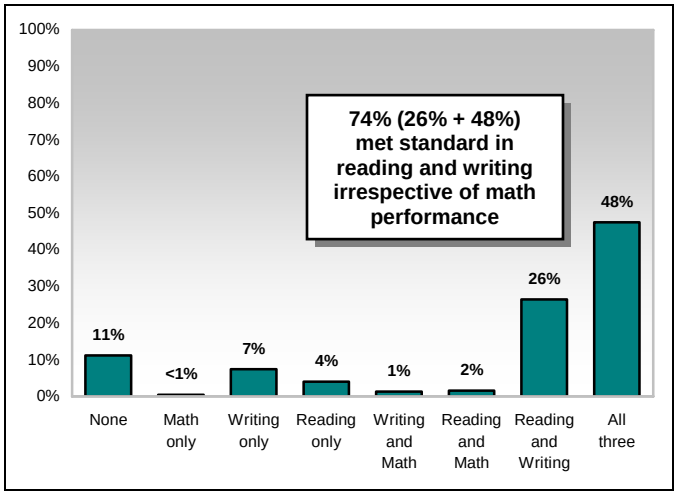
Exhibits 18 and 19 indicate that approximately half of students in the class of 2009 met standard in reading, writing, and math: 49 percent of students slated for all three WASL assessments met standard in each subject, whereas 48 percent of all students met standard on either the WASL or WAAS. In addition, 27 percent of students met standard in reading and writing but not math. Approximately 10 percent of students did not meet standard in any subject; this percentage includes students who did not complete the WASL as scheduled in spring 2007.

Exhibit 18
Percentage of Students Meeting Standard on the WASL in Spring 2007
Class of 2009



Note: Based on the number of students in the class of 2009 who were enrolled as 10th graders in 2006–07 and who were slated to take the WASL (n=71,567).

Exhibit 19
Percentage of Students Meeting Standard on the WASL or WAAS in Spring 2007
Class of 2009



Note: Based on the number of students in the class of 2009 who were enrolled as 10th graders in 2006–07 and who were slated to take either the WASL or the WAAS (n=75,197).

Exhibit 20 reports met-standard rates for the class of 2009 by various student characteristics. Again, we calculate met-standard rates as the percentage of (1) all students who were slated to take the WASL, and (2) students who have completed the WASL. We report met-standard rates individually for reading, writing, and math, as well as for two combinations of subject-area assessments: reading and writing, and reading, writing, and math. Because some students did not complete the WASL as scheduled in spring 2007, met-standard rates are higher as a percentage of completers than as a percentage of students who were slated to take the WASL.

Exhibit 20
WASL Met-Standard Rates by Student Characteristics
Class of 2009

Student Characteristics	WASL Subjects				
	Reading	Writing	Math	Reading and Writing	Reading, Writing, and Math
Met Standard (Percentage of Slated)					
All Students	80.4	83.4	52.1	75.6	49.2
Female	83.6	87.7	50.5	79.7	48.1
Male	77.2	79.1	53.6	70.9	49.8
Asian	85.4	87.2	61.0	81.0	57.7
African American	66.0	73.0	25.3	59.7	23.5
American Indian	66.9	71.3	32.7	59.8	29.9
Hispanic	65.8	68.3	26.9	57.8	25.2
White	83.5	86.4	57.3	78.9	53.8
Low Income	68.3	72.4	33.4	61.0	30.4
Not Low Income	86.7	89.1	61.8	82.7	58.5
ELL	44.0	43.9	14.2	32.4	10.8
Not ELL	82.3	85.5	54.0	77.5	50.9
Disability	37.9	42.6	11.5	28.9	8.6
No Disability	83.5	86.3	55.0	78.4	51.6
Met Standard (Percentage of Completed)					
All Students	87.3	90.0	56.6	85.0	56.9
Female	90.2	93.9	54.5	88.5	55.5
Male	84.3	85.9	58.6	79.7	58.3
Asian	90.6	92.3	64.8	87.7	64.1
African American	74.9	81.9	28.5	70.5	29.4
American Indian	78.1	82.0	38.6	73.6	39.8
Hispanic	74.1	76.6	30.5	67.8	31.0
White	89.8	92.3	61.7	87.2	61.9
Low Income	77.6	81.4	38.0	72.4	38.3
Not Low Income	91.9	94.1	65.5	89.7	65.4
ELL	52.1	51.8	16.6	40.1	13.9
Not ELL	88.9	91.7	58.5	86.2	58.9
Disability	46.8	52.6	14.3	38.1	13.4
No Disability	89.8	92.2	59.2	86.7	59.0

Exhibit 21 displays met-standard rates as a percentage of students who were slated to take the WASL for two combinations of subject-area assessments—reading and writing with and without math—by student characteristics.

- 81 percent of Asian students who were slated to take the reading and writing portions of the WASL met standard compared with 79 percent of White students. In contrast, 60 percent of African American and American Indian students met standard in reading and writing. Hispanic students had the lowest met-standard rate (58 percent).

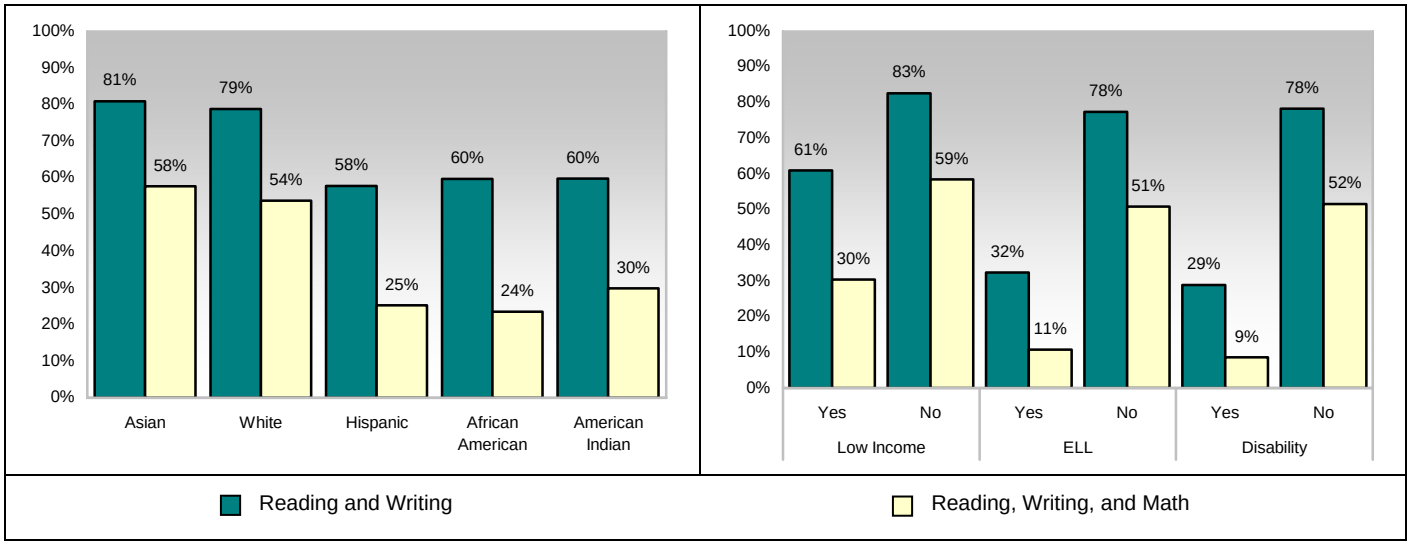
- Met-standard rates decline substantially when performance in math is also considered, but gaps by race/ethnicity persist. Approximately 1 in 2 Asian and White students met standard in reading, writing, and math compared with roughly 1 in 3 American Indian students and 1 in 4 African American and Hispanic students.

- Low-income students had lower met-standard rates than their higher-income peers. Among students who were eligible for free/reduced-price meal benefits, 61 percent met standard in reading and writing; this compares with 83 percent of non-eligible students. The corresponding met-standard rates for reading, writing, and math are 30 and 59 percent.

- Performance gaps are also large for students who did and did not receive transitional bilingual instruction: 32 percent of English language learners and 78 percent of other students met standard in reading and writing; met-standard rates in reading, writing, and math combined were 11 and 51 percent for ELL and non-ELL students, respectively.

- Differentials in WASL performance are largest for students with and without documented disabilities. Approximately 29 percent of students with at least one disability and who were slated to take the "regular" reading and writing WASL assessments met standard, while 78 percent of their non-disabled peers met standard. When performance in math is also considered, only 9 percent of students with one or more disabilities met standard compared with slightly more than half (52 percent) of non-disabled students.

Exhibit 21
Percentage of Students Meeting Standard on the WASL
Class of 2009



Note: As a percentage of students slated to take each combination of assessments.

Compared with performance on the WASL in spring 2006, 10th graders who completed the WASL in spring 2007 had slightly higher met-standard rates in reading and math. For most groups of students, met-standard rates in writing improved considerably from 2006 to 2007. Overall, 84.3 percent of students in the class of 2008 met standard in writing on the spring 2006 WASL;[19] in spring 2007, 90 percent of students in the class of 2009 met standard on the writing WASL. Met-standard rates in writing increased by as much as 13 percentage points for ELL students.

Exhibits 22 and 23 report the level of WASL performance for students who did not meet standard in spring 2007. There are four levels of WASL performance: 1 (Below Basic), 2 (Basic), 3 (Proficient), and 4 (Advanced). Students with a level 3 or 4 score are deemed to have met standard. The results show that, with the exception of math performance among "disadvantaged" students (low-income students, English language learners, students with a disability, and non-Asian minorities), a higher percentage of students had level 2 scores than level 1 scores; however, math scores were distributed more evenly between levels 1 and 2 compared with reading or writing scores. This pattern indicates that more students were "far misses" in math than in reading or writing.

As mentioned, for four groups—low-income students, English language learners, students with at least one disability, and non-Asian minorities—a higher percentage of students had level 1 scores than level 2 scores in math. For example, 18.6 percent of students with at least one documented disability received a level 2 score in math compared with 67.8 percent of disabled students with a level 1 math score.

Exhibit 22
Level of Performance Among Students Who Did Not Meet Standard on the WASL, by Student Characteristics
Class of 2009

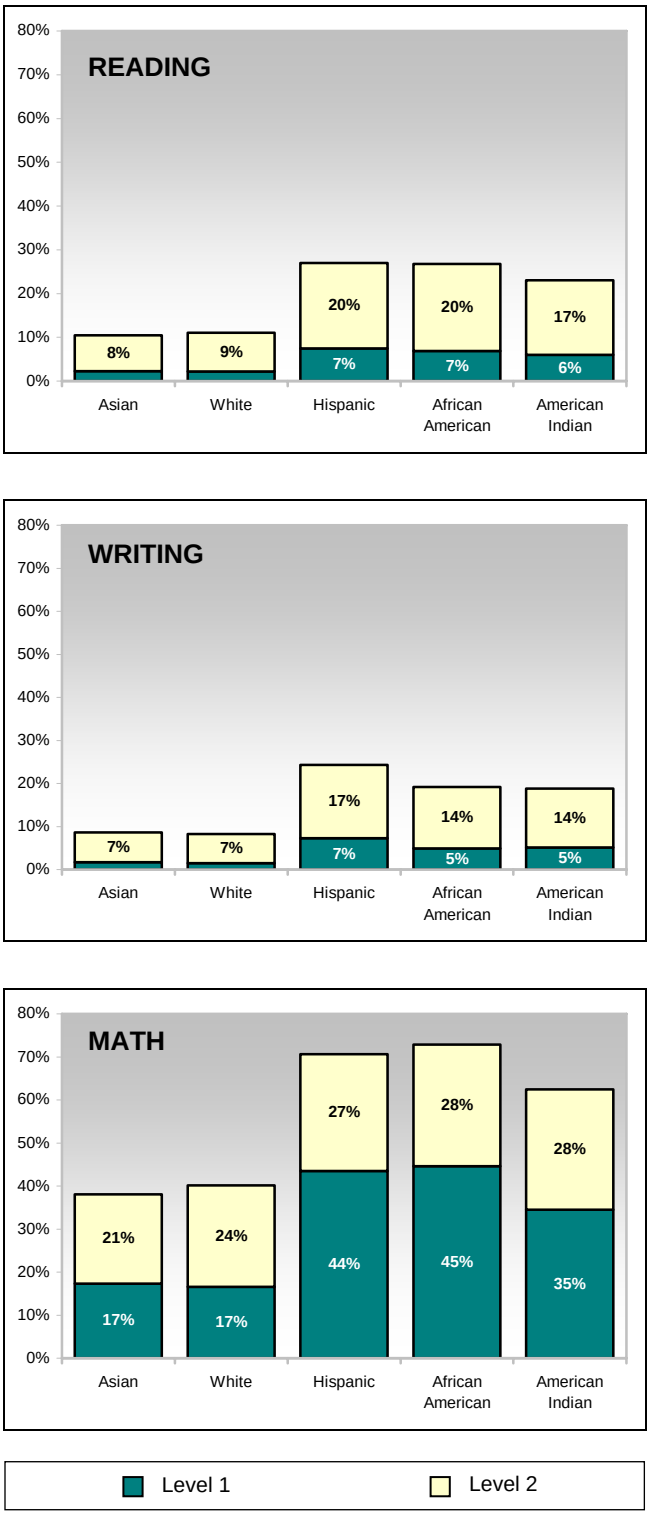
Student Characteristics	WASL Level	WASL Subjects		
		Reading	Writing	Math
All Students	Level 1	3.1	2.4	21.4
	Level 2	10.7	8.4	24.1
Gender				
Female	Level 1	2.0	1.4	22.1
	Level 2	8.6	5.2	25.5
Male	Level 1	4.1	3.4	20.6
	Level 2	12.7	11.5	22.6
Income				
Low Income	Level 1	6.1	5.1	35.6
	Level 2	17.5	14.4	27.7
Not Low Income	Level 1	1.6	1.0	14.2
	Level 2	7.2	5.4	22.3
Language				
ELL	Level 1	16.8	18.2	63.7
	Level 2	31.9	30.8	20.2
Not ELL	Level 1	2.4	1.6	19.3
	Level 2	9.6	7.3	24.3
Disability				
Disability	Level 1	20.6	16.9	67.8
	Level 2	33.8	31.5	18.6
No Disability	Level 1	1.9	1.5	18.3
	Level 2	9.1	6.9	24.4

[19] R. Barnoski & W. Cole. (2007). *Tenth-grade WASL performance in spring 2006: How individual student characteristics are associated with performance.* Olympia: Washington State Institute for Public Policy, Document No. 07-02-2201.

Exhibit 23

Level of Performance Among Students Who Did Not Meet Standard on the WASL, by Race/Ethnicity

Class of 2009



As we did for WASL performance among students in the class of 2008, we conducted a logistic regression analysis of met-standard rates for students in the class of 2009. The standardized logistic regression coefficients in Exhibit 24 describe the net association of each student characteristic with WASL met-standard rates—that is, the "independent" association of each characteristic after accounting for all other characteristics in the analysis. The analysis is restricted to students who completed the relevant WASL assessment(s) in spring 2007.

Once again, the statistics reported in Exhibit 24 indicate the relative strength of each characteristic's association with met-standard rates: the larger the absolute value, the stronger its association with meeting standard. Positive and negative values indicate that the characteristic is associated, respectively, with increased and decreased rates.

GPA has the strongest association with WASL performance: on average, as GPA increases, the likelihood of meeting standard on the WASL correspondingly increases. Even after accounting for GPA, however, several student characteristics are associated with meeting standard on the WASL.

After GPA, two characteristics—disability status and receipt of bilingual instruction—have the strongest association with WASL performance, although the strength of these associations is roughly half that of GPA. The association of these characteristics with WASL performance is negative: on average, being an English language learner or having a disability reduces the likelihood of meeting standard in reading, writing, and math, even after adjusting for gender, race/ethnicity, income, and GPA.

Eligibility for free/reduced-price meal benefits is also consistently associated with lower met-standard rates, although the strength of this association is about half that of ELL or disability status.

The relative association between race/ethnicity and WASL performance is weak compared with GPA, disability status, and receipt of bilingual instruction. Nevertheless, the results show that African American and Hispanic students continue to have somewhat lower met-standard rates in math, even when other factors are controlled.

Exhibit 24
Relative Association of Student Characteristics
With WASL Met-Standard Rates
Class of 2009

Student Characteristics	WASL Subjects				
	Reading	Writing	Math	Reading and Writing	Reading, Writing, and Math
Female	+2	+11	-22	+5	-19
African American	-7	-5	-12	-6	-11
American Indian	-2	-3	-3	-3	-3
Hispanic	-1	-2	-10	-1	-8
Asian	0	0	-1	0	-1
Low Income	-11	-12	-11	-12	-11
ELL	-21	-23	-19	-23	-22
Disability	-27	-27	-28	-26	-27
GPA	+57	+49	+78	+58	+80
AUC	0.842	0.837	0.852	0.838	0.851

Standardized logistic regression coefficients (times 100).

WASL performance by gender varies across subjects. Holding other characteristics constant, met-standard rates for female students are slightly higher in reading, substantially higher in writing, and much lower in math than the corresponding met-standard rates for male students.

The AUC statistic, which measures how well the student characteristics in our analysis account for meeting standard on the WASL, are in the 0.800s range. AUCs of this magnitude indicate that, in combination, the student characteristics in Exhibit 24 are strongly predictive of WASL performance. Knowing a student's gender, racial/ethnic background, family income, language use, disability status, and GPA allows one to make a reasonably strong prediction about his or her performance on the WASL.

D. Conclusions Regarding the Statistical Analysis of the WASL

Based on adjustments the Legislature made to Washington's high school assessment system during the 2007 session, **we estimate that between 84 and 87 percent of students in the class of 2008 have satisfied the WASL graduation requirement** (i.e., met standard in reading and writing as well as completed the math assessment).

- As of spring 2007, **approximately 64 percent of students had satisfied the requirements needed to graduate with a Certificate of Academic Achievement** (CAA) by meeting standard in reading, writing, and math.

- An additional **23 percent of students were on track to graduate without a CAA**, either because they met standard in reading and writing but not math, or because they met standard using one or more alternate assessments for students in special education. Students in special education who meet standard on an alternate assessment rather than on the "regular" WASL earn a Certificate of Individual Achievement (CIA).

The **Certificate of Academic Achievement** (CAA) and **Certificate of Individual Achievement** (CIA) "tell families, schools, businesses, and colleges that an individual student has mastered a minimum set of reading, writing, and math skills by graduation."

These certificates are earned by meeting standard in reading, writing, and math on the WASL (for the CAA) or the WAAS (for the CIA). The CIA is intended for students in special education who are unable to take the WASL.

Until 2013, students can earn a high school diploma without a CAA or CIA if they meet the state's reading and writing standards and also earn additional math credits.

Source: Office of Superintendent of Public Instruction
www.k12.wa.us/graduationrequirements/ CAA-CIA.aspx

We reiterate that these results slightly underestimate the percentage of students who have satisfied the WASL graduation requirement because they do not include students who met standard using an approved alternative assessment (i.e., a Collection of Evidence (COE); SAT, ACT, or AP scores; or the subject-area GPA cohort option). Section II describes the number of students who have met standard on an alternative assessment.

Conversely, the results may overestimate the percentage of students who are eligible to graduate because they do not include information as to whether other high school graduation requirements—the culminating project, a high school and beyond plan, and credit requirements—have been met.

Students in the class of 2009 were scheduled to take the WASL for the first time in spring 2007. **Roughly three-quarters (75.6 percent) of students in the class of 2009 satisfied the WASL graduation requirement** after only one attempt at taking the WASL. A smaller proportion of students—49.2 percent—satisfied the requirements for a CAA by meeting standard in reading, writing, and math.

Nevertheless, for students in both the classes of 2008 and 2009, **completion and met-standard rates continue to vary by student characteristics**. Met-standard rates in particular are lower for African American, American Indian, and Hispanic students relative to Asian and White students; for low-income students relative to their higher-income peers; for English language learners relative to English-speaking students; and for students with one or more disabilities relative to non-disabled students.

SECTION II: ALTERNATIVE ASSESSMENT OPTIONS

The 2006 Legislature directed the Institute to "explore options to augment the current system of assessments to provide additional opportunities for students to demonstrate that they have met the state learning standards."[20] In its interim report on this topic, the Institute reviewed alternative assessment options using criteria established in the legislative study assignment.[21]

This section focuses on alternative assessment options. We expand the findings of the interim report and present results for students who participated in an alternative assessment during the 2006–07 school year. New findings regarding the cultural appropriateness of student assessment options, the initial implementation of the Collection of Evidence (COE), and "multiple measures" assessment systems are also included.

A. ALTERNATIVE ASSESSMENT OPTIONS FOR HIGH SCHOOL GRADUATION

This section describes the alternative assessment options available to public K–12 students in Washington State and presents data on student performance on these assessments to date.

1) Alternative Assessment Options Defined

Generally, alternative assessments can be classified into three broad categories: substitute exams, performance-based assessments, and grade-based options.

- **Substitute exams** are tests that may be taken in lieu of the regular exit exam. Students who pursue this option may substitute their scores from a different exam, usually a nationally available standardized test, for their state exit exam results. State or local policy determines how substitute exam scores correspond to scores on the state exam.

- **Performance-based assessments** allow students to complete tasks, rather than tests, to demonstrate academic achievement. These tasks may include (but are not limited to) oral presentations, written essays, and hands-on activities.

- **Grade-based options** involve the use of teacher-assigned grades in specific courses or a student's overall grade point average (GPA) as a measure of achievement. Depending on state policy, student grades may be used as a substitute for exit exam scores or in combination with exam scores. For example, an exit exam score may be indexed with the GPA to determine eligibility for graduation. This combination of assessment results is sometimes referred to as a "multiple measures" approach.[22] Multiple-measures assessment systems are discussed later in this report.

[20] SSB 6618 § 2(1). Appendix A contains the full text of the legislation.
[21] These criteria are: comparability to the content and rigor of the WASL, reliability, challenges to implementation, potential for standardization across the state, costs, and cultural appropriateness. A. Pennucci. (2007). *Alternative assessment options for high school graduation: Interim report—revised*. Olympia: Washington State Institute for Public Policy, Document No. 07-02-2202.
[22] L. Darling-Hammond, E. Rustique-Forrester, & R. Pecheone. (2005). *Multiple measures approaches to high school graduation*. Stanford, CA: Stanford University, School of Education, School Redesign Network.

2) Alternative Assessment Options in Washington State

Beginning with the class of 2008, public high school students in Washington must meet standard on the WASL or an alternative assessment to graduate. In 2006 the Legislature approved three alternative assessment options—one for each of the three categories described above—for students who did not meet standard on the 10th grade WASL twice:

- **PSAT/SAT/ACT** scores (math only), a substitute exam option;
- **Collection of Evidence (COE)**, a performance-based option; and
- **GPA Subject-Area Cohort,** a grade-based option.[23]

High school students who have taken but not met standard on the 10th-grade reading, writing, or math WASL assessments can use any of these options.

A previous report issued by the Institute reviewed the alternative assessment options as they existed under ESSB 6475 passed by the 2006 Legislature.[24] In 2007, the Legislature modified the options available to students as well as the requirements for eligibility.[25] ESSB 6023:

- Phased out the use of PSAT scores as an approved alternative for math;
- Expanded the SAT/ACT option to include reading and writing in addition to math;
- Permitted subject-area Advanced Placement (AP) exams to be used as alternative assessments;
- Restricted the GPA subject-area cohort option to students with a cumulative GPA of at least 3.2; and
- Adjusted the eligibility requirements so that, effective July 22, 2007, students who do not meet standard on the WASL once, rather than twice, may pursue an approved alternative assessment option.

The alternative assessment options currently available to Washington students are summarized below.

Washington's Substitute Exams: SAT, ACT, and AP. A student who does not meet standard on the reading, writing, or math WASL can substitute his or her score from the relevant section of the SAT or ACT.[26] The State Board of Education (SBE) sets cut scores based on an OSPI analysis of the SAT and WASL results.[27] Cut scores are the minimum test score considered to "meet standard." School districts submit a form to OSPI to document students who become eligible to graduate under this option; OSPI also provides a form that allows students to obtain fee waivers for taking the exams.[28]

[23] ESSB 6475, Chapter 115, Laws of 2006.
[24] Pennucci (2007).
[25] ESSB 6023.
[26] Students can use SAT or ACT scores obtained either prior to or after taking the WASL.
[27] J. Willhoft. (2006). *Using mathematics portion of SAT, ACT, or PSAT as an alternative for the Certificate of Academic Achievement.* Olympia: Office of Superintendent of Public Instruction. See <www.sbe.wa.gov/meetings/lastmeeting/nov06/SAT-ACT-PSATcuts.ppt>.
[28] For a copy of the documentation form, see: <http://www.k12.wa.us/assessment/CAAoptions/pubdocs/1632.pdf>; for a copy of the fee waiver form, see <http://www.k12.wa.us/assessment/CAAoptions/pubdocs/1633.pdf>.

The 2007 Legislature authorized AP exams as alternative assessments. AP exams in calculus or statistics may be used as an alternative to the math WASL; AP exams in English language and composition may be used as an alternative to the writing WASL; and AP exams in English literature and composition, macroeconomics, microeconomics, psychology, U.S. history, world history, U.S. government and politics, or comparative government and politics may be used as an alternative to the reading WASL. AP exam scores range from 1 to 5; students must earn a score of 3 or higher to qualify.[29] The cut scores for Washington's substitute exams are displayed in Exhibit 25.[30]

Exhibit 25
Cut Scores for Washington State Substitute Exams Approved as Alternative Assessment Options

Exam	Scale	Reading Cut Score	Writing Cut Score	Math Cut Score
PSAT*	20–80	N/A	N/A	47
SAT	200–800	350	380	470
ACT	1–36	13	Not yet set	19
AP	1–5	3	3	3

*No longer an option after August 2008.

Washington's Performance-Based Assessment: Collection of Evidence. Collections of Evidence (COE) are collections of classroom work samples prepared by students. Teachers oversee the compilation process. OSPI developed the reading, writing, and math content guidelines and administrative protocols for COE implementation.[31] The COE is scored by a panel of trained public school teachers under contract with OSPI.

The legislation that authorizes the COE option contains special provisions for career and technical education (CTE) students.[32] CTE student work samples must be relevant to a CTE program and address the Essential Academic Learning Requirements (EALRs). Students using this option must also attain a state or nationally recognized industry certificate or credential. Implementation of the COE option is described in greater detail later in this report.

Washington's Grade-Based Option: GPA Subject-Area Cohort. Under this option, a student's GPA in English/language arts or math classes is compared with the average GPA of six or more students who attended the same school, took the same subject-area courses, and met or slightly exceeded standard on the WASL. If the student's subject-area GPA is equal to or greater than the other students' average GPA, the student is deemed to have met standard in that subject area. Students must earn a cumulative GPA of 3.2 or higher to be eligible for this option. Guidelines have been adopted in Chapter 392-501 of the Washington Administrative Code (WAC). OSPI developed two electronic tools—an MS Access database and an MS Excel spreadsheet—to assist districts in their calculations.[33]

[29] An AP exam score of "5" represents A-level work in a relevant college course, a "4" is equivalent to mid-A to mid-B college work, and a "3" represents mid-B to mid-C work. College Board. (2007). *Advanced Placement report to the nation 2007*. New York: College Board, pp. 1, 3.
[30] For more information visit <http://www.k12.wa.us/assessment/CAAoptions/Equivalency.aspx>.
[31] These guidelines and protocols were approved by SBE on October 27, 2006. Subject area guidelines and examples can be found at <http://www.k12.wa.us/assessment/CAAoptions/CollectionofEvidence.aspx>.
[32] ESSB 6475 § 1 (6)(b).
[33] For more information, visit <http://www.k12.wa.us/assessment/CAAoptions/Comparison.aspx>.

3) Washington State Alternative Assessment Results in 2006–07

As of the end of October 2007, 612 Washington students met standard by taking an alternative assessment:[34]

- 354 (58 percent) by submitting a COE.
- 180 (29 percent) using the SAT/PSAT/ACT option.
  - ✓ 26 students took the SAT,
  - ✓ 7 students took the ACT, and
  - ✓ 147 students took the PSAT.[35]
- 79 (13 percent) using the GPA Subject-Area Cohort option.

Exhibit 26 displays these results by subject area. Most students (584, or 95 percent) who used an alternative assessment in 2006–07 did so to meet the math standards. The state's graduation requirements were modified in 2007 to allow students to earn math class credits as an alternative option. Because students must still meet standard on reading and writing assessments, there could be a shift toward more use of reading and writing options in 2007–08.

Exhibit 26
Number of Students Who Met Standard Using an Alternative Assessment Option as of October 2007

Subject	COE	SAT/PSAT/ACT	GPA Subject-Area Cohort	Total
Reading	15	N/A	3	18
Writing	7	N/A	3	10
Math	332	180	73	585
Total	354	180	79	613

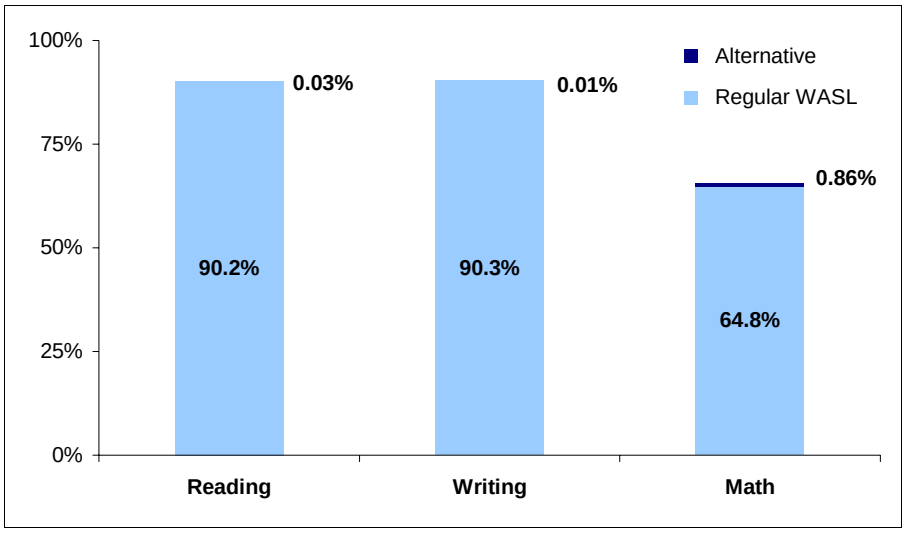
Data source: OSPI.

