

Relationships Between Performance on the 2003  
National Assessment of Educational Progress and  
Washington Assessment of Student Learning

Kathryn L. Sprigg

Washington State NAEP Coordinator

Office of Superintendent of Public Instruction

Olympia, Washington

Paper presented at the annual meeting of the American Educational Research Association

April 15, 2005, Montreal, Canada

Abstract

This study compared relationships between student performance on the 2003 National Assessment of Educational Progress (NAEP) and Washington Assessment of Student Learning (WASL). The assessments were similar in design and process; it was hypothesized that they were predictive of one another and that results from the NAEP may be useful for informing policy decisions at the state level.

The study used data from matched samples of students taking the 2003 NAEP and WASL. Data were compared using linear regression analyses. In all analyses WASL scores were used as the dependent variable, the first NAEP composite plausible values were used as independent variables.

The NAEP and WASL assessments had different scoring scales; however, they were positively correlated. The correlation between grade 4 reading scores was moderately strong. Regression resulted in an effect size of  $r = .602$ ; the unstandardized  $B$  coefficient was .338, statistically significant at  $p = .000$  (one-tailed). The correlation between grade 4 mathematics scores was .684; the unstandardized  $B$  coefficient was .839. The effect size was stronger than reading and  $r^2$  was .468, statistically significant at  $p = .000$  (one-tailed).

Analyses were also performed for students taking the 2003 eighth grade NAEP. These results were compared with students' 2002 seventh grade WASL scores. The analyses showed similar results to fourth grade. Correlation for reading performance was .613; mathematics performance correlated higher at .764. Regression resulted in  $r^2$  values of .376 for reading and .583 for mathematics; both were statistically significant at  $p = .000$  (one-tailed).

### Introduction

In 1993, the Washington State Legislature passed Engrossed Substitute House Bill 1209, adopted as the Education Reform Act of 1993, RCW28A.630.885 (MacQuarrie, 2003; Office of Superintendent of Public Instruction, 2002b). The purpose of the legislation was to raise academic standards by building a foundation of basic skills in reading, writing, communication, and mathematics, and to “assure students more fully master these basic skills as the tools to success in later learning, life, and work” (Office of Superintendent of Public Instruction, 2002b). The law was intended to provide students with the chance to develop the skills and abilities to “become responsible citizens, contribute to their own economic well-being and to their families and communities, and enjoy productive and satisfying lives” (Office of Superintendent of Public Instruction, 2002b). Mandates of the act included the development of curriculum standards and an assessment system to measure acquisition of those standards. By 1997, Washington had developed the Essential Academic Learning Requirements (EALRs) and was in the process of implementing the Washington Assessment of Student Learning (WASL), benchmark assessments aligned to the EALRs, at grades 4, 7, and 10.

In 2001, the United States Congress passed Public Law 107-110, the No Child Left Behind Act of 2001 (NCLB Act). Part of the provisions of the act mandated that all states participate in biennial reading and mathematics assessments in grades 4 and 8; the vehicle for this participation was the National Assessment of Educational Progress (NAEP). Other parts of the legislation required that states develop a statewide testing program to assess reading and mathematics in grades 3 through 8 and once in high school. Further, states were required to maintain adequate yearly progress (AYP) towards the goal of closing the achievement gap

between advantaged and disadvantaged students (Office of Superintendent of Public Instruction, 2002b). The NAEP was expected to be used to provide confirmatory evidence of state reported AYP.

### Background

Linn (1994) described state-by-state progress as educators began to implement minimum competency testing reforms as early as the 1970's. By 1983, 34 states had some form of testing, but the focus tended to be on basic skills that were generally accepted as reasonable for purposes of establishing high school graduation requirements. Testing also was given a significant role in reforms advocated in "A Nation at Risk" (1983), one of the reports that set off a flurry of educational reform efforts across the country in the early 1980s. After the report was published, nearly every state introduced reforms that either mandated new test requirements or expanded the use of existing testing programs; testing and evaluation became a primary engine for carrying out educational policy (Linn, 1994).

Subsequent reform efforts continued to emphasize accountability in educational systems (Linn, 1994), although some new requirements were added. The most important of these included changes in the types of questions used on the assessments, inclusion of all students in the assessments, and the development and adoption of standards that both direct the focus of the assessments and define performance levels. The new types of assessments have been referred to as authentic, direct, or performance-based assessments (Linn, 1994). The fundamental difference between the older, standardized tests and the newer, performance-based assessments was the change in scoring methods. The older tests consisted of multiple-choice items, which could be scored by machine; the newer tests combined several types of assessments including the construction of short and extended responses and

conducting experiments. These newer assessments are scored by panels of judges trained to look for responses judged appropriate according to an approved scale (Linn, 1994). High-stakes testing and assessment has tended to shape learning and instruction and the performance-based assessments are believed to be more compatible with new, constructivist theories of learning that consider learners as active participants in their knowledge acquisition process, rather than simply passive receivers of information (Linn, 1994; Shephard, 1991).