Impact on Overall Met-Standard Rates for the Class of 2008. To date, the number of students who met standard on an alternative assessment has had little impact on overall met-standard rates. Using the same data presented in Section I, the following analysis is based on the 67,653 students in our research database who had an assigned graduation date of 2008, were enrolled as 11th graders in a Washington public high school during the 2006–07 school year, and were slated to take the WASL in all three subjects (reading, writing, and math) in spring 2006.

After the spring 2007 WASL administration, 90 percent (n=61,002) of these 67,653 students had met standard on the reading WASL, 90 percent (n=61,073) on the writing WASL, and 65 percent (n=43,862) on the math WASL. As of the end of October 2007, fewer than 1 percent of the 67,653 students had met standard via an alternative option within each subject area (see Exhibit 27). These figures, however, do not include students who met standard on the August 2007 retake. As discussed later, the number of students using the reading and writing alternative options is expected to increase next year.

[34] As reported by the Office of Superintendent of Public Instruction. Additionally, 14 students have met standard through the score appeals process.
[35] As noted above, the PSAT option is being phased out and will not be available after August 2008.

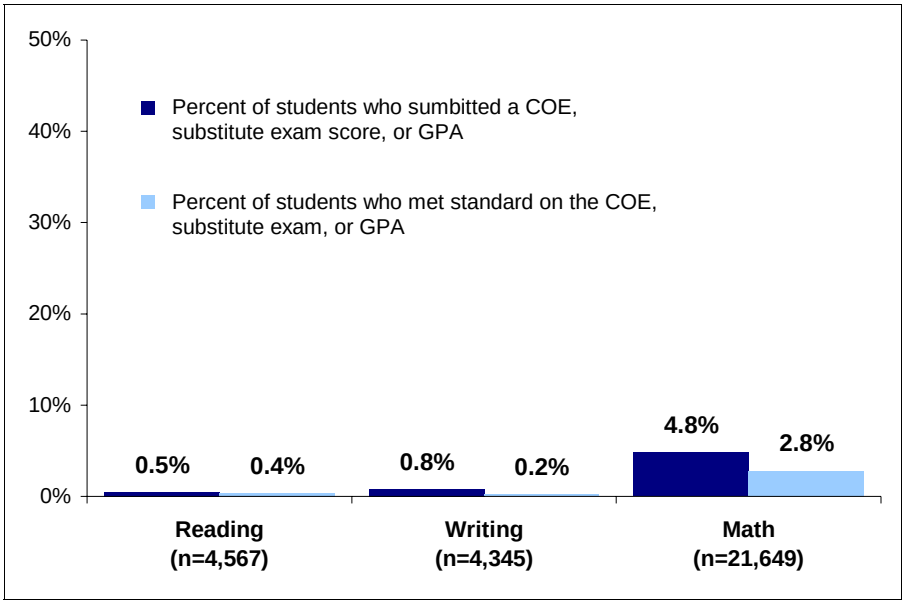
Exhibit 27
Percentage of Class of 2008 Students Who Met Standard Using an Alternative Assessment Option
As of the end of October 2007



Source: Institute analysis of OSPI data. For all three subject areas n=67,653.

Alternative assessments have thus far contributed very little to overall met-standard rates because, to date, relatively few eligible students have pursued an alternative option (see Exhibit 28). Of students who have completed but did not meet standard on the regular WASL at least once, less than 5 percent pursued an alternative option in math, and less than 1 percent in reading and writing. Again, this trend could change next year.

Exhibit 28
Percentage of Eligible* Students Who Pursued and Met Standard on an Alternative Option
As of the end of October 2007



Source: Institute analysis of OSPI data.

* "Eligible" refers to students who took but did not meet standard on the WASL. The "n" for each subject area refers to the number of students who are eligible to take an alternative assessment.

B. Findings Regarding Legislative Review Criteria

This section summarizes the Institute's review of "options to augment the current system of assessments to provide additional opportunities for students to demonstrate that they have met the state learning standards."[36]

1) Interim Report Findings

Findings from the Institute's February 2007 report, which reviewed the following alternative assessment options, are updated here:[37]

- Collection of Evidence (COE);
- College admissions and placement exams;
- Overall and subject-area cohort GPA;
- Comprehensive achievement tests; and
- Segmented math exams.

We evaluated each option according to criteria established in SSB 6618. That legislation directed the Institute to consider alternative assessment options in terms of their comparability to the content and rigor of the WASL, challenges to implementation, potential for standardization across the state, costs, and cultural appropriateness. These review criteria are described in Exhibit 29.

Exhibit 29
Legislative Review Criteria for Alternative Assessment Options

Content and Rigor Comparability: Compliance with RCW 28A.655.061(1): "alternative assessments for each content area shall be comparable in rigor to the skills and knowledge that the student must demonstrate on the Washington assessment of student learning."

Reliability in measuring a student's ability to meet state learning standards.

Challenges to implementation, including any legislative action necessary for implementation.

Whether assessment procedures or methods can be **standardized** across the state.

Costs for implementation.

The **cultural appropriateness** of each option.

SSB 6618, Laws of 2006
(See Appendix A for text of the legislation.)

We also reviewed each option according to its potential to increase met-standard rates. As SSB 6618 § 2(1) states, the Institute is to "explore options to augment the current system of assessments to provide additional opportunities for students to demonstrate that they have met the state learning standards."

The findings are summarized below.

[36] SSB 6618 § 2(1). Appendix A contains the full text of the legislation.
[37] One option the Institute has not studied is end-of-course exams, which is currently under review by the State Board of Education.

Content Comparability. The Institute's initial finding with respect to content comparability of alternative assessment options has not changed. The February 2007 report found that the "content and format of tests [that are] designed for different purposes…may not perfectly match the skills and knowledge measured by the WASL…. Unless test items are purposefully selected to align with the skills and knowledge measured by the WASL, the content match will generally be much less than 100 percent."[38]

Assessments created without explicit attention to Washington's standards will not be as aligned in content to the WASL as options intentionally developed to align with Washington's standards. The COE and segmented math exams are examples of alternative assessments developed based on the EALRs and Grade Level Expectations (GLEs).

Rigor Comparability. Regarding rigor comparability, the Institute's initial finding also has not changed: "through the process of establishing cut scores, a level of rigor comparable to the WASL can be achieved."[39] In other words, while the rigor of individual test questions may not be equal, one can determine equivalency among overall scores on different tests through analysis of students' average performance on the exams in question. This finding applies to substitute exams. For the COE, staff at OSPI established a comparable level of rigor by developing detailed guidelines that directly align the COE guidelines with the state standards.[40]

Less is known about the comparability of alternative assessment options based on students' course grades, which is an option in Washington (the GPA Subject-Area Cohort option). Local curricula, instructional practices, and other factors (e.g., the inclusion and relative weight of classroom assessments, homework assignments, and attendance) teachers use to determine grades vary. Whether a certain GPA demonstrates student mastery of state learning standards depends on the extent to which those standards are taught and assessed in each class.

As noted in the Institute's interim report, met-standard rates are unlikely to increase substantially when assessments have comparable rigor, because similar levels of academic preparation and skill are required for students to perform well.

Reliability. The alternative assessment options we reviewed are considered to be reliable: they have a sufficient number of test items to measure a student's knowledge and skills consistently. The only exception is the subject-area GPA cohort option.

The subject-area GPA cohort measure is less reliable than the overall GPA option because it is based on fewer courses and fewer students. An overall GPA option is based on <u>all</u> courses and <u>all</u> students in a given grade, whereas the subject-area GPA cohort option is based on as few as six students and may include only a handful of high school math courses.

Challenges to Implementation. Nationally available standardized tests are relatively easy to implement and have comparatively low costs. Assessment options that involve classroom work (such as the COE) or state-led test development (such as segmented math exams) are more complex and time consuming to implement. As discussed later, these implementation challenges are largely short-term and, according to educators interviewed, can be overcome with sufficient time for and attention to planning.

[38] Pennucci (2007): p. 7.
[39] Ibid, p. 6.
[40] J. Wilhoft. (2007). *Standard-setting for Collection of Evidence*. Presentation to the Washington State Board of Education August 13, 2007. Olympia: Office of Superintendent of Public Instruction.

Potential for Standardization. SSB 6618 directs the Institute to consider "[w]hether the procedures or methods could be standardized across the state." All assessment options reviewed for this study have standardized processes.[41]

Costs. Different alternative assessment options incur different costs, and there is a tradeoff between content/rigor alignment and cost. It is more costly for the state to develop a test from scratch, based on Washington's standards, than to purchase an off-the-shelf exam that comes with already-developed test items, protocols, and scoring capacity. Off-the-shelf tests, however, are not fully aligned with Washington's standards.

Exhibit 30 summarizes our findings on alternative assessment options.

Exhibit 30
Alternative Assessment Options Review: Summary of Findings

<table>
<tr><th></th><th>Comparable Rigor</th><th>Comparable Content</th><th>Reliable</th><th>Low Costs</th><th>Easy to Implement</th><th>Standardized Process</th></tr>
<tr><td colspan="7">Options with lower potential to increase met-standard rates</td></tr>
<tr><td>Nationally available, standardized tests
College admissions, placement, and credit-bearing exams: SAT, PSAT, ACT, ASSET, COMPASS, ACCUPLACER, AP, IB
Comprehensive achievement tests: ITED, NAEP, MAP
Career and Technical Education exams: Industry Certification, ACT Workkeys</td><td>Yes (based on cut score)</td><td>No (partial overlap; varies by test)</td><td>Yes</td><td>Yes</td><td>Yes</td><td>Yes</td></tr>
<tr><td>Subject-Area GPA Cohort</td><td colspan="2" rowspan="2">Depends on curriculum, instruction, and how teachers assign grades</td><td>Yes (less)</td><td>No (initial)
Yes (ongoing)</td><td>No (initial)
Yes (ongoing)</td><td>Yes</td></tr>
<tr><td>Overall GPA</td><td>Yes (more)</td><td>Yes</td><td>Yes</td><td>Yes</td></tr>
<tr><td colspan="7">Options with higher potential to increase met-standard rates</td></tr>
<tr><td>Collection of Evidence</td><td>Yes</td><td>Yes</td><td>Yes</td><td>No</td><td>No</td><td>Yes</td></tr>
<tr><td>Segmented Math exams</td><td>Yes</td><td>Yes</td><td>Yes</td><td>No (initial)
Yes (ongoing)</td><td>Yes</td><td>Yes</td></tr>
</table>

2) Cultural Appropriateness of Assessment Options

In addition to the review criteria in Exhibit 30, the 2006 Legislature directed the Institute to review the cultural appropriateness of alternative assessment options. This review reveals an achievement gap—lower average scores for certain groups than others—across all assessments for which data are available. The analysis below raises more questions about cultural appropriateness than can be answered with available information. In particular, we cannot determine precisely what factors cause these systematic achievement gaps—but it is unlikely that the cause is the assessments themselves.

We begin by defining terms. Because the concept of "cultural appropriateness" is nebulous, it is important to examine its layers of meaning and define its role in student assessment. We then present data comparing student performance on three assessments—the SAT, COE, and GPAs—with WASL performance by race/ethnicity. A discussion of the policy implications of these findings follows.

[41] Even the subject-area GPA cohort option has a standardized method for calculating average grades. The content and rigor may vary, but the process is standardized.

Conceptual Issues

An assessment is usually judged to be culturally appropriate to the extent that it is "fair" for students from different cultural groups, "neutral" with respect to their cultural differences, and hence "unbiased."

Fairness refers to the testing process, the conditions under which assessments are administered to students. Fairness can be understood in terms of equality or equity.

- ***Equality:*** An identical (i.e., standardized) test is administered uniformly to all students.
- ***Equity:*** Assessments are tailored to the specific needs and background characteristics of individual students.[42]

The SAT exemplifies a test that is generally considered to be fair in terms of equality: all test takers regardless of their characteristics complete a uniform test under tightly controlled conditions.[43] Conversely, in Washington State, the WAAS-Portfolio option for students in special education typifies an assessment that is viewed as fair in terms of equity. The individualized WAAS-Portfolio is reserved for students whose disabilities preclude them from taking the pencil-and-paper WASL.

In addition to the testing process, fairness can refer more generally to the schooling process. Students, under this criterion, should have an equal opportunity to learn the content being tested; as such, "tests of abilities need to take into account how children are schooled."[44]

Neutrality, in contrast to fairness, concerns testing outcomes. In particular, neutrality refers to predications for student performance. A test can be unconditionally or conditionally neutral.

- ***Unconditional neutrality:*** The same level of performance is predicted for students regardless of their characteristics.
- ***Conditional neutrality:*** The same level of performance is predicted for students with similar background characteristics and past levels of achievement.[45]

Test bias occurs when student performance on a test is attributable to factors other than mastery of the content or skills being assessed. Stated formally, bias refers to "a systematic error in measuring knowledge and skill."[46] With respect to the WASL and its alternative assessments, three different kinds of bias are potentially relevant: labeling bias, methodological bias, and content bias.

---

[42] T.C.M. Lam. (1995). *Fairness in performance assessment*. Greensboro, NC: ERIC Clearinghouse on Counseling and Student Services (ED391982).

[43] Nevertheless, because of unequal access to for-profit coaching classes, the SAT fails on another dimension of equality: "equal opportunity to prepare for a test." 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, p. 75.

[44] R.J. Sternberg. (2007). Who are the bright children? The cultural context of being and acting intelligent. *Educational Researcher*, *36:* 148-155.

[45] R.F. Ferguson. (1998). Teachers' perceptions and expectations and the Black-White test score gap. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap.* (Pp. 273-317). Washington, DC: Brookings Institution Press, p. 275. Ferguson also identifies a third benchmark for evaluating bias, based on unobserved potential. Students from different subgroups who have equal "potential" are predicted to perform similarly.

[46] N.S. Cole. (1997). *The ETS gender study: How females and males perform in educational settings*. Princeton, NJ: Educational Testing Service, p. 4.

- "***Labeling bias*** arises when tests claim to measure one thing but actually measure something else."
- "***Methodological bias*** arises when a test assesses mastery of some skill or body of information using a technique or method that underestimates the competence of one group relative to another."
- "***Content bias***…arises when a test claims to measure something that could in principle be measured in an unbiased way, but fails to do so because it contains questions that favor one group over another."[47]

We consider each type of bias in turn.

Labeling bias poses the following question: Does an assessment measure what it purports to measure? The WASL does not measure "aptitude" or "intelligence" per se, but rather "student learning"—in particular, mastery of Washington State's learning standards. To the extent that the WASL and its alternative assessment options are aligned with state learning standards, it does not suffer from labeling bias.[48] The question, however, then becomes whether the EALRs and GLEs are culturally biased. This question is beyond the scope of the Institute's study assignment.[49]

Methodological bias is present when the format of an assessment systematically favors particular subgroups of examinees despite comparable levels of student learning. The research as to whether some groups of students consistently perform better on multiple-choice or open-ended questions is extensive, but the results are mixed.[50]

Content bias refers to an assessment that "contains content or language that is differentially familiar to subgroups of examinees."[51] For example, some researchers allege that the method by which SAT items are selected routinely yields questions that favor whites over minority students.[52] During the past 20 years, "exhaustive efforts have been made by the Educational Testing Service (ETS) to purge such cultural biases from the SAT. Before a test is ever given to students, it is picked over with a fine-tooth comb by experts in American language and culture."[53]

Recent changes to the SAT place an "increased emphasis...on learned material rather than aptitude."[54] These changes shift the burden of cultural appropriateness from tests to schools. If tests measure learned material, then opportunities to learn are pivotal.

---

[47] The quoted passages defining the various types of bias are from: C. Jencks. (1998). Racial bias in testing. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap.* (Pp. 55-85). Washington, DC: Brookings Institution Press, pp. 55-57.

[48] See, e.g., D. Pavelchek, P. Stern, & D. Olson. (2002). *Relationship between the WASL and placement tests used by community and technical colleges.* Olympia: Social & Economic Sciences Research Center, Washington State University; College Board. (2006). Report for the State of Washington on the alignment of the SAT and PSAT/NMSQT to the Washington Essential Academic Learning Requirements. New York: College Board.

[49] On the cultural appropriateness of EALRs and GLEs, see Office of Superintendent of Public Instruction. (2003). *Bias & fairness review: Washington State education standards & grade level content expectations,* <www.k12.wa.us/CurriculumInstruct/pubdocs/ BiasFairnessReviewReadingMath.doc>.

[50] 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; Cole, 1997.

[51] R. Hambleton & J. Rodgers. (1995). Item bias review. *Practical Assessment, Research & Evaluation*, *4*. Available online at <PAREonline.net/getvn.asp?v=4&n=6>.

[52] R. Freedle. (2003). Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores. *Harvard Educational Review*, *73*: 1-43.

[53] For example, see: After many years of repair, the test content of the SAT now appears fair: The way the results are used is the problem. *Journal of Blacks in Higher Education*, *15*: 28-29, p. 28.

[54] How changes in the SAT will affect college-bound Blacks. *Journal of Blacks in Higher Education*, *36*: 12-13, p. 13.

Similarly, "[e]very WASL question goes through extensive analysis by a Bias and Cultural Fairness Committee of specially trained educators before inclusion. Before any question 'counts' on the WASL, each question goes through review to ensure there is no cultural bias in the exam. Each question also is given a trial run with students to make sure the question does not pose special difficulty for students from different backgrounds."[55]

Sometimes the review fails to identify problematic items before the WASL is administered to students. A recent example concerns a writing prompt on 10th-grade reading WASL that some Latino/a students found insensitive.[56] Some stakeholders representing different racial/ethnic groups who were consulted for this study indicated that recommendations from Bias and Cultural Fairness Committees are not always implemented.[57]

Language Barriers.[58] Most K–12 classes and assessments in the United States are conducted in English. For English language learner (ELL) students, it is extremely difficult to untangle language proficiency from the academic knowledge and skills that statewide assessments intend to measure. Also, newly arrived immigrant ELL students may be especially unfamiliar with the context of test items.[59] On average, ELL students' test scores are substantially lower than those of English speaking students.[60]

No policy or research consensus has emerged regarding the most effective way to mitigate language barriers in assessment. National assessment standards recommend that test developers focus on reducing grammatical complexity and simplifying vocabulary; in other words, "modifying the language but not the content of the test item."[61] Also, as discussed in later sections of this report, ensuring that all students are taught the knowledge and skills that are tested on assessments is critical to improving student performance.[62]

---

[55] Office of Superintendent of Public Instruction, "Frequently Asked Questions about the WASL." Available at <www.k12.wa.us/assessment/ WASL/FAQ.aspx#3>.

[56] See: <www.k12.wa.us/Communications/pressreleases2007/ReadingWASLquestion.aspx>.

[57] Several stakeholders on separate occasions raised this issue.

[58] This brief summary does not consider language proficiency tests such as the Test of English as a Foreign Language (TOEFL) or the Washington Language Proficiency Test (WLPT). These tests measure students' English language proficiency, not their academic knowledge and skills.

[59] J. Abedi, C. H. Hofstetter, & C. Lord. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research*, *74*, p. 5.

[60] For example, in 2007, 38 percent of 10th-grade ELL students met standard on the reading WASL and 11 percent met standard in math. For non-ELL students, the met-standard rates were 82 percent in reading and 56 percent in math.

[61] Abedi et al., 2004, p. 17. Dr. Jamal Abedi, an education researcher at the University of California, Davis, has written extensively on this topic and also reviewed this report. He, as well as researchers representing the American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME), recommends that test developers implement this "simple language" principle, which is commonly referred to as a component of "universal design." The universal design approach is also recommended as a policy to mitigate assessment barriers faced by special education students. For more information, see: S. Thompson, C. Johnstone, & M. Thurlow (2002). *Universal design applied to large scale assessments*. NCEO Synthesis Report 44, Minneapolis: University of Minnesota, National Center on Educational Outcomes; AERA, APA, & NCME (1999); J. Abedi. (2006). Language issues in item-development. In S. M. Downing and T. M. Haladyna (Eds.), *Handbook of test development*. New Jersey: Lawrence Erlbaum Associates; J. Abedi. (2002). Standardized achievement tests and English language learners: Psychometrics issues. Educational *Assessment 2002*, *8*: 231-257; J. Abedi, C. Boscardin, & H. Larson (2000). *Summaries of research on inclusion of students with disabilities and limited English proficient students in large-scale assessment*. Los Angeles: University of California: CRESST.

[62] This conclusion is based on the Institute's own review of the literature as well as input received by education researchers consulted for this study in the fall of 2007. The researchers consulted were Dr. Jamal Abedi (School of Education, University of California, Davis, and the National Center for Research on Evaluation, Standards, and Student Testing (CRESST)); Dr. Paul Sandifer (former director, Office of Student Performance Assessment, South Carolina Department of Education); Dr. Themla Jackson (of Foresight Consultants and the OSPI Multi-Ethnic Think Tank in Washington State); Dr. Susan Brookhart (School of Education, Duquesne University, and editor, Educational Measurement, Issues, and Practice); Dr. John Hollingsworth (DataWORKS educational consulting); and Dr. Edward Kifer (Center for Advanced Study of Assessment at Georgetown College).

Results and Analysis

In the following analysis, we examine whether (and to what extent) achievement gaps for a variety of assessment and outcome measures are consistent. Data from selected assessments are presented below. Other assessments, where available data on student performance were disaggregated by subgroup, are included in Appendix C.

For the sake of illustration, we limit our discussion to student performance by race/ethnicity. It is important to note that "culture" is neither reducible to nor synonymous with racial/ethnic identities but can also include one's socioeconomic class,[63] immigrant status,[64] membership in the Deaf community,[65] and other group affiliations. Moreover, the racial/ethnic designations used in this report—an artifact of the way educational data are typically collected—are not monolithic. "African American," "American Indian," "Asian," "Hispanic," and "White" are pan-ethnic categories that comprise many different subgroups, each with distinct cultures.[66] Analyzing student performance using these categories nevertheless provides a useful starting point for addressing the cultural appropriateness of different assessment options.

Our analysis shows that there is a racial/ethnic achievement gap on every assessment for which we have data. Although we are unable to determine the cause of these achievement gaps, their consistency across a variety of student outcomes suggests either that the assessments are culturally inappropriate in broadly similar ways or that factors other than cultural inappropriateness of the assessments are responsible.

WASL Performance by Race/Ethnicity. Exhibit 31 presents a summary of WASL performance by race/ethnicity over time. Although met-standard rates have increased for every racial/ethnic group in recent years, a persistent achievement gap separates Asian and White students from American Indian, African American, and Hispanic students. This disparity is especially pronounced for math. Between 1999 and 2006, the met-standard rate for Asian students exceeded that for White students by an average of 2.7 percentage points. Conversely, compared with White students, average met-standard rates were 23.9, 28.5, and 30.7 percentage points lower, respectively, for American Indian, Hispanic, and African American students.

[63] B. Bernstein. (1977). Social class, language and socialization. In J. Karabel & A. H. Halsey (Eds.), *Power and ideology in education*. (Pp. 473-487). New York: Oxford University Press; P. Bourdieu & J. Passeron. (1990). *Reproduction in education, society and culture* (2nd ed.). London: Sage; A. Lareau. (2002). Invisible inequality: Social class and childrearing in Black families and White families. *American Sociological Review*, *67*: 747-776; P. DiMaggio. (1982). Cultural capital and school success: The impact of status culture participation on the grades of U.S. high school students. *American Sociological Review*, *47*: 189-201.

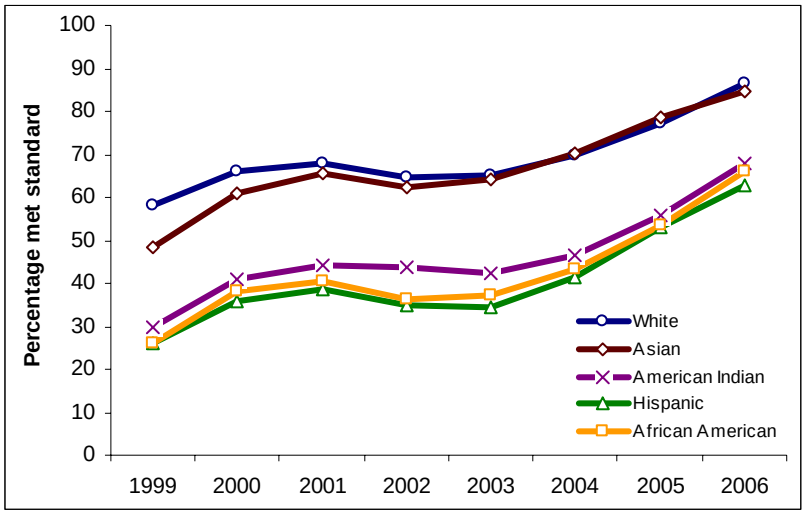
[64] J. Ogbu. (1987). Variability in minority school performance: A problem in search of an explanation. *Anthropology and Education Quarterly*, *18*: 312-334; A. Portes & R. G. Rumbaut. (2001). *Legacies: The story of the immigrant second generation*. Berkeley: University of California Press.

[65] B. McLain & A. Pennucci. (2002). *Washington School for the Deaf: Models of education and service delivery*. Olympia: Washington State Institute for Public Policy, Document No. 02-06-2202; J. Gilliam & S. Easterbrooks. (1997). Educating children who are deaf or hard of hearing: Residential life, ASL, and Deaf culture. ERIC Digest #558 (ED414676); C. Padden & T. Humphries. (1988). *Deaf in America: Voices from a culture*. Cambridge: Harvard University Press.

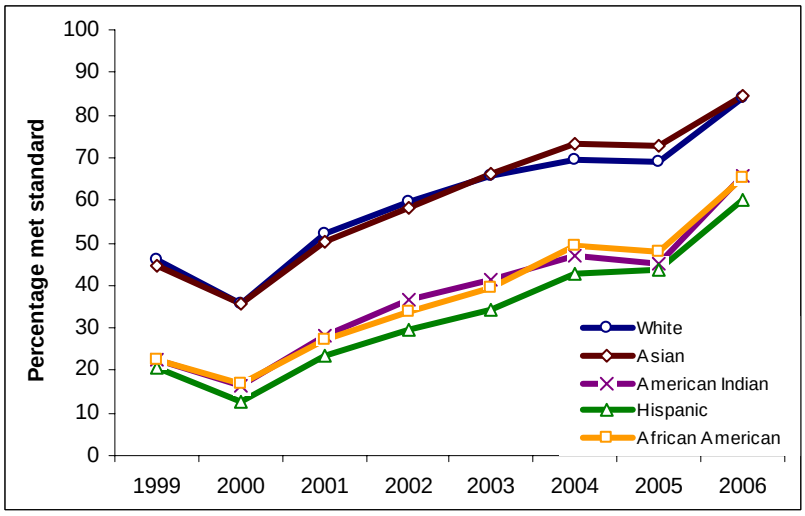
[66] See, e.g., S. Cornell & D. Hartmann. (1998). *Ethnicity and race: Making identities in a changing world*. Thousand Oaks, CA: Pine Forge Press; M.C. Waters. (2000). *Black identities: West Indian immigrant dreams and American realities*. Cambridge: Harvard University Press; M.C. Waters. (1990). *Ethnic options: Choosing identities in America*. Berkeley: University of California Press; Y.L. Espiritu. (1992). *Asian American panethnicity: Bridging institutions and identities*. Philadelphia: Temple University Press.

Exhibit 31
10th-Grade WASL Met-Standard Rates by Race/Ethnicity, 1999–2006

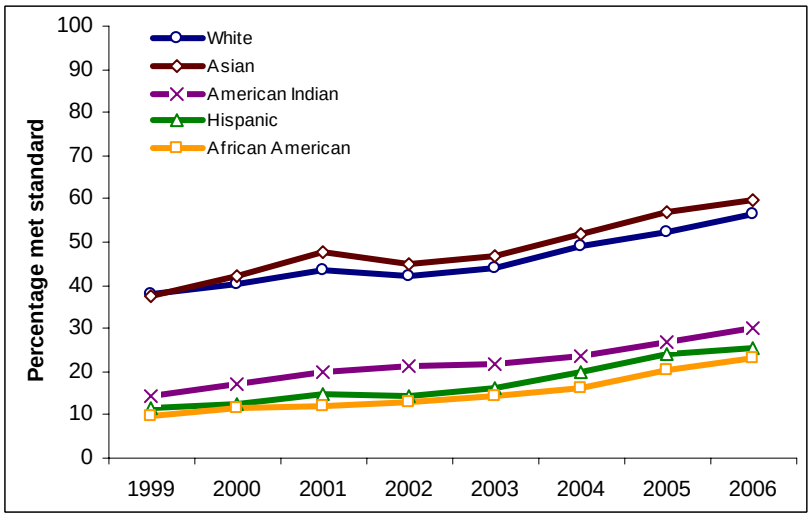
Reading



Writing



Math



Source: OSPI

In previous reports, the Institute analyzed the relative strength of associations between various student characteristics, including race/ethnicity, and met-standard rates in spring 2006.[67] These analyses statistically matched students on relevant background characteristics and prior academic performance to isolate the association between race/ethnicity and WASL performance.

After taking a student's gender, socioeconomic status, family background, language skills, study habits, and GPA into account, the racial/ethnic gap in WASL scores becomes smaller.[68] For example, once other demographic factors and GPA are controlled, the gap between White and Hispanic met-standard rates decreases from 30 to 13 percentage points.[69]

Comparison With Other Assessments. How do WASL performance gaps compare with other assessment and outcome measures? This section focuses on achievement gaps with respect to the SAT and student GPAs, both of which are currently approved for use as alternative assessment options. Appendix C reports student performance trends by race/ethnicity for a variety of additional assessments and outcomes (e.g., standardized tests, statewide assessments, grades, and graduation and dropout rates), including other approved alternatives such as the ACT and AP exams.

SAT. In its interim report on alternative assessments, the Institute concluded that standardized exams such as the SAT have a low potential to increase WASL met-standard rates because of the moderately strong correlation between WASL performance and SAT scores. On average, students who do poorly on the WASL will also do poorly on the SAT.[70] Our analysis suggests that this conclusion likely holds within subgroups of students as well.

Exhibit 32 presents average SAT verbal, math, and writing scores by race/ethnicity nationwide over time. As with the WASL, Asian and White students receive higher scores on the SAT than American Indian, Hispanic, and African American students. However, trends in SAT performance contrast with WASL performance in one respect.[71] Unlike the achievement gap on the WASL—with American Indian, Hispanic, and African American student performance differing collectively from White and Asian students—achievement gaps on the verbal and math portions of the SAT are less polarized.[72]

Between 1987 and 2006, Asian test takers had higher verbal and math scores than American Indian students, who in turn received higher scores than Hispanic students. African Americans had the lowest average SAT scores of any racial/ethnic group. White students earned higher SAT-Verbal scores than Asian students, while Asian students earned higher SAT-Math scores than White students.

---

[67] W. Cole & R. Barnoski. (2007). *Tenth-grade WASL in spring 2006: Association between poverty and WASL performance by race/ethnicity*. Olympia: Washington State Institute for Public Policy, Document No. 07-01-2205; W. Cole & R. Barnoski. (2007). *Tenth-grade WASL in spring 2006: Relative strength of associations between student characteristics and met-standard rates*. Olympia: Washington State Institute for Public Policy, Document No. 07-01-2206.

[68] A similar finding is reported in L.V. Hedges & A. Nowell. (1999). Changes in the Black–White gap in achievement test scores. *Sociology of Education*, *72:* 111-135. Hedges and Nowell concluded that the achievement gap between Black and White students on seven achievement tests was only partially reduced when social class, family background, and type of community were controlled in statistical analyses.

[69] Moreover, an analysis of WASL performance for students in the class of 2008 found that race/ethnicity accounts for a very small proportion of the explained variance in reading, writing, and math scores—approximately 0.2 to 1.0 percent—after considering gender, poverty, language barriers, disabilities, and GPA. An analysis of the ACT found that race/ethnicity accounts for approximately 1 to 2 percent of the variance in ACT scores controlling for GPA, coursework, parents' education, language barriers, and school attended. See J. NobleM. Davenport, J. Schiel, & M. Pommerich. (1999). *Relationships between the noncognitive characteristics, high school course work and grades, and test scores for ACT-tested students*. ACT Research Report Series 99-4. <http://www.act.org/research/reports/pdf/ACT_RR99-04.pdf>.

[70] Pennucci (2007).

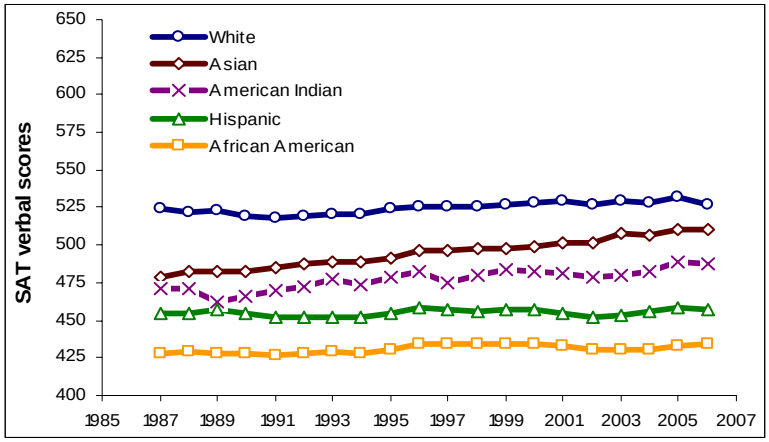
[71] Comparisons of this sort should only be made with the understanding that the SAT is "taken only by particular groups of self-selected students. Therefore, aggregate results of their performance on these tests usually do not reflect the educational attainment of all students in a school, district, or state." Most students who take the SAT intend to enroll in college and are therefore not representative of all students. See <http://www.collegeboard.com/about/news_info/cbsenior/yr2006/related.html>.

[72] SAT achievement gaps by race/ethnicity should be interpreted with caution because many test takers do not report their race/ethnicity. See: D. Whittington. (2004). The achievement gap: Should we rely on SAT scores to tell us anything about it? *Education Policy Analysis Archives*, *12*(12). Available at: <http:epaa.asu.edu/epaa/v12n12/>.

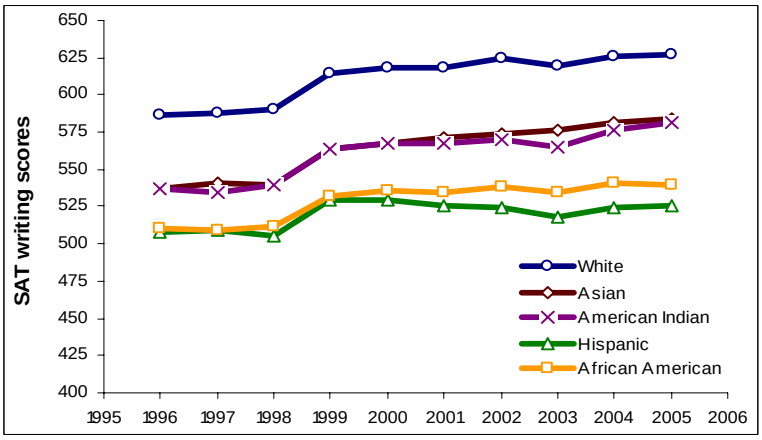
Unlike the verbal and math portions of the SAT, performance on the SAT-Writing assessment is characterized by a three-way achievement gap. White students outperformed Asian and American Indian test takers, followed by African American and Hispanic students. In 2005, the College Board revised the SAT-Writing assessment to include an essay as well as a multiple-choice component. Exhibit C.1 in Appendix C charts average SAT essay scores nationwide by race/ethnicity in 2006. Average scores for Asian and White students were identical (7.5 out of 12), followed by American Indian (7.0), Hispanic (6.8), and African American (6.5) students.

Exhibit 32
SAT Scores by Race/Ethnicity, United States

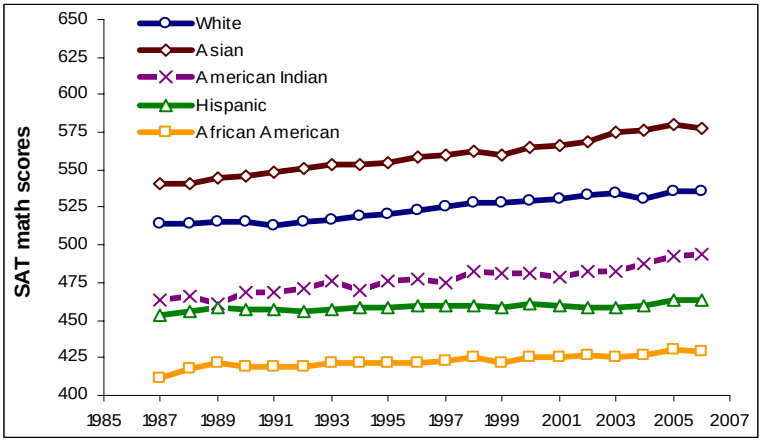
SAT-Verbal Scores (1987–2006)



SAT-Writing Scores (1996–2005)



SAT-Math Scores (1987–2006)

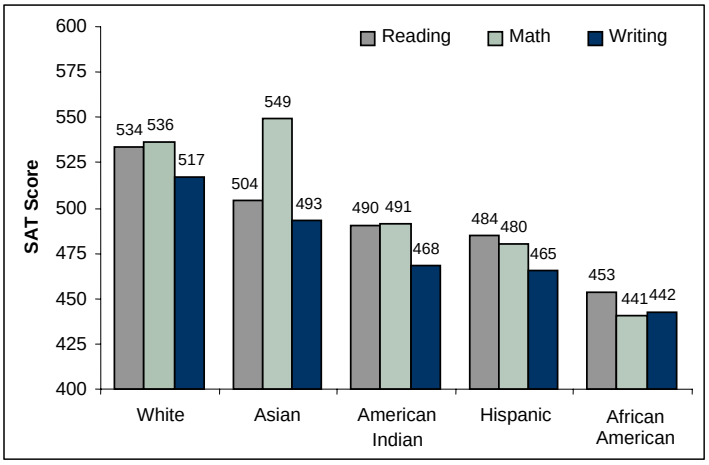


Source: J.L. Korbin, V. Sathy, & E.J. Shaw. (2007). *A historical view of subgroup performance differences on the SAT reasoning test*. New York: The College Board (Research Report No. 2006-5).

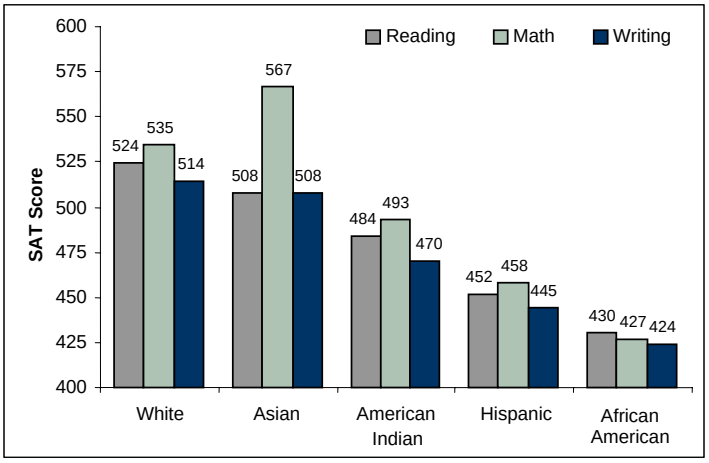
Exhibit 33 compares average SAT scores by race/ethnicity for students in Washington State with student performance nationwide in 2006. Overall, SAT scores in Washington exceed the national average, but performance varies by race/ethnicity. White test takers in Washington surpass their counterparts in the United States by an average of approximately 10 points on the critical reading component of the SAT but perform at the national average in math and writing. Compared with their peers nationwide, Asian students in Washington perform below the national average in reading, writing, and math, as do American Indians in math and writing. Conversely, Hispanic and African American test takers in Washington do much better than the national average for their respective racial/ethnic groups, in some cases outperforming their counterparts nationwide by more than 20 points.

Exhibit 33
SAT Scores by Race/Ethnicity,
Public School Students, 2006

Washington State



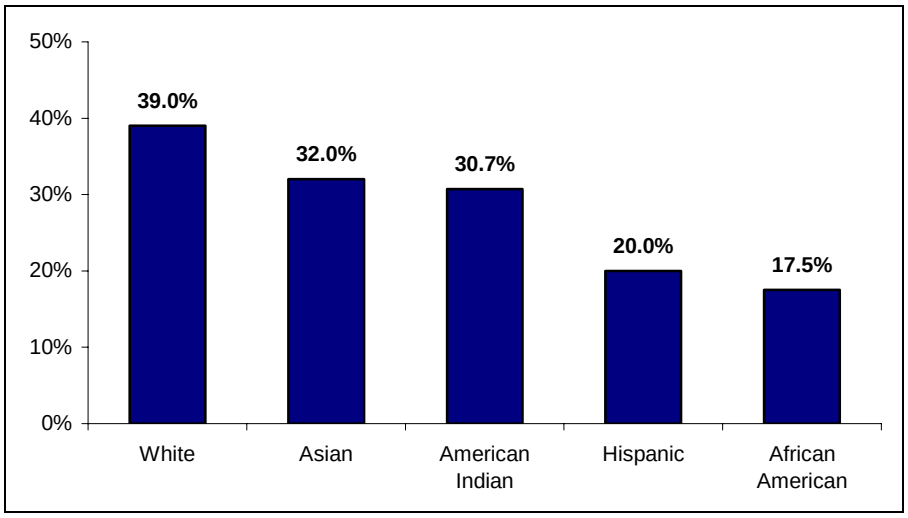
United States



Source: OSPI <www.k12.wa.us/ Communications/pressreleases2006/ SAT2006.aspx>.

Exhibit 34 depicts the percentage of students by race/ethnicity who did not meet standard on the math WASL in 2002 and 2003 and who would have been eligible to substitute SAT-Math scores based on the current cut score (470). Thirty-nine percent of White students who did not meet standard in math had SAT-Math scores above the minimum threshold (n=3,367) compared with 32 percent of Asian (n=435), 31 percent of American Indian (n=63), 20 percent of Hispanic (n=162), and 18 percent of African American (n=174) students.

Exhibit 34
Percentage of Students Who Did Not Meet Standard on the Math WASL With SAT-Math Scores at or Above the Cut Score



Source: Institute analysis of OSPI data.

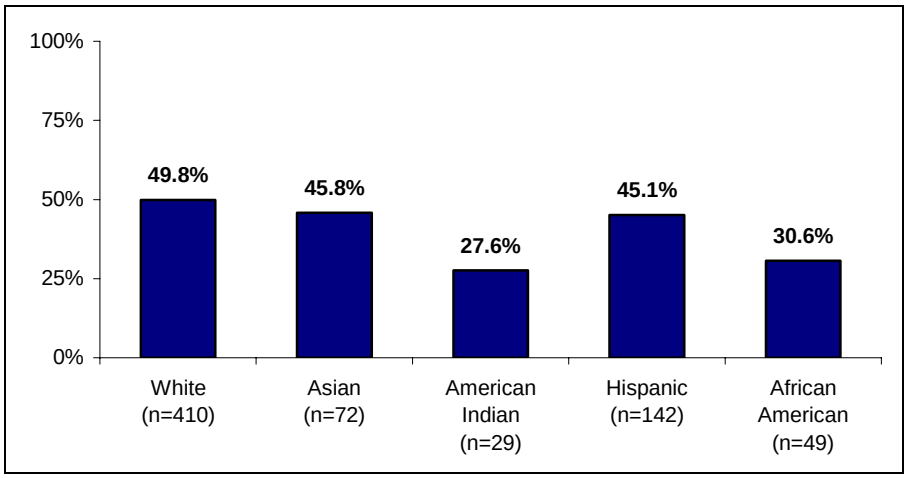
***Collection of Evidence*.** Are racial/ethnic achievement gaps less pronounced for performance-based assessments than for standardized assessments? Raymond Reyes, Associate Vice President for Diversity and former professor of education at Gonzaga University, suggests that performance-based or "authentic" assessments such as portfolios, presentations, speeches, experiments, and essays "are closer to traditional Indian experience than taking standardized tests."[73]

However, other education researchers warn that performance-based assessments are susceptible to the same kinds of cultural biases typically attributed to selected-response tests and, therefore, require a comparable degree of scrutiny.[74]

While we could not locate subgroup results for any performance-based assessments in other states, Washington's recent implementation of the Collection of Evidence (COE) provides an initial look at student performance. Exhibit 35 displays met-standard rates, by ethnicity, for 10th-grade math COEs submitted in June 2007.[75] Of students who submitted a COE in June 2007, 46 percent met standard. American Indian and African American students have lower met-standard rates, but, in contrast with other assessments, Hispanic students' math COE met-standard rates are comparable to Asian and White students' rates.

[73] R. Reyes. (1998). *A Native perspective on the school reform movement.* Portland, OR: Northwest Regional Educational Laboratory.
[74] American Educational Research Association, 1999, p. 74; R.L. Linn, E.L. Baker & S.B. Dunbar. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, *20*: 15-21.
[75] Too few reading and writing collections were submitted to permit subgroup analysis.

Exhibit 35
Math COE Met-Standard Rates by Race/Ethnicity
Washington State, 2007



Source: Institute analysis of OSPI data.

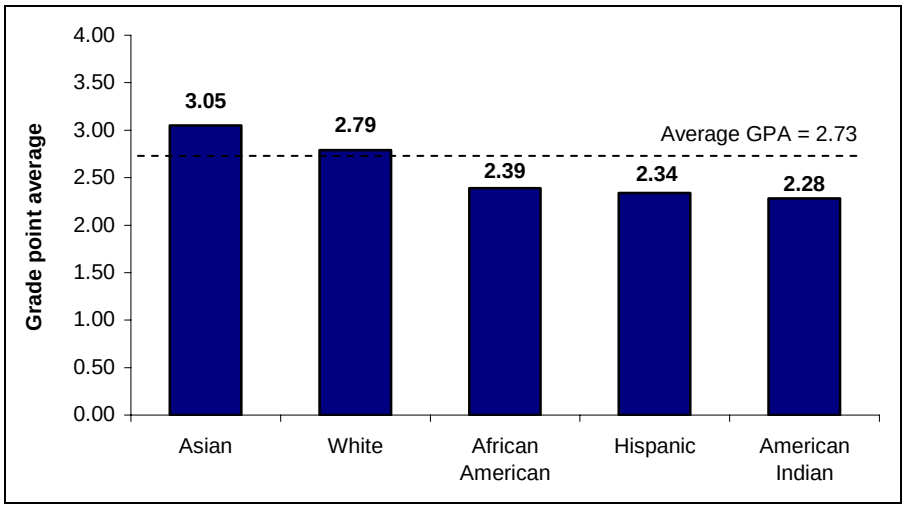
Similarly, in 2007, English language learner (ELL) students in Washington:

- Were more likely to submit a math COE than non-ELL students who registered (60 percent compared with 40 percent); and
- Had a higher met-standard rate than non-ELL students (53 percent compared with 46 percent).

These COE results should be interpreted with caution because they are based on small numbers of students in the first year of implementation. Trends may not hold in other subject areas or if greater numbers of students submit COEs.

Grade Point Average. ESSB 6023 mandates that students earn a grade point average (GPA) of at least a 3.2 to be eligible for the grade-based cohort option. Exhibit 36 displays GPAs for Washington State 10th graders in spring 2006. Asian students had the highest GPAs, well above the overall average of 2.73. As the largest group, White students had GPAs nearest the average. Average GPAs were substantially lower for African Americans, followed by Hispanics and American Indians.

Exhibit 36
Student GPA by Race/Ethnicity,
Washington State 10th Graders, Spring 2006

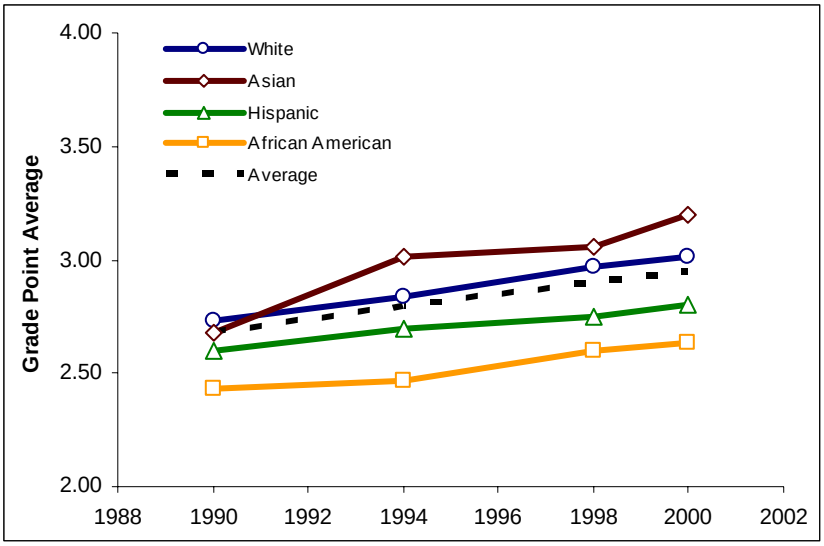


Source: OSPI, Core Student Record System

Average GPAs by race/ethnicity for high school graduates nationwide, depicted in Exhibit 37, differ slightly from Washington. GPAs for each group increase over time, but the relative performance of different groups remains stable. Except for 1990, Asian students had higher average GPAs than White students, followed by Hispanics. African Americans had the lowest GPA of any racial/ethnic group.

In 2000, the average GPA for Asian students was 3.20, followed by Whites (3.01), Hispanics (2.80), and African Americans (2.63). The national average in 2000 was 2.94. Data for American Indian students were not available. Research suggests that the "upward drift" in average high school GPAs between 1990 and 2000 reflects grade inflation.[76]

Exhibit 37
National Mean GPA of High School Graduates, by Race/Ethnicity, 1990, 1994, 1998, and 2000



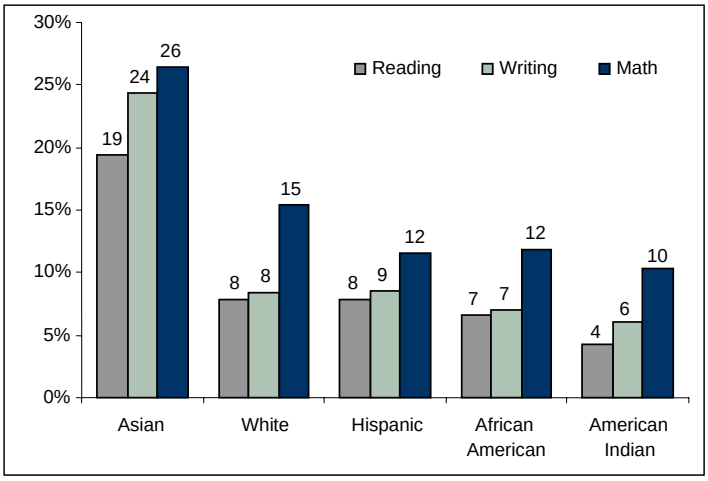
Source: NCES. 2004. *The high school transcript study: A decade of change in curricula and achievement, 1990-2000*. Washington, DC.

Exhibit 38 charts the percentage of students by race/ethnicity and subject area who did not meet standard on the WASL in spring 2006 and who had GPAs above the minimum threshold of 3.2. These percentages give a rough indication of how many students may be eligible to substitute their grades for the WASL.

- 19 percent of Asian students who did not meet standard on the reading WASL had a GPA of 3.2 or higher compared with 24 percent of Asian students who did not meet standard in writing and 26 percent who did not meet standard in math.
- At the other end of the spectrum, only 4 percent of American Indian students who did not meet standard in reading had a GPA at or above the eligibility threshold compared with 6 percent of American Indian students who did not meet standard in writing and 10 percent who did not meet standard in math.

[76] S. Alon & M. Tienda. (2007). Diversity, opportunity, and the shifting meritocracy in higher education. *American Sociological Review*, *72:* 487-511, p. 490.

Exhibit 38
Percentage of Students Who Did Not Meet Standard on a Subject-Area WASL With GPAs at or Above the Eligibility Threshold, Spring 2006



Source: Institute analysis of OSPI data.

Conclusions Regarding Cultural Appropriateness

A fairly consistent and persistent racial/ethnic achievement gap characterizes performance on a variety of assessments and outcome measures. In other words, performance differentials are not unique to any one method of assessment; rather, broadly similar gaps in achievement are replicated across a variety of measures over time. Several other studies have documented similar racial/ethnic group differences across standardized tests.[77]

Directly measuring the "cultural appropriateness" of assessments is possible in theory though likely not in practice. One could infer that an assessment is culturally inappropriate by controlling statistically for all other relevant factors (e.g., "ability," motivation, past achievement, demographics, family background, socioeconomic status, cumulative teaching experiences, cumulative curricular exposure, school-level characteristics, community-level characteristics, and so on). The unexplained residual in a regression analysis of test scores on all these factors, assuming they could be observed and measured, might then be attributed to the cultural inappropriateness of an assessment. To our knowledge, however, there is no direct way to measure cultural inappropriateness.

The mere presence of group-based differences in achievement does not by itself constitute evidence for the existence of bias.[78] Indeed, the term "bias" does not properly apply to situations in which "observed differences are not an error but a correct reflection of differences that occur on many different types of measures, in many different samples of students."[79] Thus, ***allegations that any one assessment is biased become less credible to the extent that achievement gaps are replicated across a variety of educational and assessment outcomes*.**

[77] J.L. Kobrin, V. Sathy, & E.J. Shaw. (2007). *A historical view of subgroup performance differences on the SAT Reasoning test*. New York: College Board (Research Report No. 2006-5); W.J. Camara & A.E. Schmidt. (1999). *Group differences in standardized testing and social stratification*. New York: College Entrance Examination Board (College Board Report No. 99-5); College Board. (1999). *Reaching the top: A report to the National Task Force on Minority High Achievement.* New York: College Board.

[78] R.L. Linn. (1990). Admissions testing: Recommended uses, validity, differential prediction, and coaching. *Applied Measurement in Education*, *3*: 297-318.

[79] Cole (1997): p. 4. See also College Board, 1999.

A related issue concerns the efficacy of alternative assessments in boosting met-standard rates. Alternative assessments are intended to "provide additional opportunities for students to demonstrate that they have met the state learning standards."[80] However, if alternatives merely reproduce the same racial/ethnic achievement gaps that characterize WASL performance, their potential to increase met-standard rates will be negligible. Students with low WASL performance would likely perform similarly on other assessments, without some instructional interventions.

The consistency and pervasiveness of achievement gaps suggests that assessments are not biased *per se*; rather, cultural biases may be much more deeply embedded in the structure of the educational system itself. Differences in achievement could be due to non-K–12 factors such as community characteristics or family influence, but they could also be due to students' differential opportunities to learn. More information about students' learning experiences—in particular, the alignment of curriculum and instruction with learning standards and assessments—is needed.

The alignment of curriculum and instruction with standards is a central theme emphasized by district and school staff, other stakeholders, and national experts consulted for this study, as discussed throughout the remainder of this report.

3) Collection of Evidence Implementation

The Collection of Evidence (COE) is Washington State's performance-based alternative assessment option that students can use as an alternative to the 10th-grade reading, writing, and math WASL scores for meeting proficiency requirements. This section describes the COE's first year of implementation in the 2006–07 school year, based on the following:

- Interviews with 23 district and school staff and 15 teachers (see Appendix B);
- Consultations with OSPI staff overseeing implementation;
- Observation of the scoring process and a sample of collections; and
- Analysis of data provided by OSPI.

Key findings from this implementation study follow. The COE option:

- Was initially implemented in 2006–07 with 727 submissions and 354 students (46 percent) meeting standard. Based on information provided in the interviews, the number of reading and writing COEs submitted in 2007–08 may increase substantially.
- Was complex to implement, but educators view the challenges as temporary, due to the short time frame for initial implementation.
- Requires time-intensive involvement of teachers, who reportedly benefit from training focused on the COE specifically as well as the state's learning standards generally.
- Tends to be targeted to level 2 students (those with scores near the threshold for meeting standard on the WASL).
- Appears to have manageable implementation costs at the school level but higher-than-anticipated scoring costs at the state level.

The need for alignment among curriculum, instruction, learning standards, and the COE assessment option was a recurring theme in the interviews.

[80] SSB 6618 § 2(1).

COE Participation and Results

As noted earlier, the COE was approved by the Washington State Legislature in spring 2006. The first year of implementation was during the 2006-07 school year. The first submission deadline was June 15, 2007. Sufficiency review took place from June 18 to 22. Scoring, range finding, and standard setting took five weeks, and scores were released in September 2007.

COE results should be interpreted with the following limitations in mind. The COE was not fully implemented in 2006–07, and trends could change. Additionally, it is difficult to generalize from these findings given the small numbers of students in some categories. This caution applies to reading and writing COE results in particular because of the small numbers of these COEs submitted.

Exhibit 39 presents data on the number of students, by subject area, who:

- **Registered** to submit a COE;
- **Submitted** a COE;
- Submitted a **sufficient** COE; and
- **Met standard** on the COE.

Exhibit 39
Washington State COE Participation and Results, 2006–07 School Year

	Math	Reading	Writing	Total
Registered	1,812	257	255	2,324
Submitted	727	18	31	776
Sufficient	718	18	31	767
Met standard	332	15	7	354
Percentage submitted of those registered	40.1%	7.0%	12.2%	33.4%
Percentage sufficient of those submitted	98.8%	100%	100%	98.8%
Percentage met standard of those sufficient	46.2%	83.3%	22.6%	46.2%

Source: Institute analysis of OSPI data.

COE Registrations. In the 2006–07 school year, 2,009 Washington students registered to submit 2,324 collections. (Some students signed up to complete a COE in more than one subject area). Most of these registrations (78 percent) were in math. Overall, the number of students registered to submit a COE represented 8 percent of eligible students in math (1,812 of 21,649 students); 6 percent in reading (257 of 4,567); and 6 percent in writing (255 of 4,345).[81]

COE Submissions. Of COEs registered, about one-third (33 percent) were submitted by the June 15, 2007, deadline. According to educators interviewed for this study, the two-thirds of students who registered for, but did not submit, a COE did so because they:

- met standard on the WASL between registering and the submission deadline;
- decided not to submit in response to the change in math requirements;
- decided to pursue a different alternative assessment option;
- chose not to complete the collection for other reasons.

We do not know how many students fall into each category; these are the observations of a small sample of educators interviewed for this study.

[81] The number of eligible students is based on the number of students who have taken the relevant subject-area WASL at least once without meeting standard, out of: students slated to take the WASL, enrolled in 11th-grade in 2006–07, and with an assigned graduation date of 2008.

Sufficient COEs. A "sufficient" COE meets all the OSPI-developed requirements—for instance, for math, 8 to 12 work samples covering a minimum range of content and process strands, among other criteria. Exhibit 40 summarizes the COE math sufficiency guidelines. Collections that do not meet these guidelines are classified as "insufficient" by state reviewers and are not scored. The guidelines were developed to ensure that COEs are valid and reliable measures of student learning in relation to the EALRs.

Exhibit 40
OSPI Guidelines for Sufficient Math Collections

The collection must include:

- At least 8 and no more than 12 separate work samples that together demonstrate understanding and application of the mathematics content and process strands (as defined in the EALRs and as assessed on the 10th-grade WASL).
- Work samples that demonstrate at least two different targets from each content strand.
- Work samples that demonstrate certain key targets.

Characteristics of the 8–12 work samples:

- Each must combine at least one entire content target and one entire process target.
- At least two must be produced "on-demand."
- All must include the assignment and/or problem(s).
- All must represent moderate to high complexity.
- All must include evidence of student work supporting the answer or conclusion.

The collection must not include:

- Group responses to a work sample.
- Processes or content below the 10th-grade level.
- Tasks that feature only multiple-choice questions or problems that expect single or limited responses.
- Work samples with grades displayed on them.
- Scoring guides/rubrics from district or classroom assessments.
- Teacher comments that are specific instructions aimed at improving the outcome of the work sample.
- Work that has not been produced by the student.

Other required documents:

- Checklist of sign-off forms.
- Student Information Form.
- Work Sample Documentation Form.
- Work Sample Sign-off Form.
- Identification of which content and process strands are represented.
- Verification that the assignment is attached.
- Student signature verifying authenticity of work and that it represents what he/she knows and can do.
- Teacher signature supporting authenticity of work and that it represents what student knows and can do.
- Principal sign off.

Source: OSPI. The full guidelines, including those for reading and writing, as well as forms and sample tasks, can be downloaded at: <http://www.k12.wa.us/assessment/CAAoptions/CollectionofEvidence.aspx>.

Many educators who were interviewed commented that the sufficiency requirements are intricate and difficult to follow. There are many detailed requirements, and the guidelines reportedly evolved as implementation progressed. Many of those interviewed expressed concern that, because of these complexities, some students would not graduate purely due to an administrative error. However, most COEs (98.8 percent) submitted in 2006–07 were deemed sufficient. Only nine of the 727 math COEs did not meet all the requirements for a complete collection. During the sufficiency review, OSPI staff directly contacted district assessment coordinators in an effort to correct errors and increase the number of sufficient collections.

Educators interviewed anticipate that implementation complexity will decrease. Challenges are viewed as temporary—due to the newness of the option and the short timeline for initial implementation.