In an atmosphere of educational reform and federally mandated testing, Washington State has been concerned about policy implications and the effects of the NCLB Act on its state assessment program (Heuschel & Harmon, 2003). One of the mandates of the NCLB Act requires states to demonstrate adequate yearly progress (AYP) towards the goal of all children demonstrating proficiency on state tests by 2014 (Heuschel & Harmon, 2003). The NCLB Act also mandates participation in the NAEP at grades 4 and 8, in reading and mathematics (Office of Superintendent of Public Instruction, 2002b; National Assessment Governing Board, "What is NAEP?", n.d.). Although achievement levels for the NAEP are not aligned with the WASL achievement levels, the NAEP is expected to be used as a confirmatory measure for state AYP reports. This situation has led quite naturally to interest on the part of policy developers in possible relationships between the national and state assessments. Hence, the purpose of this study was to investigate the relationships between the NAEP and WASL assessments.

### Conceptual Underpinnings for the Study

The National Assessment of Educational Progress (NAEP) and the Washington Assessment of Student Learning (WASL) are both performance-based assessments.

Literature on the NAEP and WASL indicated that the development processes used by the two assessments were similar. Each was developed by a group of individuals from business, education, government, and the public (National Assessment Governing Board, 2002; Anderson, Ensign, & MacQuarrie, 2002). Both systems include a similar ongoing process designed to maintain the integrity and validity of the assessments (Anderson et al., 2002; National Assessment Governing Board, 2002a & b; Taylor, 2002a & b; Thompson-Klenk, 2002).

Each assessment was created by groups of stakeholders that were knowledgeable about assessment and educational reform. Groups of experts each developed the frameworks for the NAEP and WASL tests using standards set through collaboration with curriculum experts. Assessment specialists developed and maintained thousands of items to use in pilot testing, then collected data and reported results on the individual tests (Anderson et al., 2002; National Assessment Governing Board, 2002a & b). The NAEP assessment tests a sample of students in each state, using the data to report results that can be used to compare individual states to past performance, states to other states, and states to the national average. Results are not available for individual students, schools, or districts (National Assessment Governing Board, "How Does NAEP Work?", n.d.). The WASL assessment tests all eligible children in Washington State (Office of Superintendent of Public Instruction, 2002a). A study of the alignment of the WASL to the NAEP has not been conducted, although interest has been indicated, and several states have also expressed interest in linking their assessments to the NAEP (NAEP Validity Studies Panel, 2002).

Because the assessment processes are similar, and the assessments themselves are similar, questions have been posed about whether NAEP scores can be used to predict

success on the WASL. While research has been conducted to examine relationships between the WASL and other tests (see for example, McGhee, 2002; Pavelcheck, Stern, & Olson, 2002), no recent study has been conducted to consider the correlation between the WASL and NAEP. With the increased emphasis on the use of the NAEP as an external measure of state progress in closing the achievement gap, there is interest in how well the NAEP can confirm student achievement in Washington State.

Although researchers warn that comparing one test to another can be problematic because tests that are dissimilar or that test different constructs can lead to inappropriate assumptions (Feuer et al., 1999), we decided to proceed with this study to investigate relationships between the WASL and the NAEP and to determine whether NAEP results might be used as an appropriate confirmatory measure for WASL scores. Specific research questions of interest were: 1. What relationships exist between student achievement on the NAEP and WASL assessments? 2. To what extent can performance on the WASL be inferred by performance on the NAEP?

#### Study Design, Methods, Data Collection, and Analyses

This study is based on Washington State and NAEP archival data made available to the researcher by the Office of the Superintendent of Public Instruction (OSPI) and the National Center for Education Statistics (NCES). Study subjects were Washington State students at schools that participated in the 2003 NAEP fourth and eighth grade reading and mathematics assessments. Effectively, all Washington State students participated in the WASL, since it is mandated by the state legislature. With NAEP, however, only a randomly selected sample of students participated. The sample population for this study was approximately 10,000.

Sophisticated sampling procedures were used to select approximately 200 schools for the 2003 grades 4 and 8 NAEP. Schools were selected to take part based on stratified random sampling techniques coordinated by Westat, the contractor that administers the NAEP assessment for the NCES (Rust, 2002). School sampling was based on two factors: level of urbanization, or location type, and minority characteristics of the student population. At each school selected, approximately 60 students were assessed in either reading or mathematics. The student scores from the WASL were available, since all grade 4 and 7 students in public schools in Washington State are required to sit for the assessment.