COE Met-Standard Rates. The reading and writing COE met-standard rates are reported in Exhibit 39: the rates are 83 and 23 percent, respectively. These rates should be interpreted with caution. The figures are based on very small numbers of students; the met-standard rates will be much more reliable after more students submit reading and writing collections.[82] Educators interviewed for this study anticipate submitting more reading and writing COEs in the 2007–08 school year, given that the reading and writing state assessment requirements remain in place.

Because over 700 math COEs were submitted and scored, we are able to report met-standard rates by subgroup using OSPI-provided COE student data. Exhibit 41 presents math COE results by gender, ethnicity, participation in the state Transitional Bilingual Instructional Program (TBIP), first language English or not, special education and disability status, and income status.

As noted earlier in this report, ELL students have promising math COE results: ELL students had higher met-standard rates than students not in the TBIP. Other notable results in Exhibit 41 are that girls have a higher math COE met-standard rate than boys (54 percent versus 38 percent) and low-income and non-low-income students have comparable rates (46 percent).

Again, these COE results should be interpreted with caution, because they are based on small numbers of students in the first year of implementation. Trends may not hold in other subject areas or if greater numbers of students submit COEs.

Exhibit 41
COE Math Results 2006–07 by Subgroup

	Registered	Submitted	Sufficient	Met Standard	Submitted, of Those Registered	Sufficient, of Those Submitted	Met Standard, of Those Sufficient
Total	1,812	727	718	332	40.1%	98.8%	**46.2%**
Girls	987	391	386	207	39.6%	98.7%	**53.6%**
Boys	796	335	331	125	42.1%	98.8%	**37.8%**
African American	149	49	49	15	32.9%	100%	**30.6%**
American Indian	54	29	29	8	53.7%	100%	**27.6%**
Asian	163	72	72	33	44.2%	100%	**45.8%**
Hispanic	309	146	142	64	47.2%	97.3%	**45.1%**
White	1,080	415	410	204	38.4%	98.8%	**49.8%**
In TBIP (ELL students)	97	58	57	30	59.8%	98.3%	**52.6%**
Not in TBIP	1,686	668	660	302	39.6%	98.8%	**45.8%**
First language not English	276	129	127	50	46.7%	98.4%	**39.4%**
First language English	1,506	597	590	281	39.6%	98.8%	**47.6%**
Special education yes	121	46	46	16	38.0%	100%	**34.8%**
Special education no	1,662	680	671	316	40.9%	98.7%	**47.1%**
Disability (any) yes	127	47	47	16	37.0%	100%	**34.0%**
Disability (any) no	1,656	679	670	316	41.0%	98.7%	**47.2%**
Low income yes	276	145	141	65	52.5%	97.2%	**46.1%**
Low income no	1,507	581	576	267	38.6%	99.1%	**46.4%**

Source: Institute analysis of OSPI data. "TBIP" stands for the state Transitional Bilingual Instructional Program.

[82] Technically, the reading and writing COEs were judged to be "proficient" rather than to have "met standard" because too few collections were submitted for these subject areas to perform a formal standard-setting process.

Role of Teachers

Assessment coordinators at the school and district levels are instrumental in assuring that COEs meet all state requirements and are sent to OSPI for scoring. Teachers play a central role in helping students assemble the work samples that make up each collection: they are responsible for working with individual students to compile work samples that meet all requirements. As noted, the weight of this responsibility was keenly felt by teachers interviewed for this study.

The 15 teachers interviewed reported that, during the COE's first year of implementation, they relied heavily on the OSPI-developed examples and guidebooks while helping students develop and select work samples. Some teachers and administrators commented that training provided by OSPI on the COE in particular, as well as the EARLs, GLEs, and WASL in general, was helpful to implement the COE option. The theme of alignment is relevant: the training and examples all help teachers ensure that instruction in general and the collections in particular are aligned with state learning standards. Many teachers and administrators described the COE as requiring explicit alignment of classroom work with the targets and strands in the EARLs and GLEs.

Target Students

According to those interviewed, districts varied in terms of which categories of students are encouraged to pursue the COE as an alternative assessment option. Many educators reported that school leaders determined which students were approached to consider the COE. Most of those interviewed indicated that level 2 students (those closest to meeting standard on the regular WASL) were encouraged to submit a COE. Due to data limitations, we are unable to report how many COEs were submitted by level 1 or level 2 students in 2007.

While a few schools focused on Career and Technical Education (CTE) students—who, as noted earlier, have special provisions in the legislation authorizing COEs—most did not, according to those interviewed.[83] The COE data do not flag these students, so we do not know how many submitted a COE. OSPI handbooks do include CTE-relevant prompts as examples.

Implementation Costs

This discussion of implementation costs is based on the interviews conducted as well as information provided by OSPI staff. This is not a systematic data analysis of costs among all districts that implemented the COE.

At the classroom, school, and district levels, COE implementation costs are generally viewed as manageable. OSPI reimburses school districts $400 for each COE judged to be sufficient.[84] Most district staff interviewed reported that they gave all or part of the reimbursement directly to the teachers who oversaw students' COE work. Some districts used the funds to support part of a staff position for coordinating the process. The state reimbursement amount was generally described as adequate. Some interviewees reported that the timing of the reimbursement is problematic because it occurs after the collections are reviewed at the state level; a few recommended that districts be reimbursed for all collections (whether sufficient or not) because each requires staff time.[85] Many educators commented that for COE implementation, teachers need more time. In the K–12 system, time, particularly for teachers to focus on individual students, is scarce.

State-level COE scoring costs, including the cost to review the sufficiency of 727 COEs and score the 718 sufficient collections during summer 2007, totaled $501,164. This amount includes the work involved in the labor-intensive scoring process as well as the administrative and technological costs involved.

[83] During interviews many educators stated that CTE students were not considered by some districts as a target population for initial implementation, due in part to the added complexity of integrating CTE and academic work into the collections.
[84] For the 2007–08 school year, the reimbursement decreased to $300 per sufficient collection.
[85] Additionally, one district administrator reported that she was unaware there was any reimbursement for the COE.

Thirty teachers—five writing scorers, five reading scorers, and 20 mathematics scorers—were hired at a rate of $125 per day to score COEs, for a total cost of $31,250. Each collection was scored by two individuals; where scores diverged, a third individual reviewed the collection to reconcile the discrepancy. The average scoring cost per COE was approximately $698; OSPI staff anticipate that for the 2007–08 school year, per-COE scoring costs will be lower (approximately $564 per collection).[86]

The Number of Reading and Writing COEs Submitted Will Likely Increase

According to those interviewed, school leaders play a key role in determining whether and how the COE is implemented. As part of this study, we interviewed administrators in five districts that did not implement the COE option in 2006–07. Among these districts, all administrators reported plans to implement the reading and writing COE in 2007–08. Interviewees in the 16 districts that participated in the first year of implementation reported plans to expand this option.

Based on these interviews, as well as the fact that the class of 2008 is the first cohort of students who must meet standard on the reading and writing portions of the WASL or an alternative assessment option to graduate, we anticipate that the number of reading and writing COEs (as well as use of other alternative options) will increase in 2007–08. Indeed, OSPI indicates that as of the end of November 2007, 5,427 students have registered to submit a COE in February 2008, more than double the number of COE registrations during the 2006–07 school year.[87]

4) Multiple Measures Assessment Systems

Multiple measures assessment systems involve using more than one source of information about individual students' academic performance when making high-stakes decisions (such as whether students may graduate from high school). The 2006 Legislature directed the Institute to review, among other options, "assessments used in other states as well as those that have been developed and those that have been proposed in Washington."[88] During the same legislative session, Washington lawmakers proposed, but did not pass, a "weighted multiple measures approach to high school graduation requirements."[89]

In this section we examine statewide multiple measures assessment systems for high school graduation. We reviewed the research literature, drawing extensively from the work of Linda Darling-Hammond of Stanford University[90] and David Conley of the University of Oregon,[91] and contracted with three educational researchers to address a set of policy questions on this topic (see inset on the next page). This section defines "multiple measures," describes the 2006 Washington proposal, and discusses a key policy consideration raised by this review.

Appendix D contains responses to the policy questions we posed, from:

- Susan Brookhart, Center for Advancing the Study of Teaching and Learning;
- John Hollingsworth, President, DataWORKS Educational Research; and
- Edward Kifer, Center for the Advanced Study of Assessment.

[86] Cost data provided by the Office of Superintendent of Public Instruction.
[87] This total includes 1,402 students registered to submit a COE in reading, 1,271 in writing, and 2,754 in math. Additionally, 29 students have registered to augment COEs that did not meet standard in 2007. Personal communication, Lesley Klenk.
[88] SSB 6618 § 2 (1)(b).
[89] HB 2923 and SB 6461.
[90] Darling-Hammond, 2005; L. Darling-Hammond, L. McCloskey, & R. Pecheone. (2006). *Analysis and recommendations for alternatives to the Washington Assessment of Student Learning*. Stanford, CA: Stanford University, School Redesign Network.
[91] The 2004 Washington State Legislature directed OSPI to "develop options for implementing objective alternative assessments" (ESHB 2195). This legislative direction initiated the comprehensive study by David Conley of the Center for Educational Policy Research. D. Conley (2006). *Study of alternative methods to the Washington Assessment of Student Learning (WASL): Feasibility study, January–August 2005, final report.* Eugene, OR: Center for Educational Policy Research.

Multiple Measures Assessment Systems: State Policy Questions

1) What does the concept of "multiple measures" mean with respect to assessments that are required for high school graduation? How might such a system be implemented?
2) What factors should policymakers consider when designing a multiple measures assessment system?
3) Is Washington's assessment system already a multiple measures system? Explain why or why not. How could Washington's system be strengthened?
4) Could a multiple measures system be implemented in a manner that is valid, reliable, cost effective, comparable in content and rigor to the WASL, culturally appropriate, and standardized across school districts?
5) Does research evidence indicate whether multiple measures would be more valid for subpopulations of students—racial/ethnic groups, low-income students, English language learners, and special education students—compared with using a single test or measure such as GPA to determine eligibility for graduation?

"Multiple Measures" Defined

Rather than using a single high-stakes exam to determine eligibility for graduation, multiple measures assessment systems use more than one source of information about individual students' academic performance—including state and national standardized tests, locally developed assessments, and classroom-based assessments such as grades and performance tasks—to make high-stakes decisions. At least 19 states, including Washington, have implemented or are in the process of developing a multiple measures system for high school graduation.[92]

According to reviews of such systems, some states use multiple measures for reasons similar to the purpose of providing alternative options: to allow students with different learning or testing styles, or different academic strengths, to demonstrate what they have learned.[93] States and districts have also developed multiple measures systems to counteract incentives for unintended consequences (for example, cheating and grade inflation)[94] and to gain a deeper, more nuanced understanding of individual student learning.[95]

In theory, multiple measures systems increase the reliability and validity of assessment systems because students are provided with "diverse opportunities to demonstrate what they have learned."[96] We found no empirical evidence of this increase, but multiple measures as a concept has broad support among assessment practitioners and researchers.[97]

[92] Darling-Hammond et al. (2005): p. 46 (footnote 17).
[93] E. Baker. (2003). Multiple measures: Toward tiered systems. *Educational Measurement, Issues, and Practice*, *22*: 13-17.
[94] D. Koretz (2003). Using multiple measures to address perverse incentives and score inflation. *Educational Measurement, Issues, and Practice*, *22*: 18-26.
[95] T. Guskey. (2007). Multiple sources of evidence: An analysis of stakeholders' perceptions of various indicators of student learning. *Educational Measurement, Issues, and Practice*, *26*: 19-27.
[96] Assuming each assessment measure is reliable and valid for state learning standards. Darling-Hammond et al, 2005, p. 13. For a detailed discussion of reliability and validity issues as they relate to multiple measures assessment systems, see: J. Ardovino, J. Hollingsworth, and S. Ybarra. (2000). *Multiple measures: Accurate ways to assess student achievement*. Thousand Oaks: Corwin Press.
[97] American Educational Research Association, 1999.

Types of Multiple Measures Systems

Multiple measures systems can be developed in a variety of ways, and there is "no clear consensus on appropriate methods of synthesizing and using information from multiple sources" of information about students' knowledge and skills.[98] Brookhart and Hollinsgworth (two of the consultants to this report; see Appendix D) identify two general types of multiple measures systems: conjunctive and compensatory. **Conjunctive** assessment systems require students to fulfill every component of high school graduation requirements—each assessment measure must be successfully completed. Washington currently has a conjunctive multiple measures system. Students must fulfill all four state requirements to graduate from high school; specifically, Washington students must:

- Earn minimum course credits in subjects required by the state and local districts; AND
- Meet standard on the WASL (or approved alternative assessments); AND
- Complete a high school and beyond plan; AND
- Complete a culminating project.

In contrast, **compensatory** multiple measures systems allow "higher performance on one measure [to] compensate for lower performance on another."[99] In one version of a compensatory system, scores on different tests can be averaged so that higher scores on one test offset low scores on another. For example, under such a system, the reading and math standards could be met by achieving an average of 400 on the reading and math WASL assessments rather than scoring at least 400 on each.[100] Alternatively, requirements could be altered so that instead of meeting all four requirements, students would need to meet three out of four (e.g., replace the "ands" above with "and/or").[101]

2006 Washington Legislative Proposal

During the 2006 legislative session, some Washington lawmakers proposed a "weighted multiple measures approach to high school graduation requirements."[102] The proposal involved changing Washington's system from a conjunctive to a compensatory model. Under the proposal, each of Washington's four graduation requirements would have been assigned a weight. The culminating project, high school and beyond plan, and WASL requirements would each count for up to 10 percent in the calculation used to determine whether students meet state learning standards overall. Each of these components would be evaluated using a four-point grading system. The fourth component (GPA) would have modified the course credit requirement by adding grade-averaged calculations; this component would carry a 70 percent weight.

Exhibit 42 on the following page details how three different types of multiple measures systems could be implemented using Washington's four high school graduation requirements, based on (1) the current conjunctive system, (2) a possible compensatory system, and (3) the 2006 legislative proposal.

[98] D. Henderson-Montero, M. Julian, & W. Yen. (2003). Multiple measures: Alternative design and analysis models. *Educational Measurement, Issues, and Practice*, *22*(2): 7.
[99] Brookhart, Appendix D, p. D-2.
[100] Scores on the writing WASL are based on a different scale.
[101] According to Brookhart, a third, hybrid type of multiple measures is called a "complementary" system in which only one of the requirements must be met, e.g., replace all of the "and's" with "or's." Appendix D, p. D-3.
[102] HB 2923 and SB 6461.

Exhibit 42
Multiple Measures Assessments Systems

Type of Multiple Measures Assessment System	Minimum Course Credits	Culminating Project	High School and Beyond Plan	Meeting Standard on the WASL*
Conjunctive Washington's current requirements	Yes required	Yes required	Yes required	Yes required
Compensatory Described by Brookhart and Hollingsworth (see Appendix D)	*A variable combination (e.g., 3 out of 4) of the following:* Yes/No	Yes/No	Yes/No	Yes/No**
Compensatory With Weights 2006 legislative proposal	70% Uses GPA*** rather than credits	10% Based on 4-point rating system	10% Based on 4-point rating system	10% Based on 4-point rating system

* In reading and writing on the regular WASL or an approved alternative.

** Could be calculated as an average exam score among subject areas rather than a yes/no pass requirement.

*** Calculated based on courses required by the state and local district for graduation; excludes electives.

Key Policy Consideration: Alignment Among Standards, Instruction, and Assessments

The Legislature directed the Institute to examine alternative assessment options in the context of providing "additional opportunities for students to demonstrate that they have met the state learning standards."[103] The key word in this sentence is "demonstrate," because assessment systems in and of themselves do not increase student achievement; rather, they measure it. According to the experts we consulted for this study, the degree of alignment among learning standards, instruction, and assessments is more important than the specific details of how multiple measures systems are implemented. Effective teaching that is aligned to grade-level learning standards is a necessary condition for increasing achievement; in the absence of alignment, no matter how many different assessments are used, student achievement will not improve. The consultants to this report wrote:

> "Do you raise test scores by testing students or by teaching them?" (Hollingsworth, p. D-14)
>
> "Multiple measures do not 'manufacture' achievement. Students should be given several opportunities to show what they know, but if they truly don't know it additional measures will not change that fact." (Brookhart, p. D-5)

This finding holds true for subgroups of students as well, according to those consulted:

> "[E]quity for student subgroups is created by having all teachers successfully teaching the same grade-level content to all students in all schools in the state." (Hollingsworth, p. D-11)
>
> "To change subgroup pass rates in large and meaningful ways, instruction as well as assessment must change." (Brookhart, p. D-5)

An additional limitation is that a multiple measures system "adds to the complexity, cost, and interpretability" of student assessment.[104] According to Kifer, features of an ideal system include using measures that are "easy to collect, represent a range of educational outcomes, provide multiple opportunities for success, and do not include much additional work for students, teachers, and schools."[105] Most important, all three consultants agreed, is for classroom instruction to be aligned to the standards that assessments are designed to measure.

[103] SSB 6618 § 2(1).
[104] Baker (2003): p. 13.
[105] Appendix D, p. D-15.

C. Conclusions and Recommendations Regarding Alternative Assessments

The Washington State Legislature has provided alternative methods for students to demonstrate their achievement of state learning goals: college admissions and college credit exams, the GPA Subject-Area Cohort option, and the Collection of Evidence (COE). In reviewing these options and the role they can play in the state's high school student assessment system, our key findings are as follows:

- As of the end of October 2007, 612 Washington students had met standard by taking an alternative assessment (180 by achieving a sufficient score on the SAT, PSAT, or ACT, 79 using the GPA option, and 354 by submitting a COE). Most of these students met standard in math.
- For an alternative assessment to match the content and rigor of the WASL, it must be geared to Washington's learning standards. "Off–the-shelf" tests are not fully aligned with Washington's standards.
- Achievement gaps among subgroups of students persist across all assessment options reviewed. This trend suggests the gap is not caused by cultural inappropriateness of the assessments but instead may be partially due to differences in students' learning experiences.
- Washington educators involved in the initial implementation of the Collection of Evidence option view the process as challenging, but manageable. In 2007, 776 students submitted a COE. The 21 school districts interviewed reported plans to expand this option in 2007–08.
- Washington currently has a multiple measures assessment system: students must fulfill multiple academic requirements to graduate. Education researchers consulted recommend that the state examine the degree to which curriculum and instruction is aligned with grade-level learning standards in order to improve student achievement across all assessment measures.

Recommendations

Interim Report Recommendations. The 2006 Legislature required that the Institute's interim report "include recommendations on at least two alternative assessment options, alternative methods, procedures, or performance measures that were reviewed."[106]

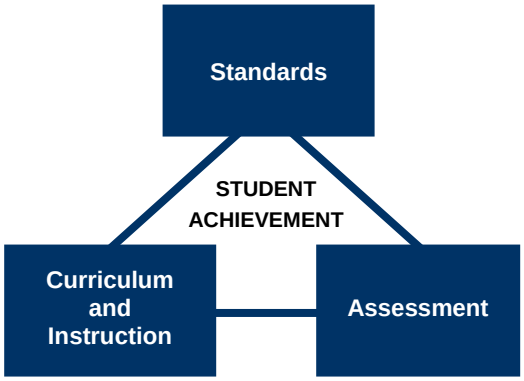
The Institute's interim report contained three initial recommendations: (1) limit the number of alternative assessment options to reduce implementation complexity, (2) balance options' potential to increase met-standard rates with their costs and implementation complexities, and (3) consider diagnostic assessments that provide feedback on student performance during the instructional process.

Our research continues to support these recommendations. Based on information provided by educators, we recommend that the total number of options be limited to reduce complexity and confusion among educators as well as students and families. When considering which additional options to authorize, a key concern (according to the legislation that directed this study) is each option's potential to increase met-standard rates. Diagnostic assessment tools are one potential way to improve student performance.

[106] SSB 6618 § 2(2).

Conclusions drawn from this final report suggest that a critical consideration in improving student learning is alignment with standards, not just for assessments, but for curriculum and instruction as well (see Exhibit 43).[107] An assessment will not appropriately measure student achievement if curriculum and instruction are not aligned with the standards upon which the assessment is based. Frequently, this concept is framed as "opportunity to learn."[108]

Exhibit 43
Alignment: A Key Factor in Student Achievement



Suggestion for Future Research. The 2006 Legislature required that the Institute's "final study shall include suggestions for any follow-up studies that the legislature could undertake to continue to build on the information obtained in this study."[109]

A clear finding from this review is the importance of alignment among standards, curriculum and instruction, and assessment. Since education reform began in the 1990s, Washington State has developed statewide learning standards and an aligned assessment system in grades 3 through 8 and grade 10. We lack information, however, about whether and to what extent current instructional practices are aligned to the state learning standards. **Therefore, as a next step, we suggest that the state undertake a systematic review of curriculum and instruction that focuses on alignment with state learning standards.**

Because individual school districts in Washington oversee curriculum and instruction, analyzing curricular alignment is a complicated task. One method available for examining curriculum alignment in a systematic and standardized fashion is the Surveys of Enacted Curriculum, a comprehensive survey of instructional practices and content administered to math, science, and English/language arts teachers. The resulting data permit researchers and educators to analyze the degree of alignment between classroom instruction, statewide standards, and assessments. Currently, 24 states (including Oregon, Idaho, and Montana) have implemented the Surveys of Enacted Curriculum.[110]

[107] Exhibit 43 focuses on factors contributing to student achievement that state policies and K–12 schools can directly influence. The model does not attempt to address other factors that impact student achievement (e.g., family and community characteristics).
[108] See, e.g., Washington State Board of Education. (2004). Appendix U: Opportunity to learn in Washington. In *Final report on the Certificate of Academic Achievement (CAA) to the Washington State Legislature, pp. 726-731*. Available at: <http://www.sbe.wa.gov/ reports/reports/CAAFinalRPT/appendix/default.htm>.
[109] SSB 6618 § 2(2).
[110] For more information on the Surveys of Enacted Curriculum, see A. Porter. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, *31*: 3-14. The surveys are administered under the auspices of the Council of Chief State School Officers. See: <http://www.ccsso.org/projects/Surveys%5Fof%5FEnacted%5FCurriculum/>.

APPENDIX A: STUDY LEGISLATION—THE GOVERNOR BOOTH GARDNER ACT (2006)

Substitute Senate Bill 6618, Chapter 352, Laws of 2006, 59th Legislature

Sec. 1. In 1993 the Washington legislature laid out a vision of a revitalized school system in Washington state. Envisioned was a comprehensive assessment system committed to high academic standards for all of its students. The Washington assessment of student learning was created as a tool to measure whether students were reaching the high academic standards. The legislature continues to support this assessment as a part of a comprehensive assessment system. Recently some alternative assessments have been developed. The legislature finds that there is interest in exploring why some students have not been able to meet the state standards and whether additional alternative methods, options, procedures, or performance measures could be used to augment the current system.

Sec. 2. (1) The Washington state institute for public policy shall conduct a study to explore options to augment the current system of assessments to provide additional opportunities for students to demonstrate that they have met the state learning standards. The study is limited to:

(a) A review and statistical analysis of Washington assessment of student learning data to increase understanding of the students who did not meet the standard in one or more areas of assessment, identify the characteristics of those students, and identify possible barriers to student success or possible causes of the lack of success;

(b) A review and identification of additional alternative assessment options that could be used to augment the current assessment system. In identifying the alternative assessment options, the institute shall include a review of alternative assessments used in other states as well as those that have been developed and those that have been proposed in Washington. The institute shall examine the use of national tests as well as career skill certification exams in their review of possible alternative assessment options.

For each of the identified alternative assessment options, the study shall at a minimum include:

(i) An estimation of the costs for implementation;

(ii) A review of the cultural appropriateness;

(iii) Whether the alternative assessment reliably measures a student's ability to meet state learning standards in one or more of the required content areas;

(iv) Whether the alternative assessment is in compliance with RCW 28A.655.061(1); and

(v) Any challenges to implementation for each of the identified alternative assessment options, including any legislative action necessary for implementation;

(c) A review and identification of additional alternative methods, procedures, or combinations of performance measures, including those proposed in Washington, to assess whether students have met the state learning standards. For each of the identified alternative methods, procedures, or performance measures, the study shall at a minimum include:

(i) An estimation of the costs for implementation;

(ii) A review of the cultural appropriateness;

(iii) Whether the method, procedure, or performance measure reliably measures a student's ability to meet state learning standards in one or more of the required content areas;

(iv) Whether the method, procedure, or performance measure is in compliance with RCW 28A.655.061(1);

(v) Any challenges to implementation for each of the identified methods, procedures, or performance measures, including any legislative action necessary for implementation; and

(vi) Whether the procedures or methods could be standardized across the state.

(2) The Washington state institute for public policy shall provide an interim report to the legislature by December 1, 2006, and a final report by December 1, 2007. The interim report shall include a preliminary statistical analysis of the information required under subsection (1)(a) of this section and shall include recommendations on at least two alternative assessment options, alternative methods, procedures, or performance measures that were reviewed under subsection (1)(b) and (c) of this section. The final study shall include suggestions for any follow-up studies that the legislature could undertake to continue to build on the information obtained in this study.

(3) The institute shall consult, at a minimum, with nationally recognized experts on assessments including representatives from recognized experts on assessments including representatives from nationally recognized centers for multicultural education, representatives of the office of the superintendent of public instruction, educators, counselors, parents, the business community, classified employees, career and technical organizations, representatives of federally recognized Washington tribes, representatives of cultural, linguistic, and racial minority groups, and the community of persons with disabilities in developing the initial list of possible alternative assessment options, alternative assessment methods, procedures, or performance measures to be reviewed under subsection (1)(b) and (c) of this section. The office of the superintendent of public instruction and school districts shall provide the institute with access to all necessary data to conduct the studies in this act.

Sec. 3. This act shall be known as the Governor Booth Gardner Act.

APPENDIX B: COLLECTION OF EVIDENCE INTERVIEWS

In the fall of 2007, the Institute contracted with the Washington State University Social and Economic Services Research Center (SESRC) to conduct interviews with district and school staff involved with implementing the Collection of Evidence (COE) in 2006–07. The WSU-SESRC conducted the interviews in October 2007. The interviews were designed to learn from K–12 educators about the initial implementation of this alternative assessment option.

As described in Exhibit B.1, 38 individuals from 21 school districts were interviewed for this study. Institute staff selected interview districts to represent diversity in terms of the following characteristics: geography, district size, WASL met-standard rates, student demographics, and degree of participation in the initial implementation of the COE (also described in Exhibit B.1). SESRC staff identified and contacted individuals to interview and provided interview tapes and notes to Institute staff.

Exhibit B.1
School Districts Represented in Interviews by 2006-07 COE Participation Status

District	Number of Individuals Interviewed
Extensive participation: most registered students submitted a COE*	
Bremerton	2
Central Valley	2
Edmonds	2
Evergreen (Clark)	2
Federal Way	1
Okanogan	3
Puyallup	3
Toppenish	1
Yelm	2
Partial participation: some registered students submitted a COE*	
Everett	3
Seattle	2
Walla Walla	2
Partial participation: some students registered, none submitted a COE*	
Aberdeen	1
Hockinson	2
Pasco	1
Tacoma	1
No participation: no students registered for or submitted a COE	
Colville	1
Granger	3
Kennewick	1
Kent	1
Vancouver	2
Total Number of Districts Interviewed	**21**
Total Number of Individuals Interviewed	**38**

* Not all registered students submitted a COE because some met standard on a retake of the WASL or a different assessment option prior to the COE submission deadline. Other registered students may have chosen not to submit for other, unknown reasons.

Exhibit B.2 lists the type of school and district staff members interviewed.

Exhibit B.2
Types of District and School Staff Interviewed

Type of Position	Number of Individuals Interviewed
Assessment Director or Specialist*	11
Assistant Principal	1
Assistant Superintendent	1
Career and Technical Education Manager	2
Chief Academic Officer	1
Counselor	1
Culminating Project Coordinator	1
Curriculum Director or Specialist	3
Literacy Coach	1
Math Coordinator	1
Teacher	15
Total Individuals Interviewed	38

* For seven of these positions, the assessment director is also the director in other areas (e.g., curriculum and instruction, research and evaluation, accountability, and/or special services).

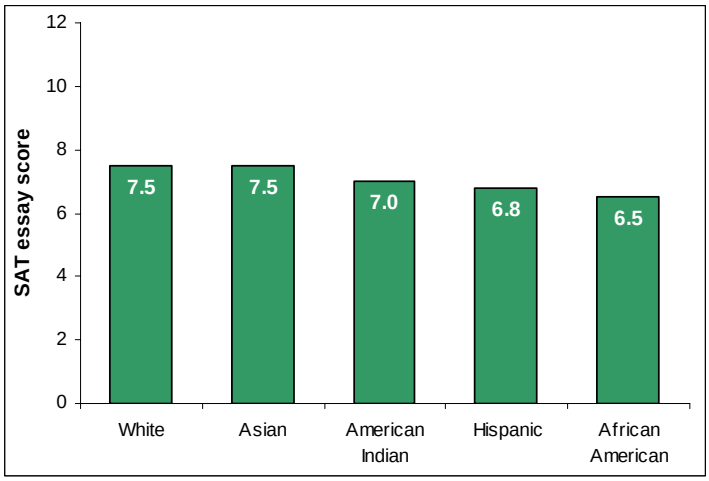
APPENDIX C: ASSESSMENT RESULTS BY RACE/ETHNICITY

This appendix presents assessment results by race/ethnicity on assessment options and achievement outcomes for which we were able to obtain subgroup data.

This compilation of data is intended to illustrate broad trends in student achievement. The results do not control for relevant characteristics that impact student outcomes (e.g., poverty, language barriers, learning disabilities). Specific differences among subgroups of students and across states should be interpreted with caution.

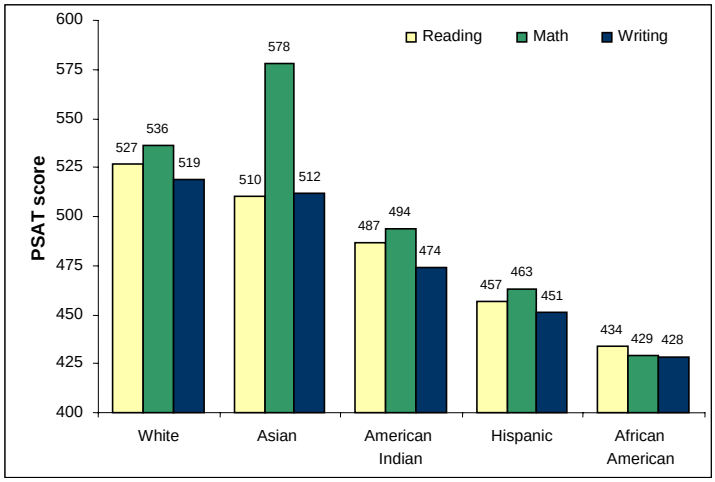
List of Exhibits

Exhibit C.1
SAT Essay Scores by Race/Ethnicity, United States, 2006



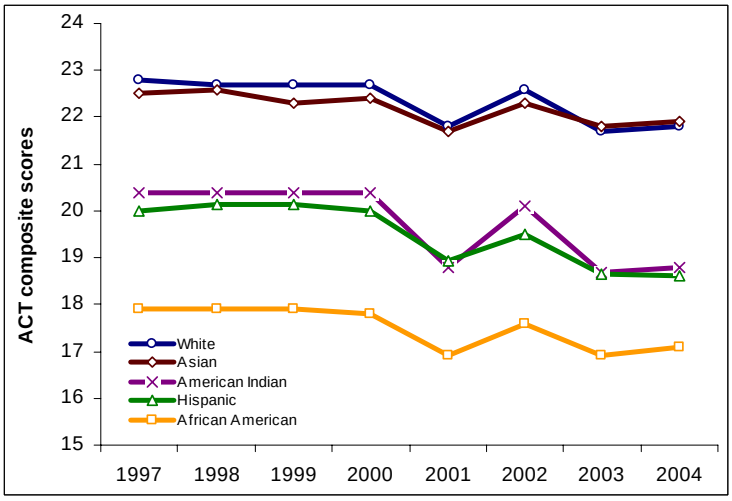
Source: K. Mattern, W. Camara, & J.L. Kobrin. (2007). *SAT Writing: An overview of research and psychometrics to date*. New York: The College Board, Office of Research and Analysis (Report No. RN-32).

Exhibit C.2
PSAT Scores by Race/Ethnicity, U.S. Public Schools, 2006



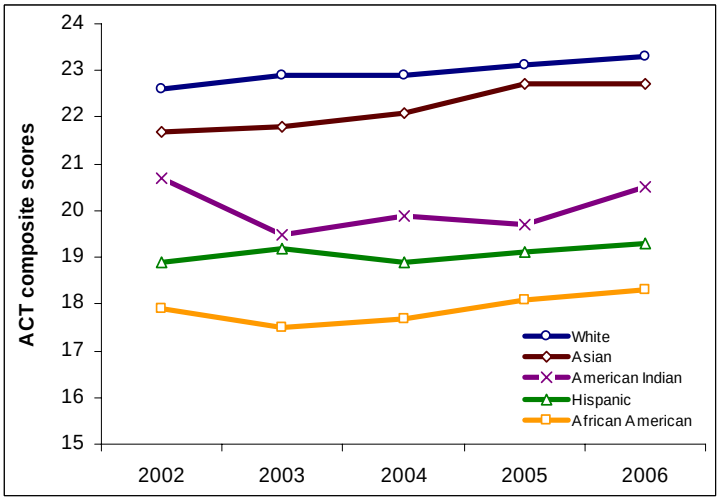
Source: Minnesota Office of Higher Education <www.mheso.state.mn.us/tPg.cfm?pageID=1442&1534-D83A_1933715A=37aefdf151fe839>.

Exhibit C.3
ACT Composite Scores by Race/Ethnicity, Washington State, 1997–2004



Source: <www.act.org/news/data/06/pdf/states/Washington.pdf>.

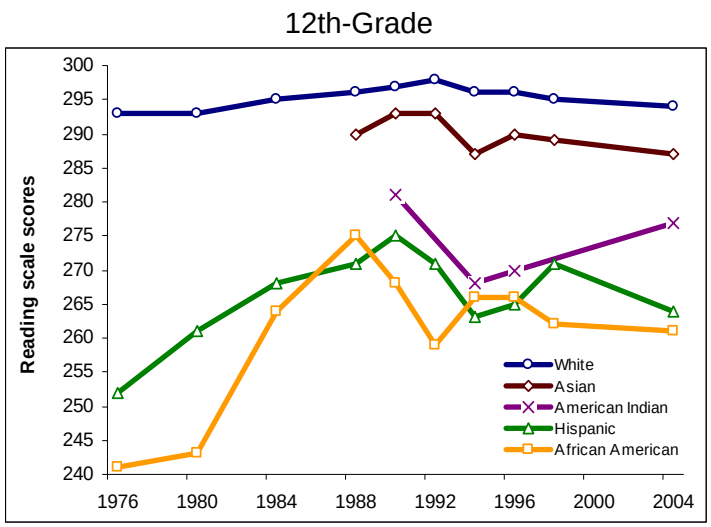
Exhibit C.4
ACT Composite Scores by Race/Ethnicity, United States, 2002–2006

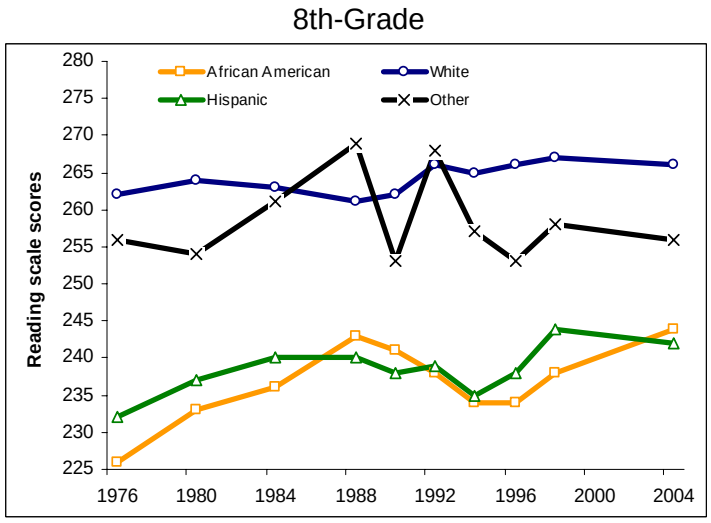


Source: <www.act.org/news/data/06/pdf/states/Washington.pdf>.

Note: Most students who take the SAT or ACT intend to enroll in college and are therefore not representative of all students. Results from these tests should therefore be interpreted with caution.

Exhibit C.5
NAEP Reading Scale Scores by Race/Ethnicity, United States, 1976–2004





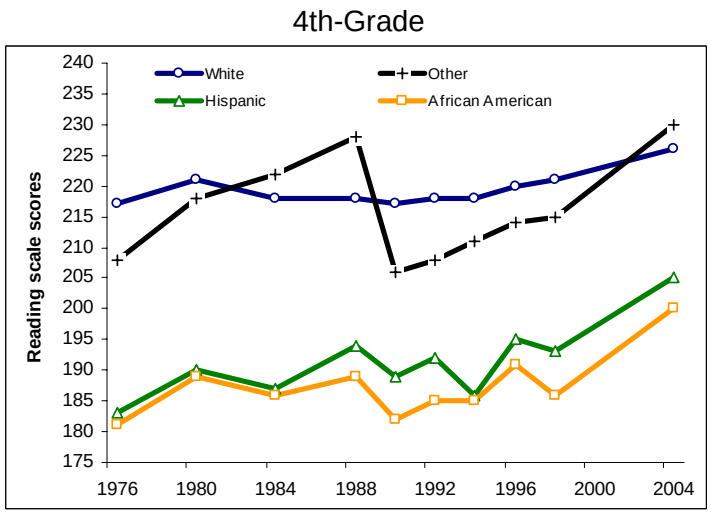
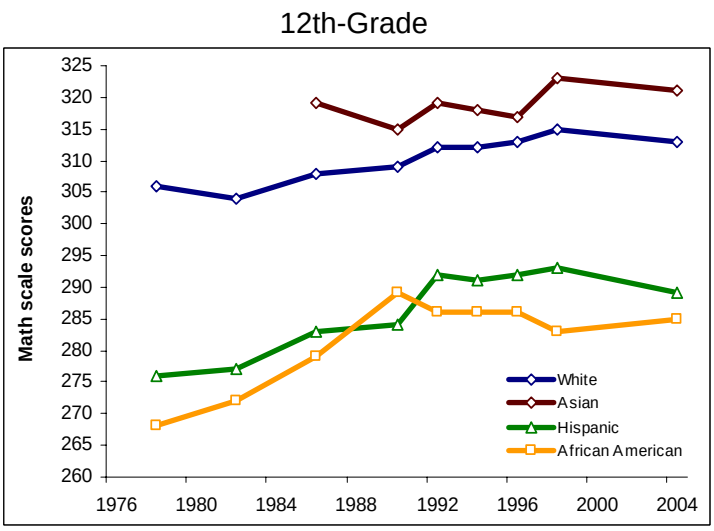
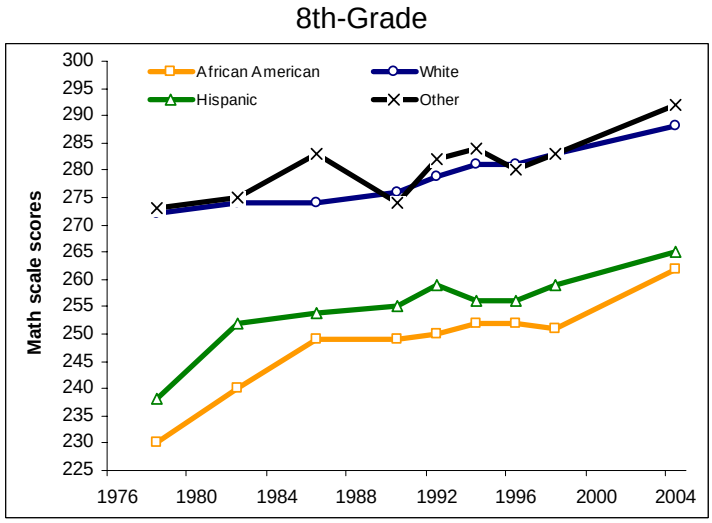
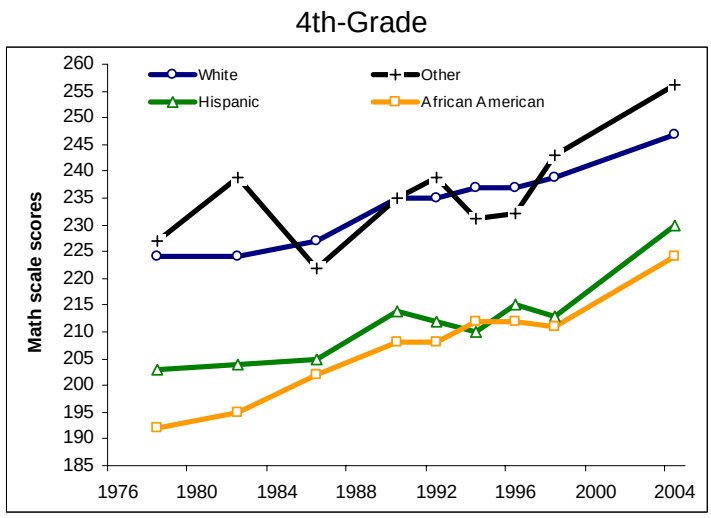


Exhibit C.6
NAEP Math Scale Scores by Race/Ethnicity, United States, 1978–2004



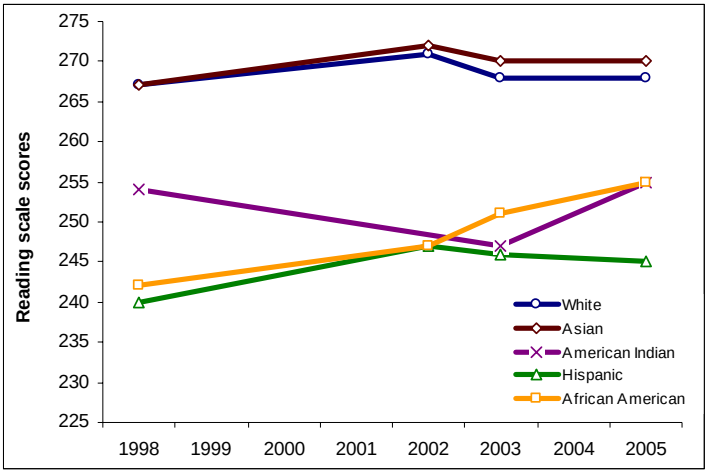




Source: <nces.ed.gov/nationsreportcard/lttnde/>.

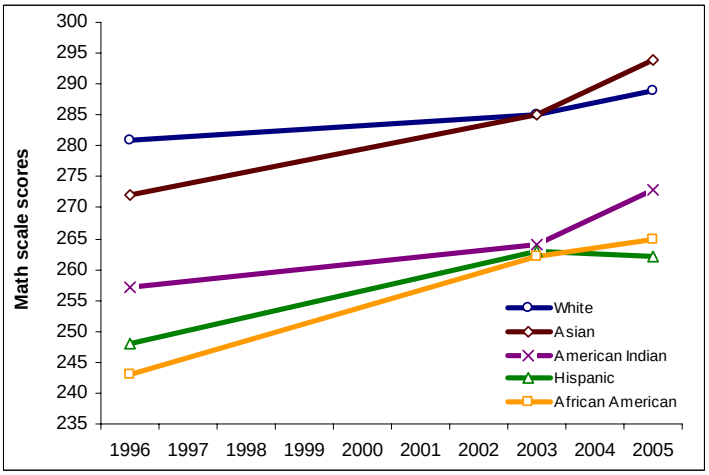
Exhibit C.7
8th-Grade NAEP Scale Scores by Race/Ethnicity, Washington State

Reading



Source:
<www.k12.wa.us/assessment/NAEP/pubdocs/2005ReportReading8.pdf>.

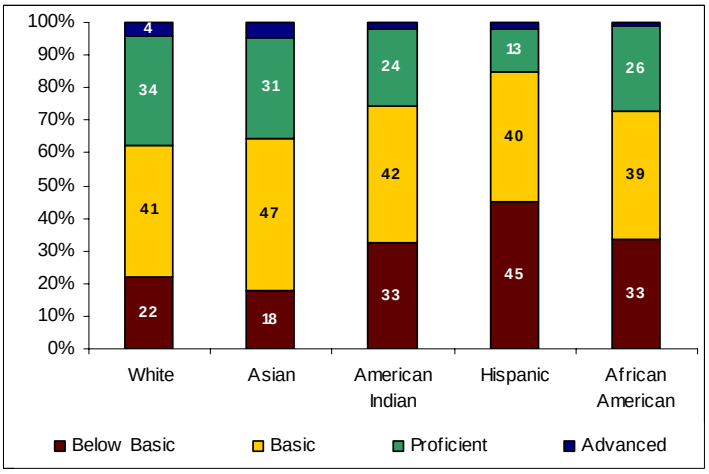
Math



Source:
<www.k12.wa.us/assessment/NAEP/pubdocs/2005ReportMath8.pdf>.

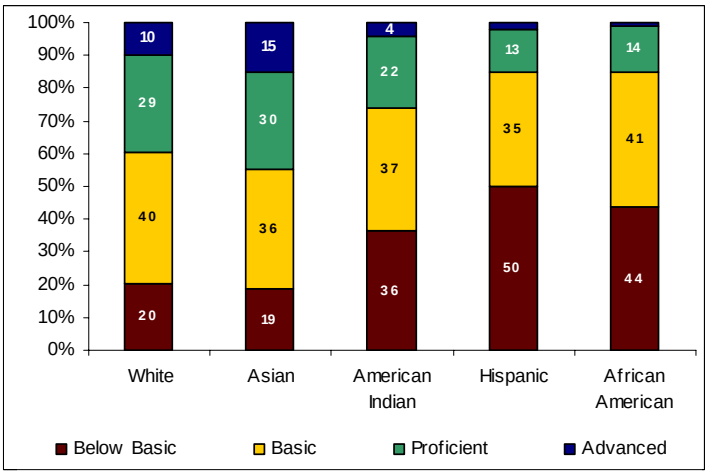
Exhibit C.8
8th-Grade NAEP Achievement Levels by Race/Ethnicity, Washington State, 2005

Reading



Source:
<www.k12.wa.us/assessment/NAEP/pubdocs/2005Grade8reading.pdf>.

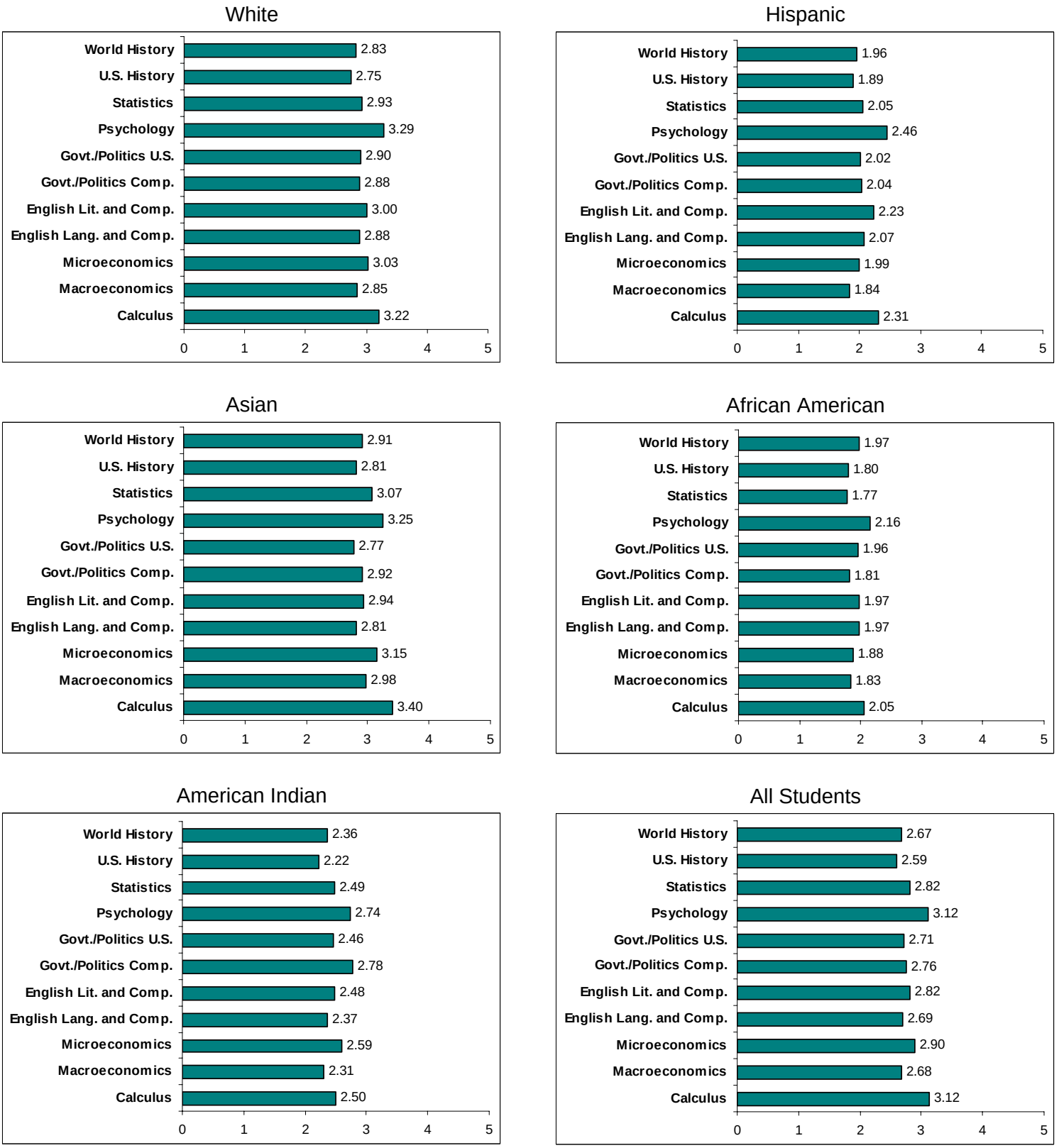
Math



Source:
<www.k12.wa.us/assessment/NAEP/pubdocs/2005Grade8math.pdf>.

Exhibit C.9

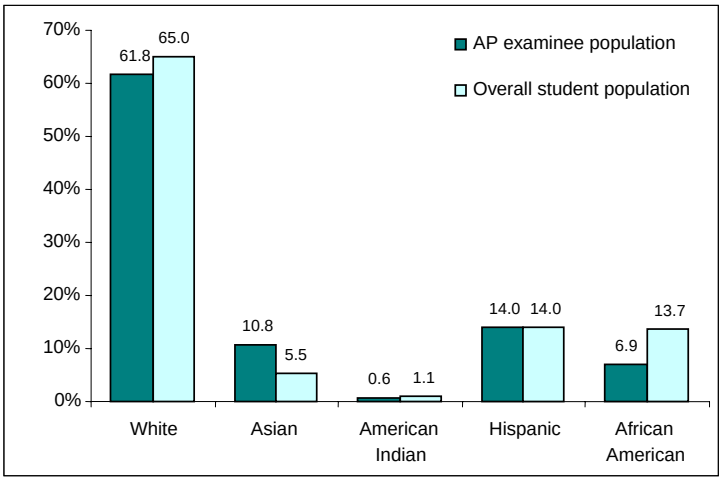
Average AP Exam Grades for U.S. Public School Students in the Class of 2006 by Race/Ethnicity



Source: College Board. (2007). *Advanced Placement report to the nation 2007*. New York: The College Board. Available: <www.collegeboard.com/ prod_downloads/about/news_info/ap/2007/2007_ap-report-nation.pdf>.

Exhibit C.10

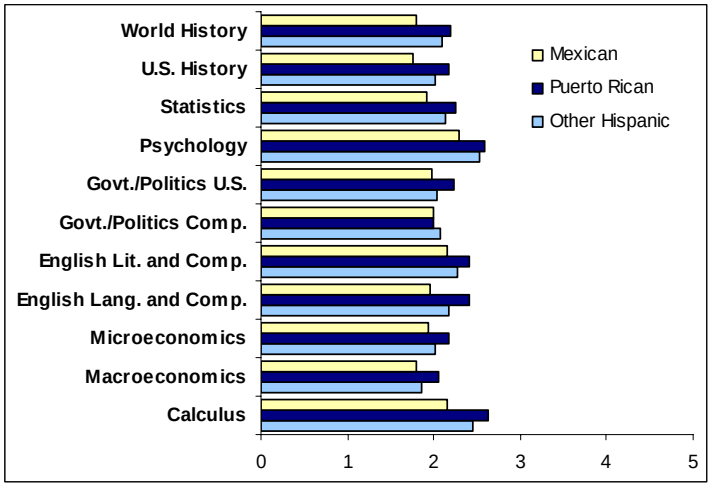
AP Examinees and Graduating Seniors by Race/Ethnicity, U.S. Public Schools, Class of 2006



Source: College Board, 2007.

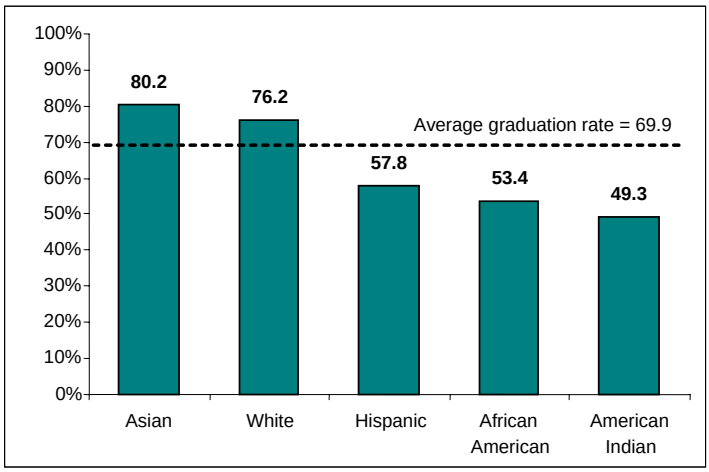
Exhibit C.11

Average AP Exam Grades for U.S. Public School Students in the Class of 2006 by Hispanic Subgroups



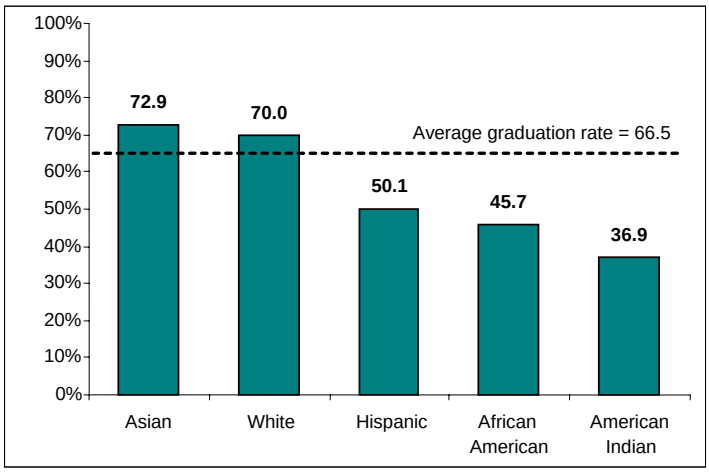
Source: College Board, 2007.

Exhibit C.12
Graduation Rate by Race/Ethnicity, United States, 2003–04



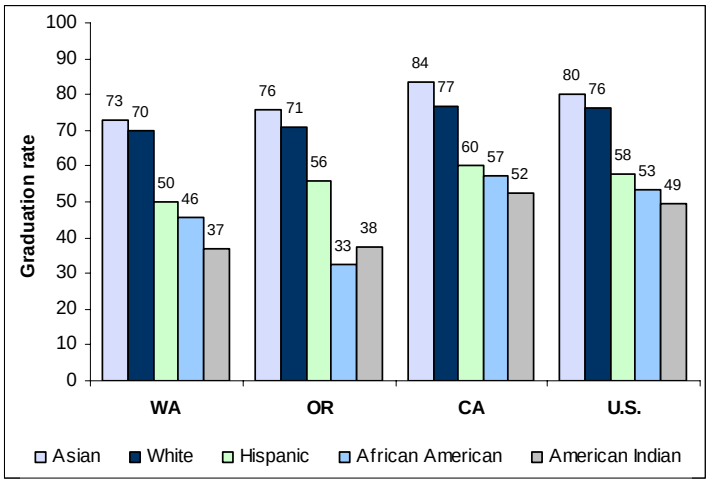
Source: *Education Week*, vol. 26, no. 40 (June 12, 2007), p. 40.

Exhibit C.13
Graduation Rate by Race/Ethnicity, Washington State, 2003–04



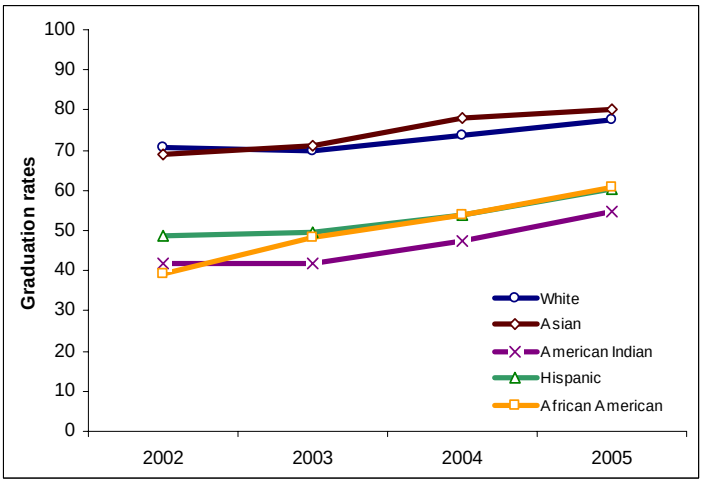
Source: *Education Week*, vol. 26, no. 40 (June 12, 2007), p. 40.

Exhibit C.14
Graduation Rate by Race/Ethnicity, Washington, Oregon, California, and National, 2003–04



Source: *Education Week*, vol. 26, no. 40 (June 12, 2007), p. 40.

Exhibit C.15
On-Time Graduation Rates by Race/Ethnicity, Washington State, Classes of 2002–05



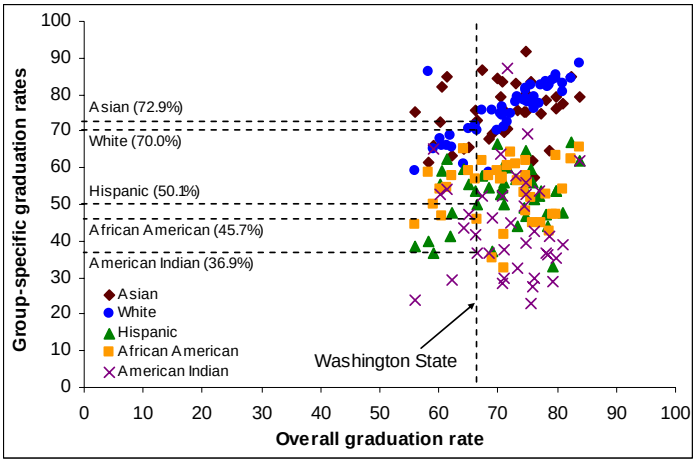
Source: Office of Superintendent of Public Instruction. (2004–2006). *Graduation and dropout statistics for Washington's counties, districts, and schools*. Olympia, WA: OSPI.