Students who took the 2003 NAEP assessment were identified using copies of the administration schedules for each NAEP sampled school. Using NAEP test booklet numbers, researchers created a list of students and extracted their 2003 or 2002 WASL scores. After creating the file, individual student identifier information was removed and the file was sent to the American Institutes for Research for addition of NAEP data.

#### Analyses of Assessment Findings

After the file was returned to Washington, scores on the NAEP and WASL were queried. Information for the general population, accessed through the Washington State Superintendent of Public Instruction website, was compared to the NAEP sample population and analyzed based on grade level comparisons. Data were compared using linear regression analyses. In all analyses WASL scores were used as the dependent variable, the first NAEP composite plausible values were used as independent variables.

Grade 4 Demographics

The number of records with NAEP fourth grade data (NAEP sample population) totaled 3,633 for reading and 3,839 for mathematics. Table 1 contains comparative demographic information.

Table 1

Demographics for Washington 4<sup>th</sup> Grade Students Compared to NAEP Participating Students

<b>2003 Demographic Information Grade 4 *</b>						
	<b>Total WASL population<sup>a</sup> N = 77,146</b>		<b>NAEP sample population<sup>b</sup> Reading n = 3,633</b>		<b>NAEP sample population<sup>b</sup> Math n = 3,839</b>	
White	54726	70.9%	2560	70.5%	2135	55.6%
Black	4620	6.0%	226	6.2%	163	4.2%
Hispanic (Latino)	9860	12.8%	425	11.7%	882	23.0%
Asian Pacific Islander	5859	7.6%	288	7.9%	231	6.0%
Native American	2081	2.7%	97	2.7%	147	3.8%
Other	---	---	23	0.6%	280	7.3%
<b>Gender</b>						
Male	39833	51.6%	1832	50.4%	1972	51.4%
Female	37313	48.4%	1801	49.6%	1867	48.6%
<b>Assessment Accommodations</b>						
Students with Disabilities Reading	8989	12.0%	349	9.6%		
SwD Math	9102	11.8%			446	11.6%
Limited English Proficient Reading	5529	7.4%	224	6.2%		
LEP Math	5587	7.2%			219	5.7%
<b>Free / Reduced Lunch Program</b>						
Eligible	24806	33.1%	1351	37.1%	1381	36.0%
Not Eligible			1886	51.9%	1997	52.0%
Information Not Available			400	11.0	461	12.0%
<b>Title 1 Program</b>						
Participating	10503	14.0%	1174	32.3%	1176	30.6%
Not Participating			2446	67.3%	2649	69.0%
Information Not Available			13	.4%	14	.4%

<sup>a</sup> Available: <http://reportcard.k12.wa.us>

<sup>b</sup> Available: <http://nces.ed.gov/nationsreportcard>

\* Information missing or not available for some records, not all sum to sample population totals

As shown in the table, while Westat's sampling methodology produced generally representative demographics for reading data, there were differences in the mathematics sample population. Most notably, the percentage of Hispanic students taking the NAEP fourth grade reading assessment (11.7%) was lower than the state's population (12.8%), but a substantially higher percentage of Hispanic students (23%) took the NAEP mathematics assessment. This difference resulted in lower percentages of White, Black, and Asian/Pacific Island students taking the mathematics assessment. Native American students represented a higher percentage of the sample population as well (3.8%). Also of note, NAEP includes a race/ethnicity demographic category not represented in WASL assessments, other; the percentage of mathematics students choosing this option was 7.3%.

Demographics for students requiring assessment accommodations also differed; a smaller percentage of special education students (-2.4%) took the reading assessment, while the percentage of special education students taking the NAEP mathematics assessment (11.6%) was similar to the total WASL population of special education students who took the mathematics WASL (11.8%). English language learners represented 6.2% of the NAEP reading sample and 5.7% of the mathematics group, compared with 7.4% and 7.2%, respectively, of the total WASL population. Information for comparisons of socio-economic factors, represented by Title 1 and free or reduced lunch participation, while different, was not considered as reliable since Washington had not yet systematically collected the data. It does appear, however, that a larger number of students in Title 1 programs were tested for the NAEP than the percentage represented in the general WASL population. The percentage of students eligible for free or reduced lunch was closer; 33.1% of the general WASL population was eligible, 37.1% for NAEP reading, and 36% for NAEP mathematics. These

differences may be a random result of sampling bias, the result of state data collection issues, or could be because schools excused children in special education and English language learners from the NAEP reading assessment in larger numbers than from the WASL.

After completing one to one matches for fourth grade NAEP with WASL records, a total of 3,049 records remained for reading; 3092 were matched for mathematics data.