Exhibit C.16
Graduation Rates by State and Race/Ethnicity, 2003–04

	All Students	African American	American Indian	Asian	Hispanic	White
AL	59.0	49.9	65.3	66.2	36.5	65.2
AK	65.1	59.3	47.2	65.7	55.7	70.5
AZ	68.4	57.8	36.7	68.1	54.8	58.5
AR	72.2	64.1	44.8	b	b	74.9
CA	70.7	57.3	52.3	83.7	60.0	76.7
CO	74.6	57.9	52.8	80.4	53.8	81.4
CT	79.8	63.5	b	79.3	53.6	85.5
DE	62.0	b	b	b	41.4	69.0
DC	58.2	58.5	b	61.3	39.7	86.3
FL	60.5	46.7	b	82.2	59.0	66.0
GA	56.1	44.4	23.8	75.4	38.6	59.2
HI	64.1	65.0	43.6	65.3	59.8	60.9
ID	77.0	b	53.5	b	52.3	77.7
IL	75.7	51.8	22.8	83.4	59.8	82.5
IN	70.9	41.6	30.0	69.8	50.2	74.6
IA	81.1	54.0	39.0	b	47.6	82.8
KA	74.4	53.0	49.5	75.3	49.0	78.5
KY	70.0	59.2	b	84.5	66.5	70.3
LA	61.4	54.6	54.2	85.0	62.5	66.2
ME	76.2	b	29.9	57.2	b	76.1
MD	74.7	62.0	56.1	91.8	64.9	81.8
MA	73.2	56.6	32.5	75.8	44.0	79.4
MI	69.1	35.2	46.2	69.4	37.0	75.9
MN	78.7	42.7	41.3	64.6	b	82.6
MS	62.1	58.0	29.2	63.4	47.5	65.6
MO	75.0	55.8	69.1	b	57.4	78.0
MT	76.2	b	42.7	b	51.5	79.4
NE	79.8	47.4	35.2	76.2	53.8	84.5
NV	54.0	a	a	a	a	a
NH	76.0	a	a	a	a	a
NJ	82.5	62.5	b	84.7	67.0	84.6
NM	60.1	54.1	52.7	72.3	55.3	68.0
NY	65.0	a	a	a	a	a
NC	66.1	57.0	41.6	75.5	53.8	71.7
ND	79.4	47.4	29.1	b	33.2	83.9
OH	74.7	48.0	39.4	75.3	46.8	78.8
OK	70.4	57.0	63.9	79.3	52.6	75.0
OR	71.1	32.7	37.6	75.7	56.0	71.1
PA	78.2	52.6	36.9	78.3	47.3	83.6
RI	70.6	57.2	28.3	53.7	54.4	74.3
SC	53.8	a	a	a	a	a
SD	78.5	b	36.4	b	44.2	82.3
TN	72.2	a	a	a	a	a
TX	67.3	61.7	52.1	86.9	57.9	75.6
UT	83.8	65.5	62.0	79.4	61.7	88.6
VT	81.0	b	b	77.7	b	80.9
VA	73.1	61.0	57.9	83.0	57.1	78.2
WA	66.5	45.7	36.9	72.9	50.1	70.0
WV	71.7	60.7	87.0	70.8	59.9	72.6
WI	77.3	44.8	b	74.8	53.6	82.6
WY	75.8	45.1	27.7	62.0	56.0	77.7
US	69.9	53.4	49.3	80.2	57.8	76.2

a Value not calculated because necessary data field(s) not reported in the U.S. Department of Education's Common Core of Data or because of very small group size.

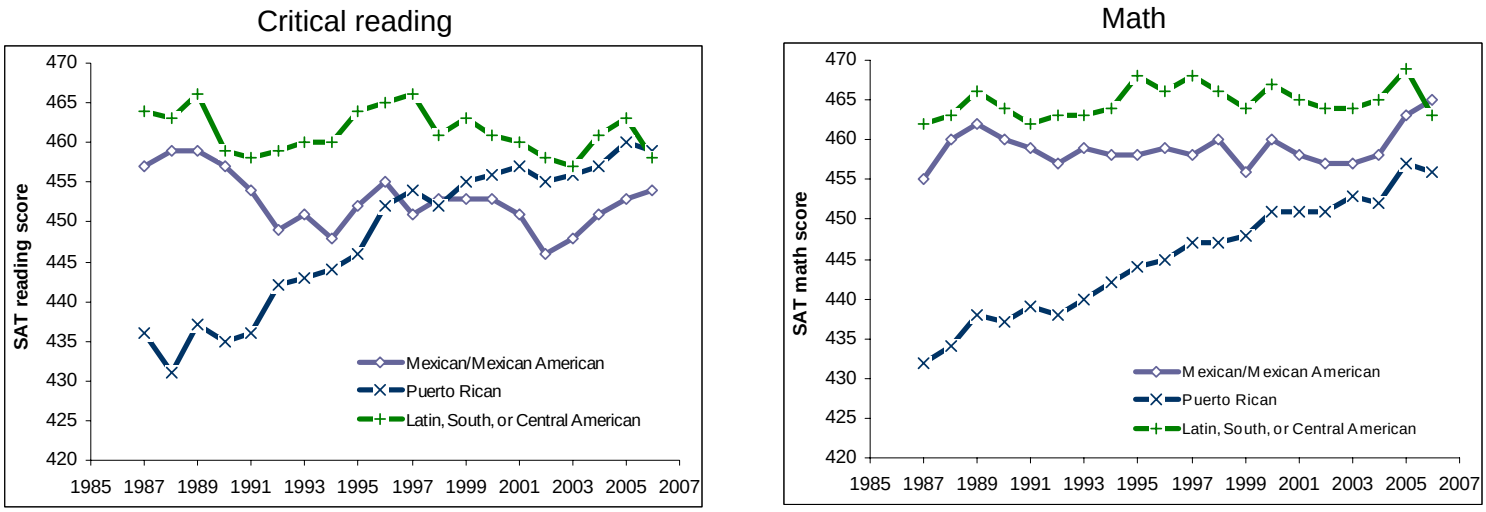
b Value not reported because of insufficient data for estimate.



Source: *Education Week*, vol. 26, no. 40 (June 12, 2007), p. 40.

Exhibit C.17

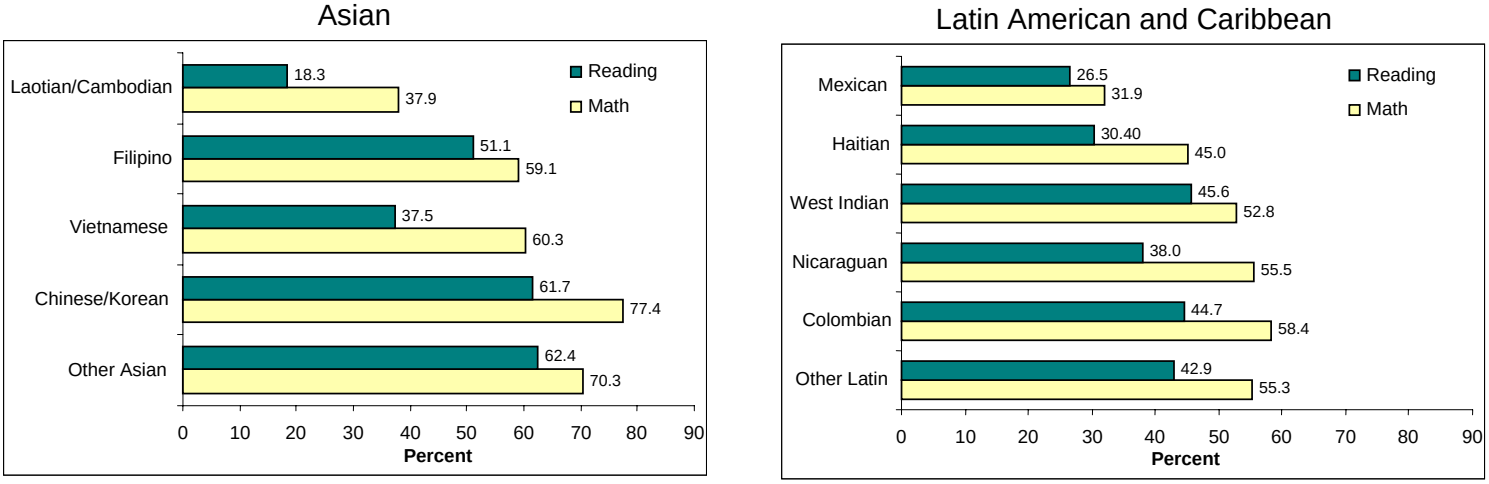
SAT Scores by Hispanic Subgroups: 1987 to 2006



Source: J.L. Kobrin, V. Sathy, & E.J. Shaw. (2007). *A historical view of subgroup performance differences on the SAT Reasoning test*. New York: The College Board (Research Report No. 2006-5).

Exhibit C.18

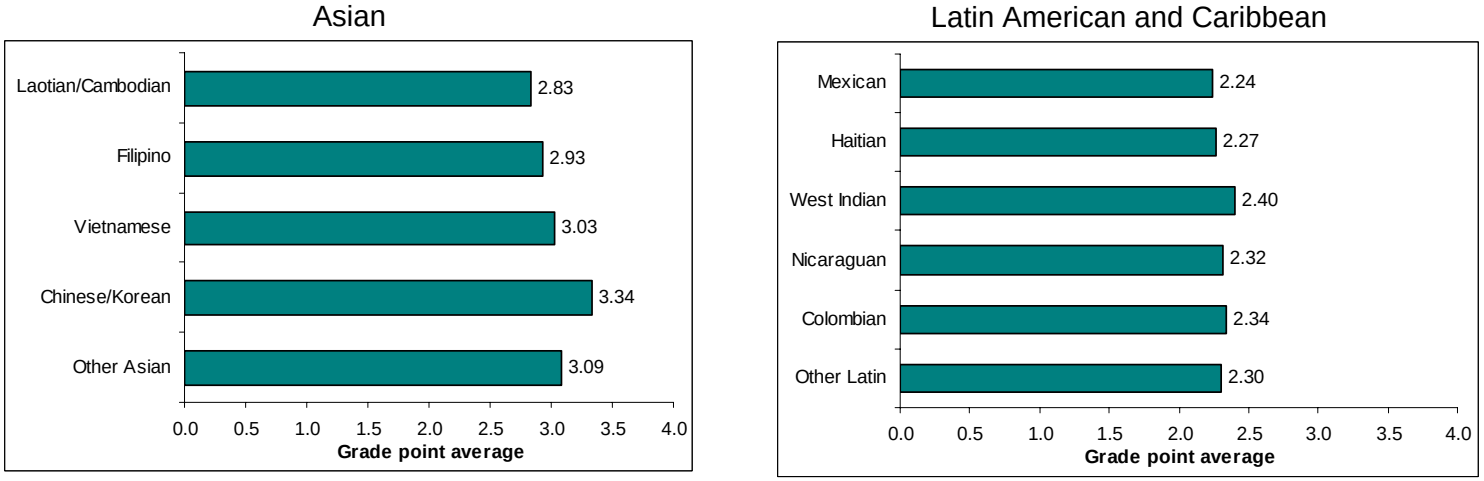
Percentile Scores in Standardized Achievement Tests, Second-Generation Immigrant Subgroups, 1992



Source: A. Portes & R. Rumbaut. (2001). *Legacies: The story of the immigrant second generation*. Berkeley: University of California Press, p. 236.
Notes: Data from the Children of Immigrants Longitudinal Study (CILS); scores from standardized Stanford achievement tests administered in the 8th and 9th grades.

Exhibit C.19

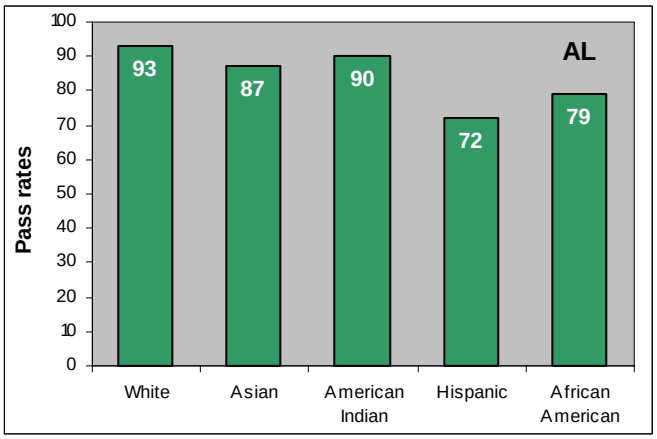
Average Junior High School GPAs by Second-Generation Immigrant Subgroups, 1992

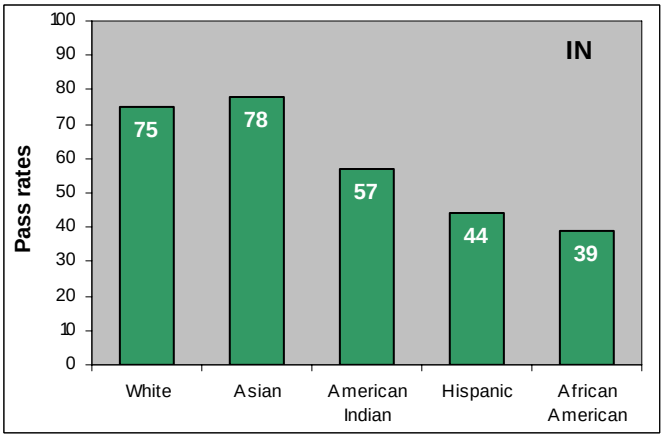


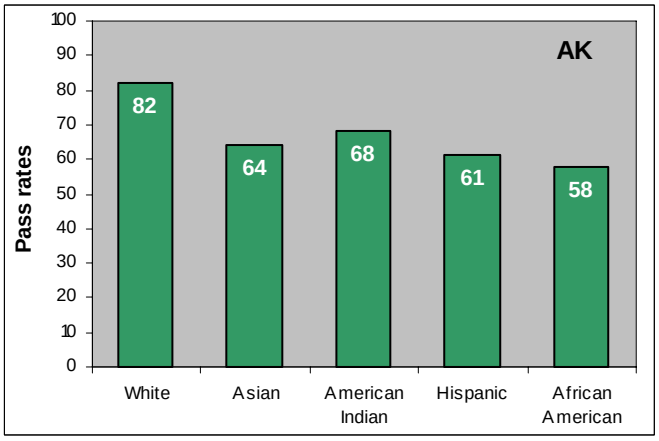
Source: Portes, 2001.

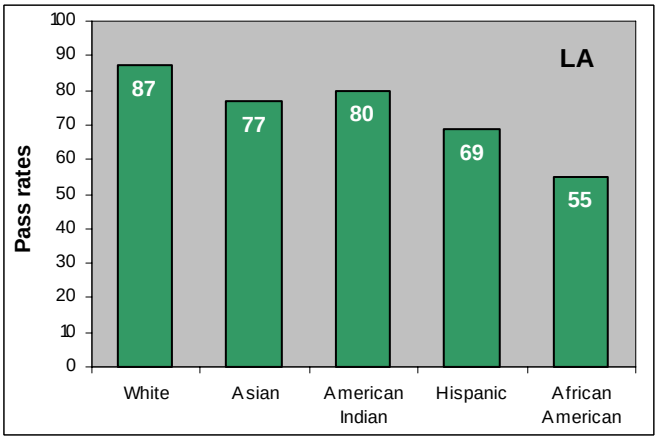
Exhibit C.20

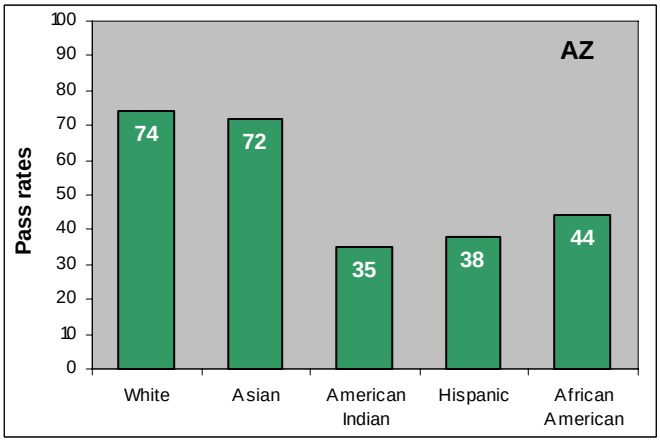
Percentage of Students Who Passed Their State's Reading Exit Exam on the First Try, 2004

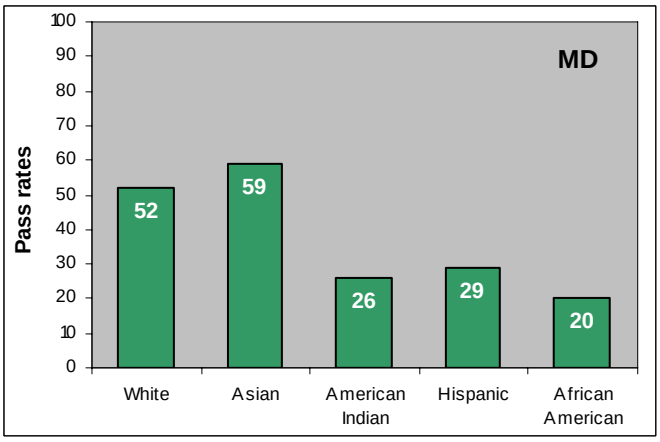


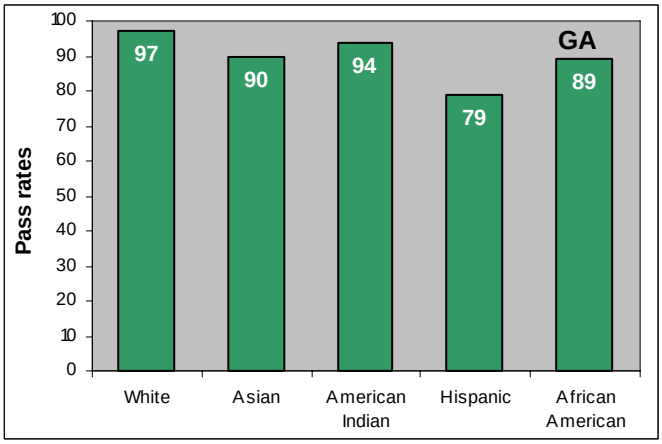


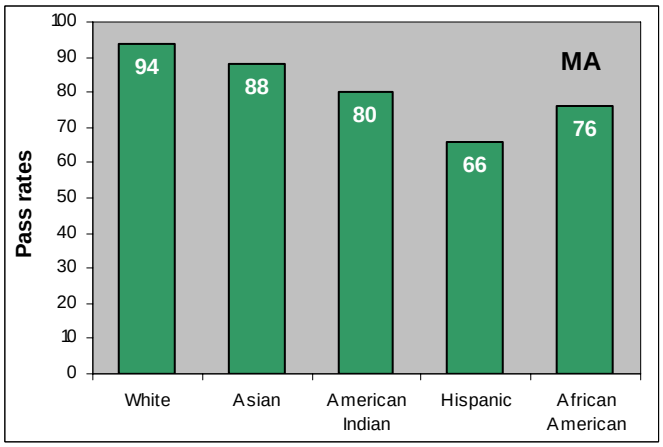




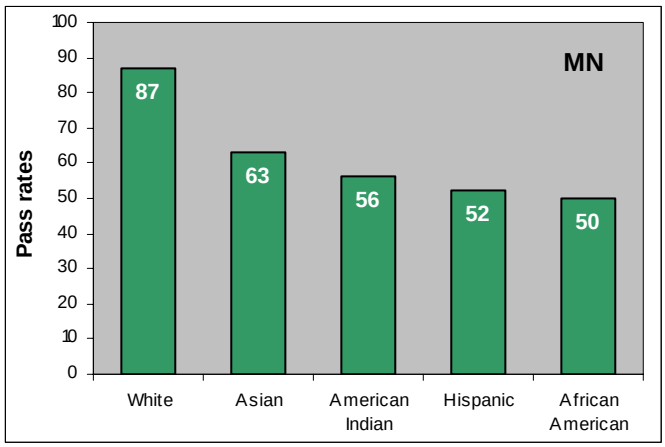


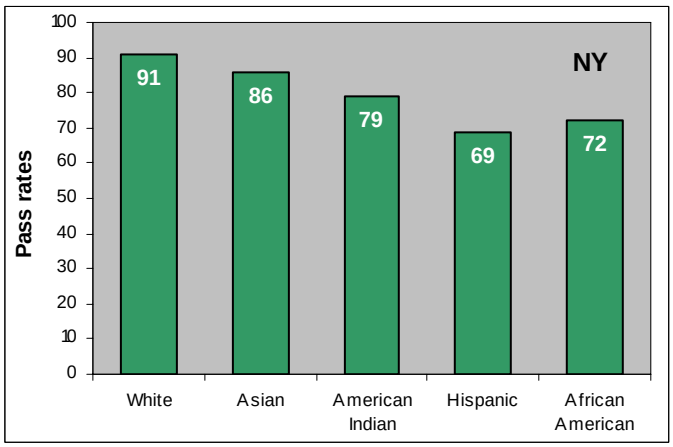


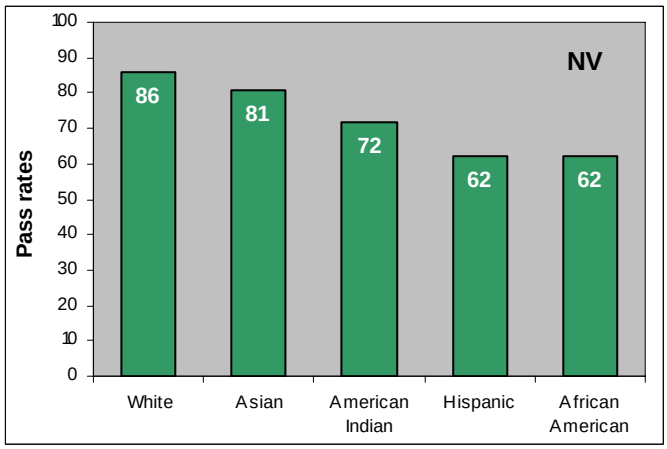


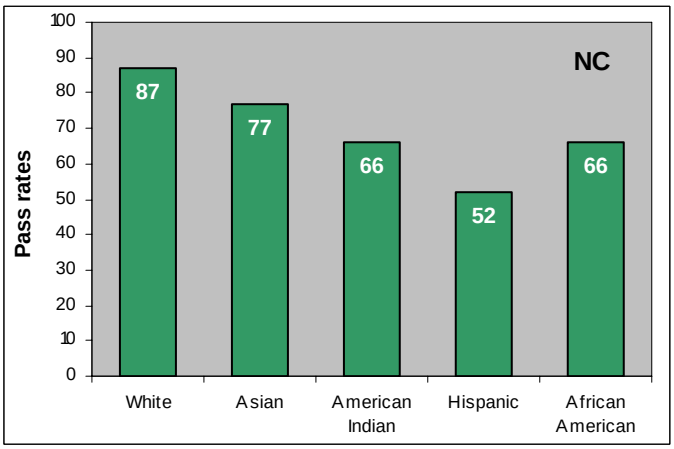


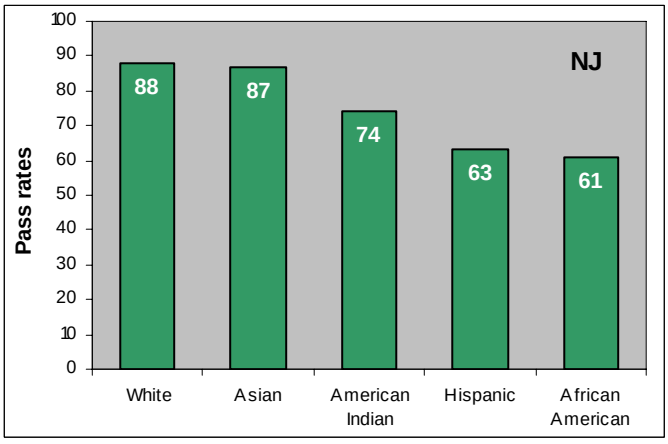
Percentage of Students Who Passed Their State's Reading Exit Exam on the First Try, 2004
(*Continued*)

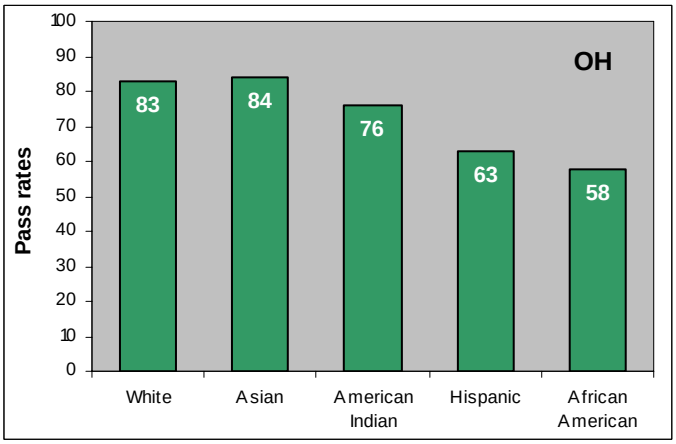


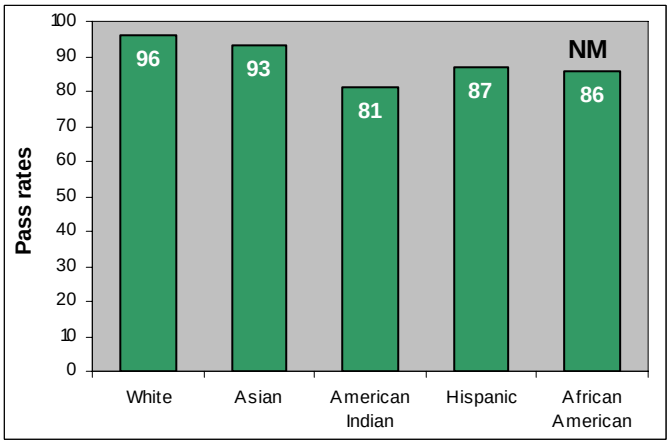


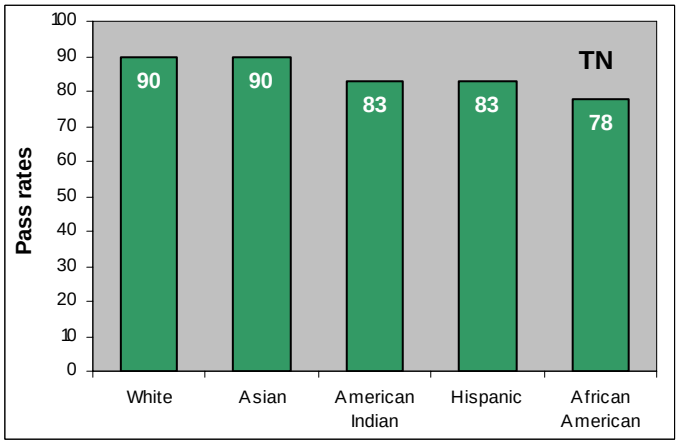




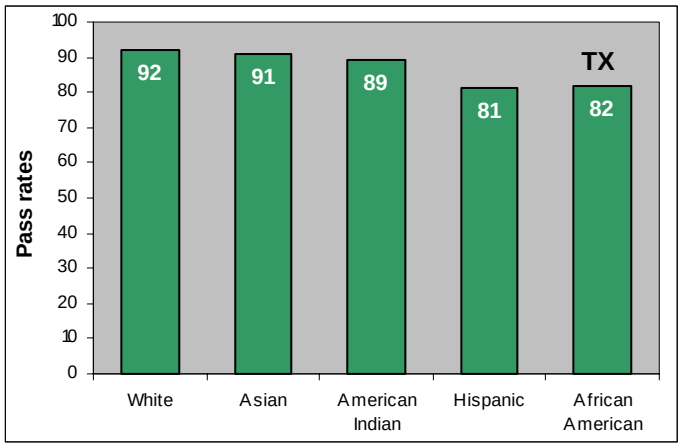


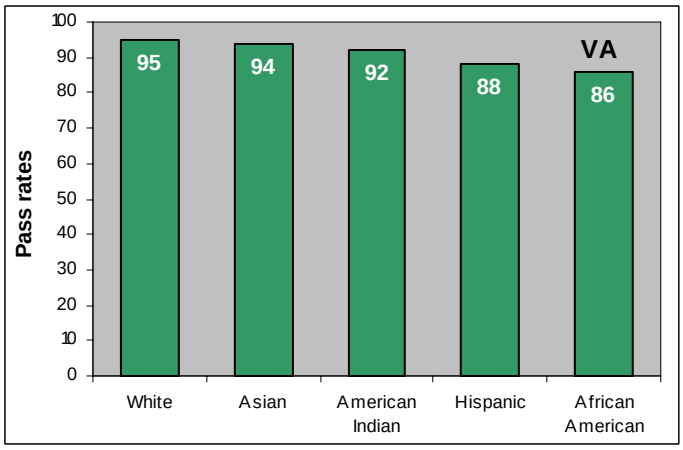


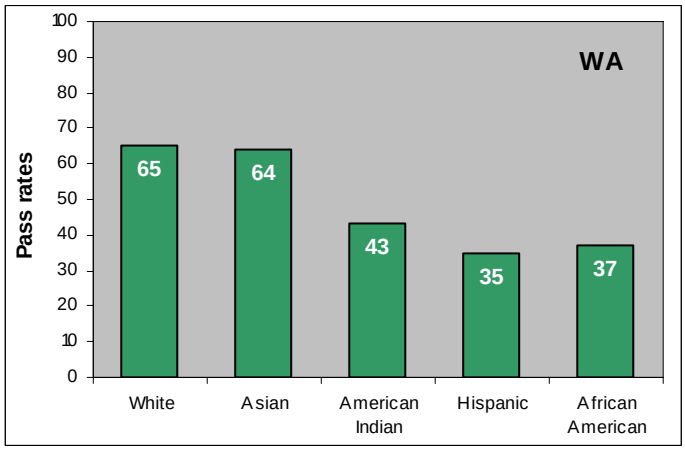




Percentage of Students Who Passed Their State's Reading Exit Exam on the First Try, 2004
(*Continued*)



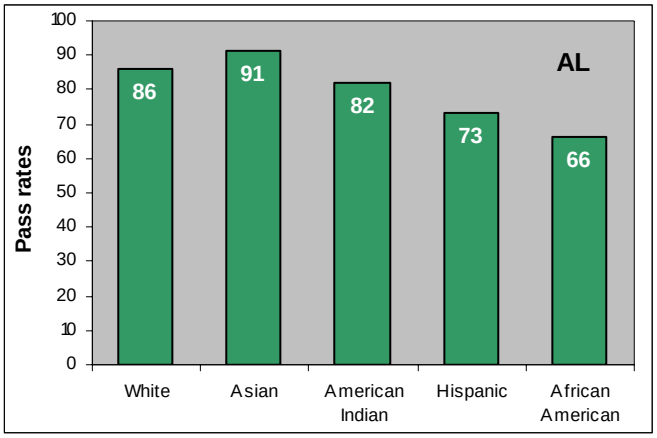


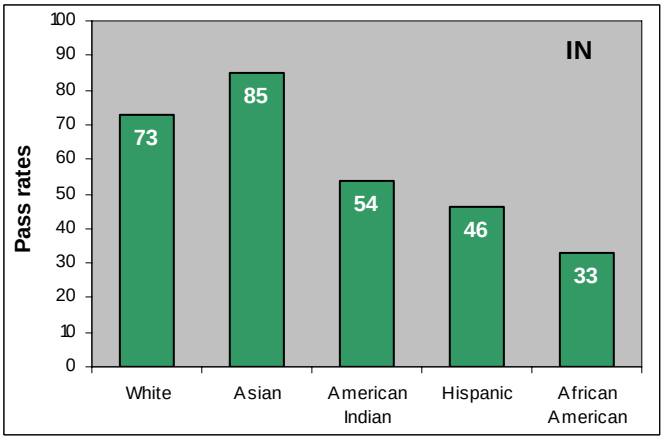


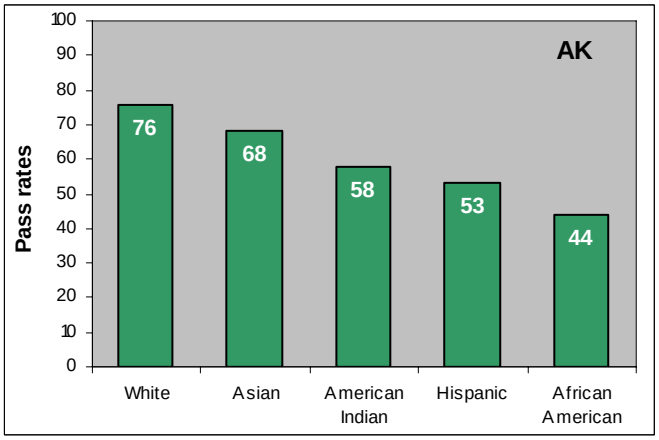
Source: National Center for Educational Statistics. (2005). *The condition of education 2005*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, p. 165.

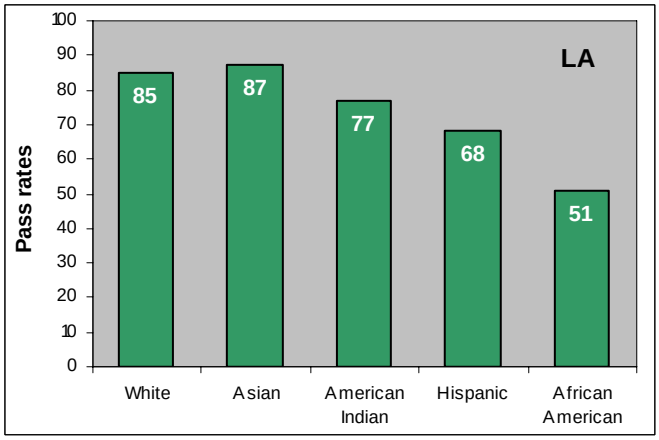
Exhibit C.21

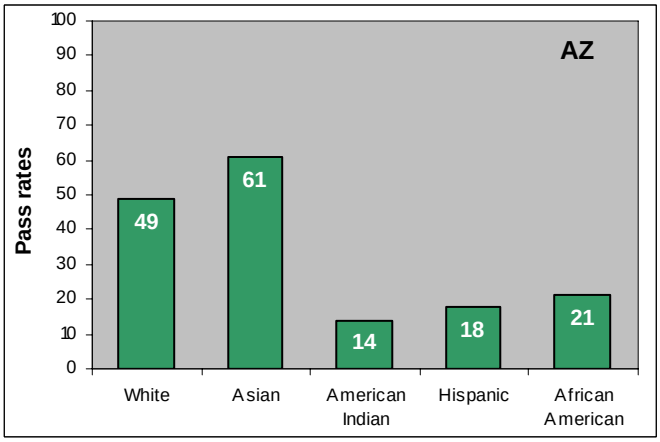
Percentage of Students Who Passed Their State's Math Exit Exam on the First Try, 2004

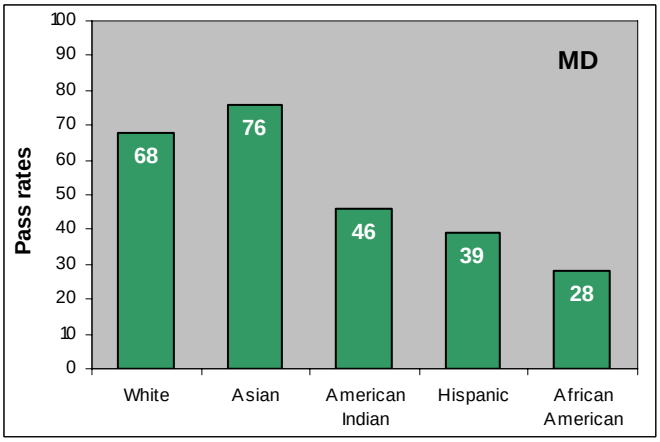


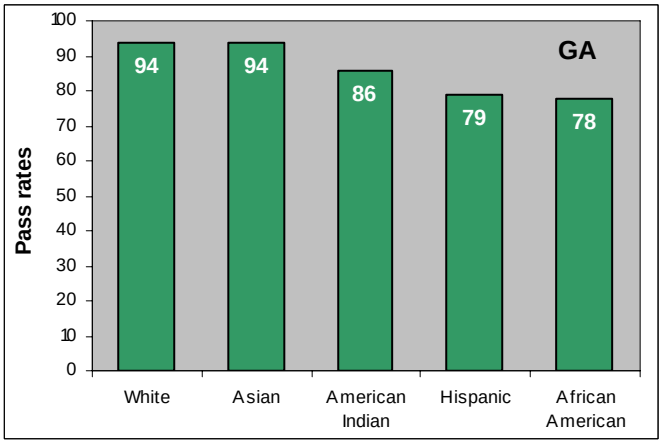


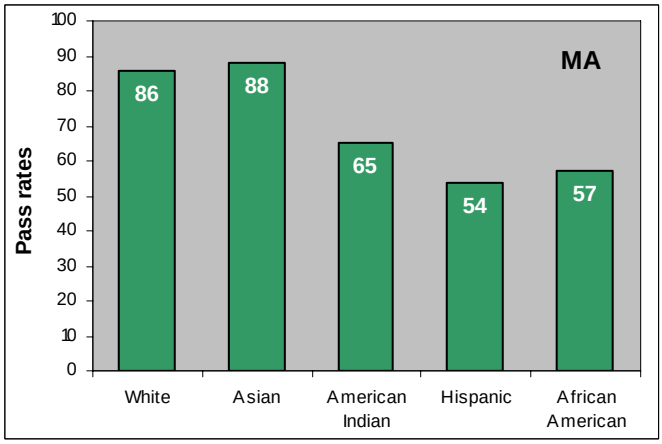




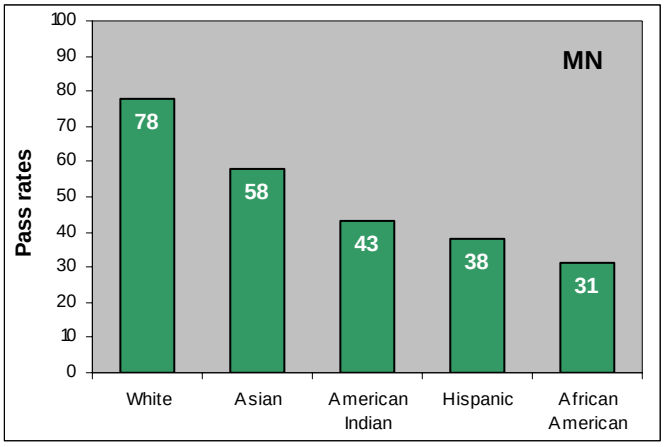


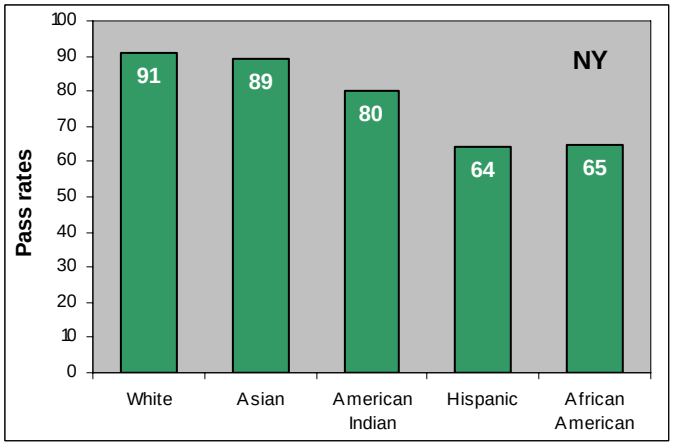


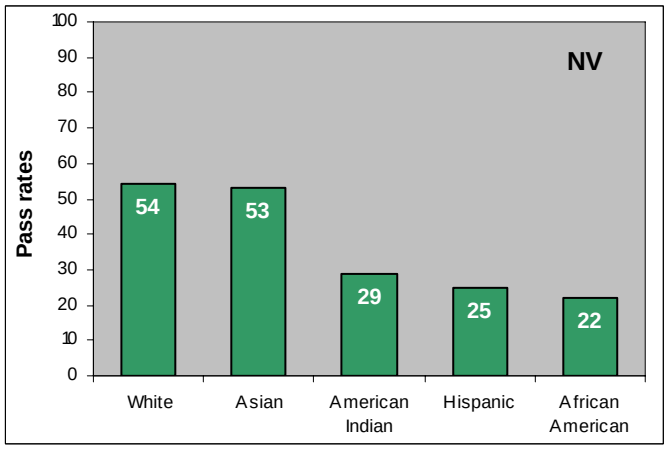


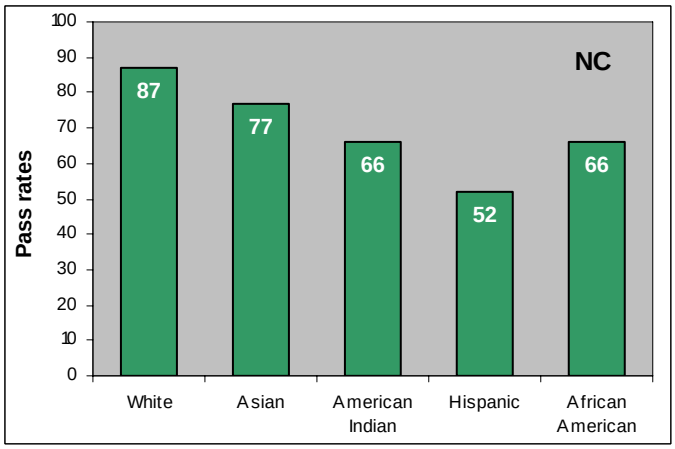


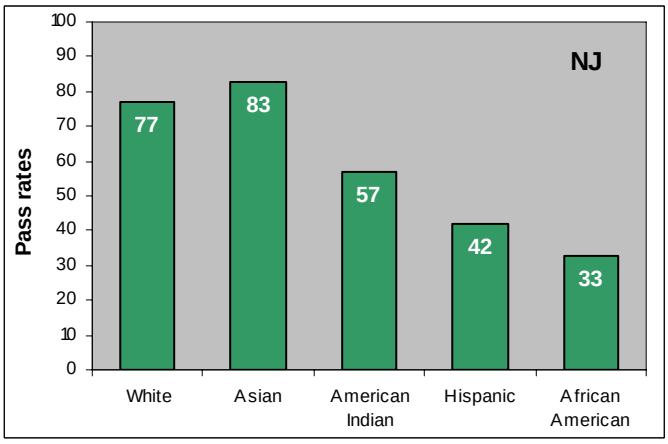
Percentage of Students Who Passed Their State's Math Exit Exam on the First Try, 2004
(*Continued*)

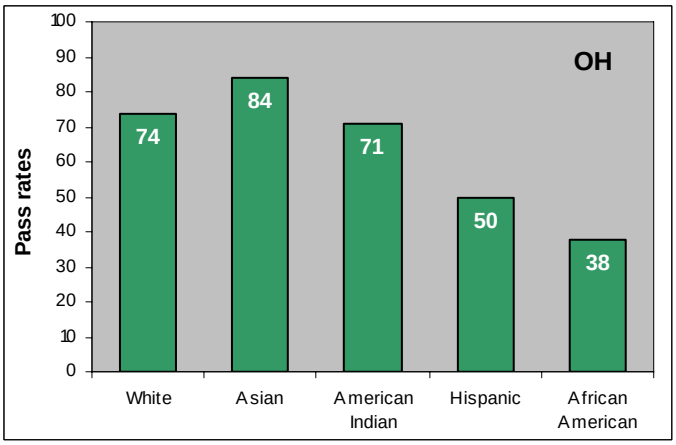


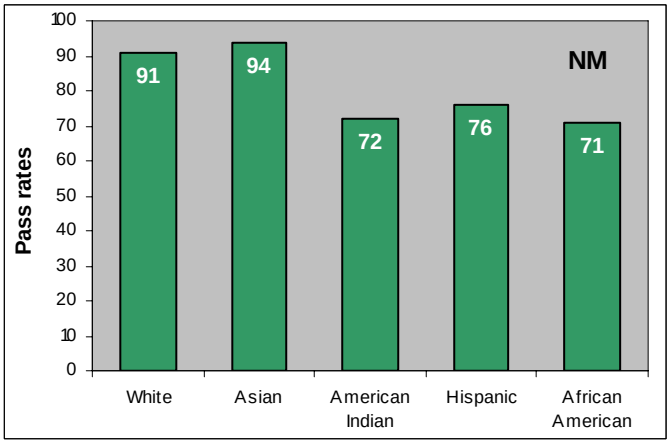


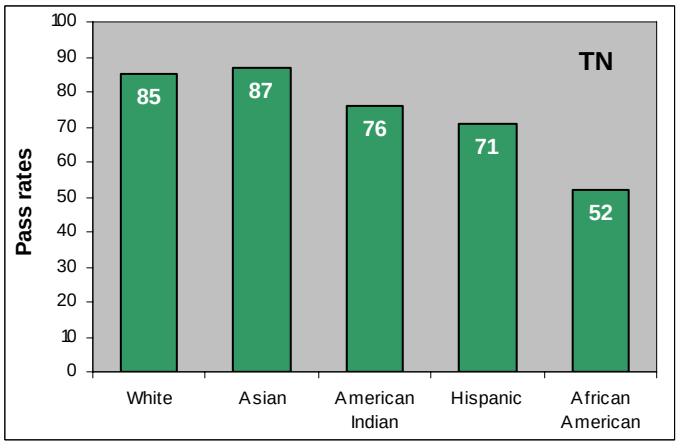




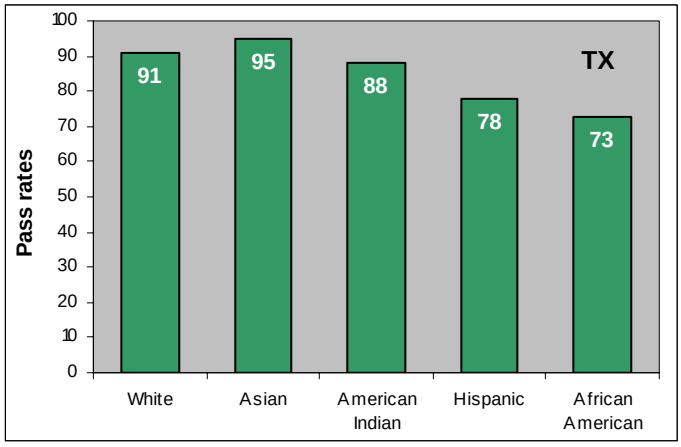


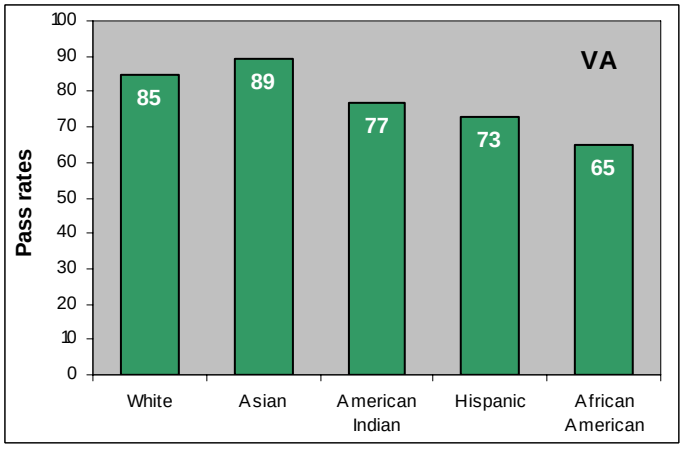


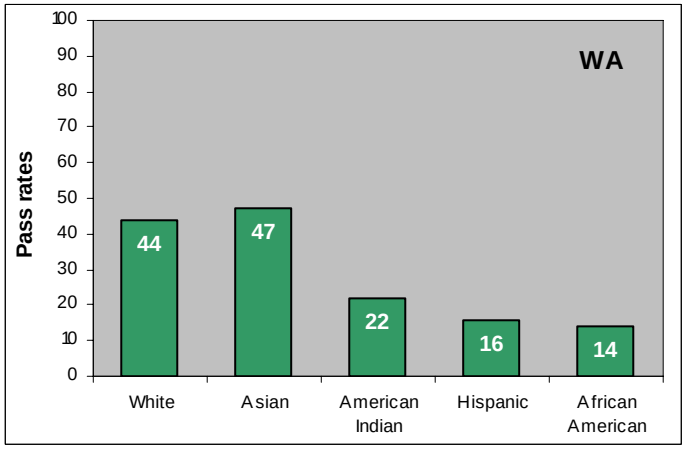




Percentage of Students Who Passed Their State's Math Exit Exam on the First Try, 2004
(*Continued*)







Source: National Center for Educational Statistics, 2005, p. 166.

APPENDIX D: MULTIPLE MEASURES POLICY PAPERS

To assist in the review of a multiple measures assessment system, the Institute contracted with educational researchers to write policy briefs addressing the following questions:

1) What does the concept of "multiple measures" mean with respect to assessments required for high school graduation? How might such a system be implemented?
2) What factors should policymakers consider when designing a multiple measures assessment system?
3) Is Washington's assessment system already a multiple measures system? Explain why or why not. How could Washington's system be strengthened?
4) Could a high-stakes multiple measures system be implemented in a manner that is valid, reliable, cost effective, comparable in content and rigor to the WASL, culturally appropriate, and standardized across school districts? (These are the review criteria set by the legislature).
5) Does research evidence indicate whether multiple measures would be more valid for specific subpopulations of students—racial/ethnic groups, low income students, English language learners, and special education students—compared with a single test or measure such as GPA?
6) Is it reasonable to expect that using a multiple measures system (or strengthening our current system) would improve high school graduation rates?
7) Is there any research we should review on this topic? If so, please list.

This appendix contains three responses from the following educational researchers:

Multiple Measures and High School Graduation

Susan M. Brookhart, Ph.D.
Brief prepared for the Washington State Institute for Public Policy
October, 2007

What does the concept of "multiple measures" mean with respect to assessments that are required for high school graduation? "Multiple measures" has been used to mean several different things, and there is not yet consensus. Two things are at issue. First, which of the following count as "multiple measures": (a) measures of different constructs; (b) different measures of the same constructs; or (c) multiple opportunities to pass the same test? Second, which of the following methods are applied: (a) conjunctive (using "and" logic, must pass all measures); (b) compensatory (where higher performance on one measure can compensate for lower performance on another); (c) complementary (using "or" logic, where achieving standard on one of the multiple measures suffices) and some would add (d) confirmatory (where additional measures are used only to validate primary measures; Chester, 2005).

For purposes of school accountability under NCLB, multiple opportunities to pass the same state test, plus a non-achievement factor like attendance, are sometimes considered "multiple measures." This is not what the professional organizations (see references) mean when they recommend using "multiple measures" to make high-stakes decisions for individuals, as is the case for high school graduation. To be true to these recommendations, multiple measures of achievement constructs should be used. This is especially relevant to high school graduation, where achievement of fundamental skills is of interest.

Does research evidence indicate whether multiple measures would be more valid for specific subpopulations of students—racial/ethnic groups, low income students, English language learners, and special education students—compared to a single test or measure such as GPA? An argument for multiple measures can be made on the basis of providing multiple modes for performance: tests, writing, projects, and so on. Some students may be better able to show what they know in one modality than another. The argument for multiple measures has been made for Native Americans (ERIC, 2001) and for English language learners (Garcia, 2000). Marwick (2004) found that when multiple measures of academic preparedness were considered in placing students into community college mathematics courses, Hispanic students were initially placed into higher-level courses – where they achieved equal or greater academic success – than when only standardized test score or only high school preparation was considered.

Carnoy (2005) studied the effects of a strong accountability system on state graduation rates, and found that while race/ethnicity was strongly related to graduation rate, when 8th grade test score level was controlled, black and Hispanic students had higher graduation rates than white students. The stronger the state's accountability system, the lower the increase in progression rates from 8th to 12th grades. He concluded, from these and other analyses, that if state accountability systems succeeded in raising academic achievement in elementary and middle schools, the result would be stronger high school performance and graduation. The existence of an exit exam in high school by itself made little difference.

The Center on Education Policy (2007, Jan.) analyzed patterns of changes in pass rates in states with high school exit exams between 2004 and 2005, both for students overall and for particular subgroups. States that had gains in overall pass rates often had even greater gains among minority subgroups, a positive step toward closing the gap that they speculate may be because states allocate resources when pass rates are on the line. However, states that had losses in overall pass rates often had greater losses among minority subgroups. Their overall conclusion was that state results for high school exit exams, overall and among subgroups, was variable.

Is Washington's assessment system already a multiple measures system? Explain why or why not. Darling-Hammond, Rustique-Forrester, and Pecheone (2005) studied multiple measures approaches to high school graduation. Their report contains appendices with state profiles of high school graduation policies for states with multiple measures approaches. Washington is listed among those states, qualifying as a "multiple measures" state based on the plans for 2008 that were in place at the time the report was written. I concur with this designation, for the reasons explained in the analysis below.

Washington's assessment system is a multiple measures system in the sense that term is currently used. It would not be a multiple measures system for everyone if graduation were based on the WASL only, and the alternatives were accessed only for students who failed the WASL. However, the plan outlined on the Washington OSPI website (www.k12.wa.us/graduationrequirements/GradRequirements.aspx) lists four measures that may be construed as a multiple measures approach according to the following analysis:

	Conjunctive (AND)	**Compensatory (COMBINED JUDGMENT)**	**Complementary (OR)**
Measures of different constructs [The "constructs" in this analysis are the WA fundamental skills.]	HS & Beyond Plan measures fundamental skill #4 more than the others; Credits and WASL measure fundamental skills #1-3 more than #4. Culminating Project measures some of all 4.	[No. All 4 steps are required for graduation.]	
Different measures of the same constructs	Course credits and grades and WASL or alternatives both measure academic fundamental skills.		Students may choose substitute exams, performance-based assessment, or GPA as an alternative to the WASL after failing once.
Multiple opportunities			Students may retake the WASL up to 4 times. Students start HS&B Plan in 8th or 9th grade and revise. Some courses may be repeated if failed.

Note that what the system currently does NOT do is use a compensatory approach, such as a Weighted Multiple Measures approach (Darling-Hammond *et al*., 2006), where an index is created such that lower performance on one measure might be offset by higher performance on another. As the analysis in the

chart shows, the current Washington system as described on the website is basically a conjunctive multiple measures system, with 4 jointly required measures that overlap in that all assess aspects of the four fundamental skills, yet for each the emphasis is different. For those students who do not pass the WASL, one of the four steps (the Certificate of Academic Achievement or Individual Achievement) becomes a complementary multiple measures system, allowing the substitution of alternatives for the WASL score. There are also multiple opportunities to take the WASL, and presumably multiple opportunities to earn the course credits, as well (if a student failed a course, they could repeat it).

A few caveats are in order. For Step 1, the High School and Beyond Plan, there is no "pass" or "criterion" language built in. It is left to local districts to decide whether a student's plan meets standard for graduation. The same caution applies for Step 4, the Culminating Project. The state website offers some example rubrics, but implementation and decisions are left to the local school district. This is different from Steps 2 (19 course credits) and 4 (Certificate of Academic Achievement or Individual Achievement, based on WASL or alternatives), where there are achievement level requirements built in: passing the courses and meeting cut scores, respectively. To the extent that passing levels are different in different districts, the system is not standard, but at least there are defined requirements for passing. The multiple measures analysis in the above chart assumes that local districts are stepping up to their responsibilities regarding the Plan and the Project, and at present the state does not check on that.

How could Washington's system be strengthened? Washington might consider two avenues, both of which would strengthen the system of high school graduation requirements.

(1) For the High School and Beyond Plan, and for the Culminating Project, each district sets policy and evaluates the final product. Therefore, the reliability and validity of the system, especially regarding the function of two of the four measures, from district to district, is unknown. As it stands now, the system could be considered to have two "measures" (credits and an exam) and two other factors for which no evidence that completion indicates achievement of state fundamental skills has been established. A study of the validity and reliability of these steps for making graduation decisions would go a long way toward strengthening evidence that all four measures function as designed in the system. A more centralized approach (e.g., state rubrics and rating) for these two factors would accomplish this goal but might weaken the local district contribution to the system, which in itself is important.

(2) The Darling-Hammond, McCloskey, and Pecheone (2006) report has already suggested, and I concur, that the alternatives to the WASL could be expanded to be offered to all students, not just those who fail the WASL. The analysis above shows alternatives to the WASL are complementary multiple measures, but this is only the case for some students (those who fail). The 2007 legislature has already come partway down this road by expanding the alternative test option to include reading and writing as well as math (it was math-only in 2006). This change makes the policy more consistent with a multiple measures approach than the math-only policy, which is more like a response to disappointing scores in one subject than an endorsement of multiple measures. Expansion of the option to other students as well as to other tests would strengthen the multiple measures aspect of the system.

Is it reasonable to expect that using a multiple measures system (or strengthening our current system) would improve high school graduation rates?

There is research that demonstrates that graduation rates decrease in states that require high school graduation exams alone. This literature was reviewed for the state of Washington by Darling-Hammond McCloskey, & Pecheone (2006). Their 47-page report, prepared at the School Redesign Network at Stanford University to address much the same question as this brief, is still very current, and is strongly recommended as background reading. Darling-Hammond and her colleagues (2006) reported that

graduation rates, calculated as the number of graduates divided by the size of the 9th grade cohort 3.5 years earlier, stayed the same (within 2 percentage points, 2 states) or declined slightly (3 states) from 1998 to 2001 for five states (Indiana, North Carolina, New York, Florida, South Carolina) that required exit exams only. Results were somewhat better for multiple measures. Darling-Hammond and colleagues (2006) found that graduation rates stayed the same (within 2 percentage points, 3 states) or rose (1 state) from 1998 to 2001 for four states (New Jersey, Wisconsin, Pennsylvania, Connecticut) that used a multiple measures approach to graduation during that time. In addition, the graduation rates for these four states were higher overall (73% to 86% in 2001) than for the five exam-only states (51% to 67% in 2001). Note, however, that state graduation rates vary widely and are affected by other economic, political, and cultural factors in addition to state graduation policies.

Theoretical support for using multiple measures includes issues of: (a) validity (various kinds of measures provide a richer representation of learning domains); (b) reliability (more than one measure leads to more confidence in a decision); (c) fairness (different learners may be more able to show what they know on different types of measures); and (d) consequences for education (the curriculum is less likely to become narrowed, as in the phrase "teaching to the test," if outcome measures are broad). All the references, both for individual and school accountability, speak to these reasons.

Is there any research we should review on this topic? A list of references accompanies this brief. The references are organized into categories according to whether the main topic was use of multiple measures for school (aggregated) accountability or individual accountability. High school graduation is an individual-level decision, so the latter are the most relevant. The articles about school-level accountability are included because a lot of the multiple measures work has been done in that context.

The Center on Education Policy has published studies of high school exit exams since 2002, most recently in September (CEP, 2007, Sept.). This report and its appendices include descriptions of each state's approach to high school graduation and an analysis of trends, including a trend toward more end-of-course exams. I commend that report to Washington as the best place to find current information about other states. CEP reports that 22 states currently mandate high school exit exams, with four more – including Washington – planning to, which would bring the number to 26. Public opinion still flares. Maryland is, like Washington, one of those four states planning to require high school exams; even as this report was in preparation, the Baltimore *Sun* ran a story about parents' and advocates' disagreement on whether special education students should have to pass the test to graduate (Oct. 22, 2007).

The abovementioned CEP study (2007, January) on subgroup pass rates shows that multiple measures systems can be associated with higher subgroup pass rates but are not guaranteed to be, and that even when pass rates rise they are often below desirable levels. In the long run, the theoretical arguments about validity, reliability, fairness, and curriculum may be the most important reasons to use multiple measures systems for high school graduation. I am firmly in favor of multiple measures systems, for these reasons – even though initial implementation of the system may be difficult and may not show an immediate rise in graduation rates.

Multiple measures do not "manufacture" achievement. Students should be given several opportunities to show what they know, but if they truly don't know it additional measures will not change that fact. Even a perfect high school graduation system won't register achievement that isn't there. To change subgroup pass rates in large and meaningful ways, instruction as well as assessment must change. The Council of Great City Schools (2000) recommends using multiple measures all through school to assure students appropriate grade placement and instruction. Carnoy's (2005) analysis of graduation rates also suggests that the way to raise graduation rates is to strengthen instruction and assessment all through school.

Multiple Measures and High School Graduation

Susan M. Brookhart, Ph.D.
Bibliography prepared for the Washington State Institute for Public Policy
October, 2007

Legal and Ethical Support for the Use of Multiple Measures

American Educational Research Association, (2000, July**).** *AERA Position Statement on High-Stakes Testing in Pre-K – 12 Education*. Available: http://www.aera.net/policyandprograms/?id=378

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

National Council on Measurement in Education (1995). *Code of professional responsibilities in educational measurement*. Available: http://www.ncme.org/about/docs.cfm

National Education Association. (2006, July). *NEA's positive agenda for the ESEA reauthorization*. Available: http://www.nea.org/lac/esea/images/posagenda.pdf

NCLB 2001, Title I, Subpart 1, Sec. 1111(b)(3)(c)(vi): "Such assessments shall involve multiple up-to-date measures of student academic achievement, including measures that assess higher-order thinking skills and understanding;" THIS LANGUAGE WAS ALSO IN TITLE I OF IASA (1994).

Articles Relevant to the Use of Multiple Measures for Individual Student Accountability, Including High School Graduation

Published pre-NCLB:

Eisner, C. (Ed.). (2000). *Ending social promotion: Early lessons learned*. Washington, DC: Council of the Great City Schools and U.S. Department of Education Office of Elementary and Secondary Education. ERIC Document No. ED 448234.

ERIC Development Team. (2001). *American Indian/Alaska Native education and standards-based reform.* ERIC Digest. ERIC Document No. ED 459039.

Garcia, P. A. (2000). *Increasing the rigor of evaluation studies of programs for English Learner students.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans. ERIC Document No. ED 441838.

Goldberg, G. L., & Roswell, B. S. (2001). Are multiple measures meaningful? Lessons from a statewide performance assessment. *Applied Measurement in Education, 14*, 125-150.

Johnson, C. R., & Taylor, R. (2001). Excellence for all in Minneapolis. *Educational Leadership, 58*(6), 55-59.

Published post-NCLB:

Carnoy, M. (2005). Have state accountability and high-stakes tests influenced student progression rates in high school? *Educational Measurement: Issues and Practice*, *24*(4), 19-31.

Center on Education Policy. Series of reports on High School Exit Exams, conducted since NCLB. Available: http://www.cep-dc.org/index.cfm?fuseaction=Page.viewPage&pageId=493&parentID=481
(2007, September). *State high school exit exams: Working to raise test scores.*
(2007, March). *"It's Different Now": How Exit Exams Are Affecting Teaching and Learning in Jackson and Austin.*

(2007, January). *State High School Exit Exams: States Continue Trend Toward Higher-Level Exit Exams, More Subjects Tested – Policy Brief 1*.
(2007, January). *State High School Exit Exams: Growth in High School Exit Exams Levels Off But Minority Students Affected Disproportionately – Policy Brief 2*.
(2007, January). *State High School Exit Exams: Gaps Persist in High School Exit Exams Pass Rates – Policy Brief 3*.
(2006, December). *High School Exit Exams: Basic Features.*
(2006, September). *The Hidden Costs of High School Exit Exams.*
(2006, August). *State High School Exit Exams: A Challenging Year*.
(2006, March). *High School Exit Exams: Standards Differ from the No Child Left Behind Act.*
(2006, February). *High School Exit Exams: Special Problems Affecting English Language Learners.*
(2006, January). *High School Exit Exams: Effects on Traditionally Underserved Students.*
(2005, August). *State High School Exit Exams: States Try Harder, But Gaps Persist*.
(2005, June). *How Have High School Exit Exams Changed Our Schools: Some Perspectives from Virginia and Maryland*.
(2004, August). *State High School Exit Exams: A Maturing Reform*.
(2004, May). *Pay Now or Pay Later: The Hidden Costs of High School Exit Exams*.
(2003, August). *State High School Exit Exams: Put to the Test*.
(2003, March). *Effects of High School Exit Exams on Dropout Rates*.
(2003, February). *Measuring the Cost of State High School Exit Exams*.
(2002, August). *State High School Exit Exams*.

Darling-Hammond, L., McCloskey, L., & Pecheone, R. (2006, September). *Analysis and recommendations for alternatives to the Washington Assessment of Student Learning*. Stanford, CA: Stanford University School Redesign Network. Available: http://www.srnleads.org/data/pdfs/wasl.pdf

Darling-Hammond, L., Rustique-Forrester, E., & Pecheone, R. L. (2005, April). *Multiple measures approaches to high school graduation*. Stanford, CA: Stanford University School Redesign Network. Available: http://www.srnleads.org/data/pdfs/multiple_measures.pdf

Marwick, J. D. (2004). Charting a path to success: The association between institutional placement policies and the academic success of Latino students. *Community College Journal of Research and Practice, 28*, 263-280.

Articles about the Use of Multiple Measures for School or State Accountability (Included here for the sake of completeness – the previous section contains the most relevant literature for high school graduation)

Published pre-NCLB:

Kolls, M. R. (1998). *Standards-based multiple measures for IASA, Title I program improvement accountability: A vital link with district core values. Rowland Unified School District.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego. ERIC Document No. ED 420681.

Novak, J. R., Winters, L., & Flores, E. (2000). *Using multiple measures for accountability purposes: One district's experience*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans. ERIC Document No. ED 443846.

Winter, Phoebe C. *Combining Information from Multiple Measures of Student Achievement for School-Level Decision-Making: An Overview of Issues and Approaches.* November 2001. Washington, DC: Council of Chief State School Officers. Available: http://www.ccsso.org/Publications/Download.cfm?Filename=CASWinter.pdf

Published post-NCLB:

Baker, E. L., (2003) Multiple measures: Toward tiered systems. *Educational Measurement: Issues and Practice, 22*(3), p. 13-17.

Chester, M. D. (2003). Multiple measures and high-stakes decisions: A framework for combining measures. *Educational Measurement: Issues and Practice, 22*(3), p. 32-41.

Chester, M. D. (2005). Making valid and consistent inferences about school effectiveness from multiple measures. *Educational Measurement: Issues and Practice, 24*(4), 40-52.

Henderson-Montero, D., Julian, M. W., & Yen, W. M. (2003). Multiple measures: Alternative design and analysis models. *Educational Measurement: Issues and Practice, 22*(3), p. 7-12.

Kim, J. S., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher, 34*(8), 3-13.

Koretz, D. (2003). Using multiple measures to address perverse incentives and score inflation. *Educational Measurement: Issues and Practice, 22*(3), p. 18-26.

Lee, J. (2007). Do national and state assessments converge for educational accountability? Meta-analytic synthesis of multiple measures in Maine and Kentucky. *Applied Measurement in Education, 20,* 171-203.

National Assessment Governing Board Ad Hoc Committee on Confirming Test Results. (2002, March). *Using the National Assessment of Educational Progress to confirm state test results.* Available: http://www.nagb.org/pubs/color_document.pdf

Porter, A. C., Chester, M. D., & Schlesinger, M. D. (2004). Framework for an effective assessment and accountability program: The Philadelphia example. *Teachers College Record, 106,* 1358-1400.

Schafer, W. D. (2003). A state perspective on multiple measures and school accountability. *Educational Measurement: Issues and Practice, 22*(3), p. 27-31.

Wilson, M. R., & Bertenthal, M. W. (Eds.). (2005). *Systems for state science assessment*. Washington, DC: National Research Council, National Academies Press.

Multiple Measures and State Student Assessment Systems

1. What does the concept of "multiple measures" mean with respect to assessments that are required for high school graduation? How might such a system be implemented? Be specific as to what assessments would be used and how they would be integrated into an overall determination of eligibility for high school graduation.

The state of Washington uses the Washington Assessment of Student Learning (WASL) for NCLB accountability purposes. Students are tested in grades 3 through 8 and grade 10. Meeting standards on the 10th grade WASL is also a high school graduation requirement.

In 2006 and 2007 the state legislature approved various alternate assessment options for students who did not meet standards on the WASL. The legislature has also considered multiple measures. **Multiple measures** refers to the combining of results from several assessments, as opposed to a single assessment, to measure student achievement. The results of each student's multiple assessments are "averaged out" to determine student performance. Typical multiple measure assessments might include state tests, GPA, nationally available standardized tests, unit tests, end-of-course tests, textbook publisher-provided tests, and district assessments.

Two models can be used to combine the scores of the multiple measures: compensatory and conjunctive. A **compensatory** model allows a high score on one assessment to average out a low score or even failing score on another assessment thereby allowing a student to pass. For example, a student receiving a failing score of "below basic" on the WASL could still pass if he received high scores on the other measures.

A **conjunctive** model requires a minimum cut score on any one assessment or on all assessments to pass. In the conjunctive model, a high score on one assessment cannot offset a score that is below the predetermined passing cut point on another assessment. In the conjunctive model, an assessment's minimum cut point is often set lower than the pass rate for the individual assessment itself. For example, students must currently achieve a "proficient" score on the WASL to pass. In a possible multiple measures conjunctive model, students could pass if they received a "basic" score on the WASL *and* they received high scores on the other measures. They would not pass if they received a "below basic" score on the WASL no matter how high they scored on the other assessments. Note: a conjunctive model that requires students to pass *all* assessments at the original cut points can only lower the overall pass rate since student must now pass all the assessments at the same time. Conjunctive models are often designed to allow a slightly lower cut point for the state test.

2. What factors (such as clarity of learning standards, curriculum and instruction alignment, etc.) should policymakers consider when designing a "multiple measures" assessment system?

For any multiple measures assessment to be valid, it must assess grade-level content. Since the publishing of our book, *Multiple Measures*, DataWORKS Educational Research has collected and analyzed over 2 million student assignments to measure alignment to state standards. The process is called Curriculum Calibration (*Multiple Measures* pages 91-92). We discovered that many high school classes are not taught at grade level. In fact, we found that students at many schools are taught two to four grades below grade level. In a study DataWORKS conducted for the South Carolina Department of Education at 800 schools, we found that less that half of high school language arts assignments were on grade level. Only 68% of Algebra I assignments were on grade level. DataWORKS has found similar results in other states including California, Florida, Alabama, and Washington DC. The South Carolina K-8 study is posted at http://ed.sc.gov/agency/offices/cso/enhance/curriculumcallibration-overview.htm. The high school results are available from DataWORKS.

The misalignment of assignments to grade-level standards raises concern in including GPA, course grades, or school and district assessments as measures in multiple measures calculations. These measures should only be used if they are truly standards-aligned.

3. Is Washington's assessment system already a multiple measures system? Explain why or why not. How could Washington's system be strengthened?

Washington's high school graduation requirements are not a true multiple measures system in which scores are combined and high scores assessments can offset or partially offset low scores on another assessment. It is a multiple requirements system: ***1) High School and Beyond Plan, 2) Credit Requirements, 3) Complete a Culminating Project, 4) Earn a Certificate of Academic Achievement or Certificate of Individual Achievement.***

Students earn the Certificate of Academic Achievement by meeting state reading, writing, and math standards on the WASL or one of the state-approved alternatives to the WASL.

One concern of the state is improving the pass rate of students. The substitute exams such as SAT/ACT scores and AP test results are probably more rigorous than the WASL itself, and students who pass these alternate assessments do show their grasp of content. However, the substitute assessments will probably have a small effect on the overall pass rate since students who pass the alternate assessments are probably the same students who could already pass the WASL. A simple study could be designed to analyze some actual student test results and compare the pass rates of students who have taken several tests to determine how these alternate assessments increase overall pass rate.

The state is using GPA and course grades as alternate assessments. GPA is already a compensatory multiple measure that averages grades. Likewise, course grading is a compensatory multiple measure where students can fail some assignments and still pass the class. However, GPA and course grades should only be used in an assessment system when they reflect student mastery of grade-level content and when consistent grading methodology is used by all teachers throughout the state.

4. Could a high-stakes multiple measures system be implemented in a manner that is valid, reliable, cost effective, comparable in content and rigor to the WASL, culturally appropriate, and standardized across school districts? (These are the review criteria set by the legislature).

The **validity** of an assessment system is created by insuring that all measurements are aligned to what the students are supposed to be learning. This is achieved by designing all measures to assess grade-level state content standards only. For example, if the state uses grades or GPA as one of the measures, then it needs to ensure that grades are standards-based. For example, if teachers review below grade-level sub-skills in class, then these assignments cannot be included in determining course grades. Redesigning report cards to be standards-based is complicated, but standards-based grading is simple: only grade-level assignments and assessments can be included in determining course grades. Also, report card grades cannot include non-standards-based factors such as effort, attendance, or behavior. The state must ensure that every high school teaches on grade level and that grading criteria is consistent throughout the state. Consistency of grading includes content standards covered plus weighting of quizzes, homework, final tests, projects, etc. If overall GPA is used as a measure, then grading must be standardized for all courses included in GPA.

The **reliability** of the system is achieved when a student would receive nearly the same score if he were to retake the test. The score should be the same if a different evaluator scores the test or a different version of the same test is given. Commercial tests are designed and field tested to be reliable. For grades and GPA to be reliable indicators, the student should receive the same grade no matter what teacher he has anywhere throughout the state.

Implementing standardized multiple measures testing will not increase student learning. Classroom instruction must be efficient and effective to provide *equal access* to the standards for all students and all student subgroups throughout the state. Of course, for any type of assessment to be fair to student subpopulations, the assessments must be free of racial, gender, and any other bias that could compromise students' ability to answer grade-level questions. However, **equity** for student subgroups is created by having all teachers successfully teaching the same grade-level content to all students in all schools in the state. Note that our Curriculum Calibration analysis of student assignments has discovered that standards-based education is not consistently implemented. Also, remediation for individual students must be in addition to grade-level classroom instruction not in place of grade-level instruction.

Cost factors for multiple measures include the added costs of collecting and analyzing multiple assessment results. Usually this is done through a database and software that uses a combining protocol to combine the multiple assessment results into a pass or fail determination. A difficulty we have found is maintaining the strict use of unique student ID numbers so assessments can be combined for each student using software. Additional costs in the future could include creating a standards-based grading protocol and training teachers to use it. When considering cost factors, keep in mind that developing a standards-based grading system in itself will not improve test scores…teaching all students on grade level will.

5. Does research evidence indicate whether multiple measures would be more valid for specific subpopulations of students—racial/ethnic groups, low income students, English language learners, and special education students—compared to a single test or measure such as GPA?

6. Is it reasonable to expect that using a multiple measures system (or strengthening our current system) would improve high school graduation rates?

7. Is there any research we should review on this topic? If so, please list.

The main purpose of multiple measures is to reduce testing errors that rely on a single high-stakes test. For example, a student might not be feeling well on the day of the test or be distracted by family or personal problems and then not do well on the state test.

We ran some scenarios using data we have from real high schools comparing state test results to multiple measures results. The measures used are shown below.

Multiple Measures Matrix

Mathematics	Weighting	Do Not Meet	Meet	Exceed
1. Math Course Grade	1/3	F-D	C	B-A
2. Math End-of-Course Test	1/3	0-69	70-79	80-100
3. Math State Test	1/3	Below Basic, Basic	Proficient	Advanced

We compared state test results to two multiple measures models: *compensatory*, in which high scores on one assessment can average out low scores on another, and *conjunctive*, where students must meet cut points in all assessments. In the conjunctive model used, students needed to score "basic" or above on the state test to pass. Students scoring "below basic" cannot pass no matter how high their scores are on the other assessments. Students failing the other assessments were deemed to not pass. The combining protocols are shown below.

State Test Only

Below Basic	Basic	Proficient	Advanced
Fail	Fail	PASS	PASS

Conjunctive Multiple Measures Combining Protocol

		Below Basic	Basic	Proficient	Advanced
80-100	A-B	Fail	PASS	PASS	PASS
	C	Fail	PASS	PASS	PASS
	D-F	Fail	Fail	Fail	Fail
70-79	A-B	Fail	PASS	PASS	PASS
	C	Fail	Fail	PASS	PASS
	D-F	Fail	Fail	Fail	Fail
0-69	A-B	Fail	Fail	Fail	Fail
	C	Fail	Fail	Fail	Fail
	D-F	Fail	Fail	Fail	Fail

Compensatory Multiple Measures Combining Protocol

		Below Basic	Basic	Proficient	Advanced
80-100	A-B	PASS	PASS	PASS	PASS
	C	PASS	PASS	PASS	PASS
	D-F	Fail	Fail	PASS	PASS
70-79	A-B	PASS	PASS	PASS	PASS
	C	PASS	PASS	PASS	PASS
	D-F	Fail	Fail	Fail	PASS
0-69	A-B	Fail	Fail	PASS	PASS
	C	Fail	Fail	Fail	PASS
	D-F	Fail	Fail	Fail	Fail

Below is a table showing the pass rates in mathematics for the schools overall and for subpopulations using the state test only and the two multiple measures. (Similar results were found for language arts.) The pass rates improved as the testing model changed from state test only to conjunctive to compensatory. The pass rate for School A went from 32.2% to 47.2% to 57.7%. School B's pass rate increased less dramatically from 30.1% to 32.6% to 36.6%.

Referring to the combining protocols above, it can be seen that the Conjunctive Combining Protocol model allows some students who scored "basic" on the state test to pass if they scored well on the other assessments. Likewise, the Compensatory Combining Protocol increased the pass rate even further since some students who previously scored "below basic" on the state test could now pass if they did well on the other assessments. However, as mentioned elsewhere, grades and tests are valid only if they are truly standards based. Often when students fail state tests yet have high grades, the classes are not teaching grade-level content. Also, although special education pass rates increased, their classes are often not based on grade-level content.

DataWORKS Educational Research
116 S. 7th Street, Fowler, CA 93625 ◆ (800)495-1550 ◆ www.dataworks-ed.com

Results of Three Methods of Measuring Student Achievement

MATH		State Test Only	Multiple Measures (Conjunctive)	Multiple Measures (Compensatory)
		Number of Students and pass rate	Number of Students and pass rate	Number of Students and pass rate
High School A	N	567	929	1096
	%	**32.2**	**47.2**	**55.7**
High School B	N	568	667	749
	%	**30.1**	**32.6**	**36.6**
English Learners	N	31	72	103
	%	14.6	27.1	38.7
Special Ed	N	25	74	138
	%	9.6	22.6	42.2
Free-Reduced	N	228	340	411
	%	26.6	35.0	42.3
African American	N	16	29	36
	%	15.1	24.8	30.8
Asian	N	92	188	198
	%	33.8	67.6	71.2
Filipino	N	71	90	100
	%	42.0	50.6	56.2
Hispanic	N	307	427	508
	%	25.6	31.6	37.6
White	N	586	845	984
	%	31.5	41.2	48.0

John Hollingsworth, President

DataWORKS Educational Research

Improving Student Achievement

Additional thoughts from DataWORKS

The state of Washington is trying to improve student achievement. Our company, DataWORKS Educational Research, was formed to use disaggregated state test results to improve student achievement. About five years ago we made a paradigm shift in our approach in accomplishing this goal: a shift away from just measuring and monitoring student achievement to measuring and monitoring how students are being taught. We asked, "Do you raise test scores by testing students or by teaching them?" When students learn more, test scores go up. This is accomplished by continually improving the effectiveness of classroom instruction.

As previously stated, our collection and analysis of over 2 million student assignments revealed that many students are taught below grade level, effectively preventing them from doing well on high-stakes state tests. In addition to looking at assignments, we went into 25,000 classrooms to measure, quantify, and analyze instructional practices being used. These observations revealed that effective research-based practices such as lesson design and delivery components, cognitive strategies (to help students remember), English learner strategies (to make English easier to understand), and differentiation (simplifying sub-skills so all students can be taught grade-level content) are not consistently implemented in the classroom. For example, effective checking for understanding is not being used in most classes. We have found that teachers call on the same hand-waving volunteers over and over and don't call on students randomly to take a sampling of the class to see if *everyone* is learning. Many teachers select students to answer before posing a question, so that other students don't even need to think of an answer. When teachers pose the question first, then provide thinking time, every student thinks of an answer whether called upon or not.

We believe the ultimate solution for improving test scores is preemptive: all students are successfully taught grade-level content every day…then test scores go up automatically. It's the continual improvement of input (curriculum alignment and classroom instruction) that produces better output (student achievement).

In summary, you need to have a fair, valid, and reliable testing system in place. Then focus on optimizing classroom instruction—the engine that improves student learning—and that will ultimately raise test scores.

John Hollingsworth, President

DataWORKS Educational Research

DataWORKS Educational Research
116 S. 7th Street, Fowler, CA 93625 ◆ (800)495-1550 ◆ www.dataworks-ed.com

Washington State's High School Graduate Requirements and Multiple Measures

Edward Kifer, Center for Advanced Study of Assessment

1. What does the concept of "multiple measures" mean with respect to assessments that are required for high school graduation? How might such a system be implemented?

We prefer to distinguish between notions of multiple measures and multiple methods. When two or more measures are used in an assessment context, one is using multiple measures. When two or more measures are used to measure the same or very similar constructs, one is usually using multiple methods. For example, the Washington Assessment of Student Learning (WASL) assesses reading, writing, and mathematics – multiple measures. It assesses reading with both constructed response items and multiple choice items and it assesses writing with both an essay and measures of conventions - multiple methods.

High school graduation requirements and exits criteria could be couched in either a multiple measures or multiple methods framework. To conform to assessment standards and commonsense, such graduation requirements should not be based on a student's performance on one measure. Either multiple measures or multiple methods or both should be included. In addition, a student should have multiple opportunities to demonstrate competence.

The particulars of the implementation of an assessment system depend on what specific measures and methods are used. Those in turn depend on the purpose or purposes of the assessment and how they are made operational. While the components of Washington's high school graduation requirements are explicit, a rationale or set of purposes for assessing some outcomes and not others is unclear. For example, "to know and apply concepts and principles" of mathematics is assessed directly in the Washington Assessment of Student Learning (WASL). It appears that those same skills are not measured in the arts in the WASL but assumed to be developed in high school course work. There seems to be no apparent measurement of "understanding the importance of work." It is helpful to give a rationale for such decisions; i.e. what to measure and how. One thing we note is that there appears to be no major emphasis on producing civic knowledge and encouraging civic participation, historically among the first defined outcomes of public schooling.

Each component of an assessment system presents measurement issues. What is to be measured, how it will be measured, and how it is weighted in a composite are both conceptual and technical issues to be addressed in the implementation phase of the policy. One needs proper rationales to deal with conceptual issues and proper methods to deal with the technical ones. We would choose components of a system according to the following criteria. The attributes are: easy to collect, represent a range of educational outcomes, provide multiple opportunities for success, and do not include much additional work for students, teachers and schools.

Decision rules for combining multiple indicators and making judgments about success have technical implications. There are important decisions to be made about how to combine and/or weight the various components of a multiple measures/multiple methods composite indicator. And, it is important to know that the influence of a component on the composite depends on how much variation there is in the indicator. For example, Washington' high school exit system includes an all or none judgment based on WASL score but includes other indicators that are allowed to vary. When there is an "all or none" indicator, it has no influence on the composite. If a person fails, they have no composite. If a person passes they have a composite score but it has no weight since every person has exactly the same score for that component. A pass means the component is present. On the other hand, if other indicators vary substantially, they will influence heavily the composite. How much an indicator varies determines its real weight in a composite.

2. What factors should policymakers consider when designing a multiple measures assessment system?

Purpose! When one is clear about the purpose or purposes of a policy other decisions are much easier to make. A serious discussion of why the state wants universal high school exit measures and why it wants it to assess each student at the state level, rather than making it a responsible of school districts, are non-trivial questions that should be addressed prior to any decision about the form of an assessment. The discussion should include potential side-effects and unintended consequences as well as main purposes.

3. Is Washington's assessment system already a multiple measures system? Explain why or why not. How could Washington's system be strengthened?

Yes. See question 1.

4. Could a high-stakes multiple measures system be implemented in a manner that is valid, reliable, cost effective, comparable in content and rigor to the WASL, culturally appropriate, and standardized across school districts? (These are the review criteria set by the legislature.)

We consider these to be good criteria for evaluating a multiple measures system. We would add to the list, however, a consideration of side-effects and potential unintended consequences. The research literature points to side-effects of high-stakes testing that include increased cheating by both teachers and students, teaching to the test and thereby narrowing the curriculum, questions about the rigor of the requirement (why cannot students pass a grade 10 test?), possibilities of score reporting errors that have important consequences for students, increased early attrition, and increased possibilities for litigation. In addition there are questions about whether alternate ways to meet requirements are comparably difficult and are fair.

It may be difficult for a new set of high school exit criteria, weighted according to some rubric, to have the same measurement properties as individual achievement tests. The WASL measures individuals in content areas with historically well-defined constructs and procedures. We suspect that WASL measurements are similarly to other tests in the same domain. These are well-known testing arenas and most professionals know where the land mines are. It is very likely, therefore, that the WASL has established desirable technical properties.

It is a much more complicated procedure to define a set of measures that reflects what should be the most basic outcomes of schooling. Even at the fundamental level, should each student leave school with equal proficiency in each of several diverse content areas? Can a student compensate for weaknesses in one area by extraordinary strengths in another? It further complicates the issue to decide the weights to be used to form a composite that reflects those outcomes. Finally, there is a problem of how a social indicator gets corrupted as it is used over time. Even with the technical issues solved, it is unlikely that results of a high school exit system a decade from now will reflect the same things it does now.

Another perspective or outside the box!

Internationally and historically high school exit examinations have been related to gaining access to higher education. Those seem not to be the purposes of the proposed Washington State system. It is not clear to us whether the purpose of the statewide requirements is a means to judge the effectiveness of schools, to assess the accumulated skills of graduates, or to set a minimum

competence level to be exceeded by each graduate. If the focus is the measurement of individuals, criteria like those above are relevant. As a consequence of that choice, the assessment is expensive, time consuming, and raises crucial fairness issues.

If, however, the state wished to measure the accumulated knowledge of its graduates, a very different assessment approach could be used. Local schools would determine who graduated or not but the state assessment would be designed to draw a picture of the accumulated knowledge of its graduates. The assessment would be NAEP-like or Gallop-like and not limited to 10th grade as is the WASL. For any given year or content area, the assessment would not be of the whole population of students; rather it would be a sample of students and schools.

The purpose of the assessment would be to describe well the outcomes of public schools in Washington and to track those outcomes across time and, perhaps, demographic groups. This is much like tracking a president's approval level across time and then breaking down the trend by party affiliation.

The design in a particular year might sample not only students and schools but also content areas and methods of measuring the outcomes. For example, one year the design would be to gather information about how well students write by assigning at random five or six different writing prompts to samples of students. In another year four of five performance assessments in science would be administered to the student samples. Each and every year could contain assessments of reading and mathematics if that were desirable. What was measured and how it or they were measured would form a set of policy questions to be answered by the many educational audiences in the state.

Accumulating this evidence across time with this sampling methodology would portray a broad array of schooling outcomes (much broader for the same given time and for less money than an assessment that does not sample) in Washington in a way that would allow policy makers to decide if what was being achieved was acceptable or needed to be changed. Across time there will be trajectories – is performance up or down? Are we getting the kind of responses that are desirable? What is considered desirable would drive what was assessed and how. This would allow Washington State to respond to criticisms about narrowing the curriculum and making crucial decisions about individuals with limited information.

In sum, this approach could provide much more and better information for less money in less time. And, it would eliminate the side-effects associated with high-stakes assessments.

The sampling approach to gathering some data does not preclude a census approach to others. There could be additional statewide surveys conducted less frequently. Also, other kinds of evidence could be collected and reported locally. There could be statewide standards setting minimums for local schools and districts in the latter scenario.

5. Does research evidence indicate whether multiple measures would be more valid for specific sub populations of students –racial/ethnic groups, low income students, English language learners and special education students –compared to a single test or measure such as GPA?

Different types of items and measures do show mean differences of varied magnitude. Boys tend to do better on some tasks; girls on others. A good assessment system includes a bias review designed to preclude systematic construct irrelevant differences in performance. Also there are statistical methods for determining whether items differentially affect performance of students

from various backgrounds and demographic groups. The bias review and the inspection of the items are used to insure that the method of measurement is not be related to mean score differences except if the method is measuring a different construct or just a part of a complex construct. Decisions about what should be measured and how should be decided by determining the purposes of the assessment.

The question of a single measure versus multiple measures is one of not making an important educational decision based on one indicator or outcome. In that sense multiple measures both meet the requirements of testing standards and are more valid.

6. Is it reasonable to expect that using a multiple measures system (or strengthening our current system) would improve high graduation rates?

   No.

7. Is there any research we should review on this topic? If so, please list.

   Below is a list of materials that might prove interesting.

References

1. AERA, APA, &NCME (1999) *Standards for educational and psychological testing.* Washington, DC. American Educational Research Association.
http://www.aera.net/AERAShopper/ProductDetails.aspx?productID=AERWSTDEPT

This is the "bible" of technical standards for testing and assessment. Although it responds to many more testing issues than posed by high school exit assessments, it does have pertinent information about what data should be collected to understand such an initiative.

2. Darling-Hammond, L., Rusttique-Forrester, E., Pecheon, R. L. (2005) *Multiple Measures Approaches to High School Graduation.* The School Redesign Network at Stanford University
http://www.srnleads.org/data/pdfs/multiple_measures.pdf

The report gives profiles of state assessment programs including those with high school graduation requirements that include some type of exit assessment. It makes a strong argument for the efficacy of multiple measures.

3. ED468488 Title: Multiple Measures: Examination of Alternative Design and Analysis Models. Authors: Henderson, Dianne L.; Julian, Marc W.; Yen, Wendy M. (This is an ERIC document.)
http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1a/61/4e.pdf

This paper defines multiple measures (there are many definitions), describes what forms the multiple measures might take, and discusses various ways that multiple measures can be combined to form a composite measurement. It discusses the consequences of choosing a particular method of weighting.

4. Educational Commission of the States
http://www.ecs.org/html/IssueSection.asp?issueid=108&subIssueid=159&ssID=0&s=What+States+Are+Doing&archive=1

As the third in an annual series on exit exam policies, this report provides state by state information, wide-ranging and thorough, on their exit exam policies. The report gives recommendations for desirable properties of such policies.

5. FairTest - The National Center for Fair & Open Testing
http://www.fairtest.org/index.htm

FairTest opposes narrowly-tailored high-stakes assessments. It weighs-in on most national assessment issues. Its views are worth considering.

6. Haney, W., et. al. (January 2004) *The National Board on Educational Policy and Testing,*The Education Pipeline in the United States 1970–2000.
http://www.bc.edu/research/nbetpp/statements/nbr3.pdf

This report suggests that exit assessments lead to lower graduation rates and early attrition. It has many features of interest and is a bit dated. Its longitudinal data provide a basis for a powerful report.

7. High Stakes: Testing for Tracking, Promotion, and Graduation (1999) *Commission on Behavioral and Social Sciences and Education (CBASSE).* The National Academies Press
http://books.nap.edu/openbook.php?record_id=6336&page=308

This is another report commissioned by a highly-esteemed organization. Although not exclusively devoted to exit assessments, it includes a discussion of issues surrounding high school graduation requirements.

8. Koretz, Daniel (2003) Using Multiple Measures to Address Perverse Incentives and Score Inflation. *Educational Measurement: Issues and Practice* 22 (2), 18–26.

This could be called a diatribe against single measures, especially those that can produce skewed results. Multiple measures are highly recommended for virtually all assessment contexts.

9. Noah, H. J. Eckstein, M.A. (1989) Tradeoffs in Examination Policies: An International Comparative Perspective , *Oxford Review of Education*. Vol. 15, No. 1, pp. 17-27

Although somewhat dated, this paper gives a context for exit examinations. In these eight systems exit examinations were related to entrance to higher education. Still, there is great variability internationally in terms of practices even though the issues surrounding such examinations are almost universal. Someone who is interested in the topic and wishes to see huge differences in practice should "google" some of the education systems that were once part of the Soviet Union.

10. Wise, Lauress L. et. al. (2005) *Independent Evaluation of the California High School Exit Examination (CAHSEE):* Evaluation Report Executive Summary, California Department of Education Sacramento, CA Contract Number: 00-07
http://www.cde.ca.gov/ta/tg/hs/documents/y3exesum.pdf

This is an evaluation report of California's controversial high school exit exam (CAHSEE). It covers an array of issues arising from the quality of the items, to scaling and equating, to pass rates. It demonstrates the range of reliability and validity issues that must be addressed when assessing outcomes in a high-stakes environment.

Edward Kifer

Center for Advanced Study of Assessment
Georgetown College
Georgetown, Kentucky
November 7, 2007