### Grade 8 Demographics

Fewer records with NAEP eighth grade data were matched than for grade 4. A query resulted in 2,625 records for reading and 2,629 records for mathematics (NAEP sample population). Demographics for the 7<sup>th</sup>/8<sup>th</sup> grade sample indicated similar results to the NAEP fourth grade mathematics sample. In the ethnicity category, ‘other’ represented approximately 7% of the population. A substantially larger percent of Hispanics were represented in the NAEP samples; this resulted in lower percentages of Whites and Blacks in the survey. Assessment accommodation percentages appeared to be closer to the general population on the 7<sup>th</sup>/8<sup>th</sup> grade demographics than they were on the fourth grade sample (see Table 1). As with the fourth grade demographics, lunch data and Title 1 data were substantially different. Some of the differences in these demographics may be a result of the difference in grade level status from seventh grade WASL to eighth grade NAEP. Table 2 contains comparative demographic information for the 7<sup>th</sup>/8<sup>th</sup> grade match.

Table 2

## Demographics Washington 7/8 Grade Students Compared to NAEP Participating Students

<b>2003 Demographic Information Grade 7/8</b>						
	<b>Total WASL population<sup>a</sup> 2002 Grade 7 N =78,779</b>		<b>NAEP sample population<sup>b</sup> Grade 8 Reading n =2,625</b>		<b>NAEP sample population<sup>b</sup> Grade 8 Math n = 2,629</b>	
White	57,443	72.9%	1,560	59.4%	1,590	60.5%
Black	4,085	5.2%	114	4.3%	127	4.8%
Hispanic (Latino)	8,261	10.5%	489	18.6%	482	18.3%
Asian Pac Islander	5,855	7.4%	189	7.2%	188	7.2%
Native American	2,097	2.7%	82	3.1%	59	2.2%
Other			190	7.2%	183	7.0%
<b>Gender</b>						
Male	40,327	51.2%	1,330	50.7%	1,308	49.8%
Female	38,492	48.9%	1,295	49.3%	1,321	50.2%
<b>Assessment Accommodations</b>						
Students with Disabilities Reading	8,671	11.0%	265	10.1%		
SwD Math	8,753	11.1%			282	10.7%
Limited English Proficient Reading	3,478	4.4%	90	3.4%		
LEP Math	3,511	4.5%			112	4.3%
<b>Free / Reduced Lunch Program</b>						
Eligible	12,168	15.4%	745	28.2%	700	26.6%
Not Eligible			1,538	58.6%	1,564	59.5%
Information Not Available			342	13.0%	365	13.9%
<b>Title 1 Program</b>						
Participating	3,871	4.9%	297	11.3%	288	11.0%
Not Participating			2,324	88.5%	2,337	88.9%
Information Not Available			4	0.2%	4	0.2%

<sup>a</sup> Available: <http://reportcard.k12.wa.us>

<sup>b</sup> Available: <http://nces.ed.gov/nationsreportcard>

\* Information missing or not available for some records, not all sum to sample population totals

### Student Performance

After completing one to one matches for 2003 fourth grade NAEP with WASL records, a total of 3,049 records remained for reading; 3092 records were matched for mathematics data. Matches for students taking the 2002 Grade 7 WASL and 2003 Grade 8 NAEP resulted in 2,075 records for reading and 2,076 records for mathematics.

Results for the NAEP assessments were available on the NCES website, as well as from the Washington State NAEP data file. Scores for the matched NAEP sample were only available from the Washington State NAEP data file. Performance across grades was considered in light of differences in scoring scales. The 2003 NAEP used a vertical scale from 0 to 500, placing fourth and eighth grade results on the same scale. The WASL scales are separate for each grade; they range from 150 to 600. While scores for the matched groups were slightly higher than for the NAEP sample population, comparisons between fourth grade reading and mathematics achievement levels were mixed.

### Grade 4 Performance

Mean NAEP reading scores for the larger NAEP sample population ( $N_2 = 3,633$ ) were approximately one point lower than mean scores for the matched reading group ( $n = 3,049$ ) who participated in both the NAEP and the WASL ( $M$ 's = 221.2 and 222.3;  $SD$ 's = 34.9 and 34.5 respectively). Mean NAEP mathematics scores for the larger NAEP sample population ( $N_2 = 3,839$ ) were approximately one point lower than mean scores for the matched mathematics group ( $n = 3,092$ ) who participated in both the NAEP and the WASL ( $M$ 's = 238.4 and 240.0;  $SD$ 's = 27.4 and 26.6 respectively). Table 3 includes the mean scores, standard deviation, and number, as well as percentages at each achievement level, of students who took the fourth grade reading or mathematics portion of the WASL and the NAEP.

Table 3

Performance of Grade 4 Washington Students Compared to NAEP Participating Students

<b>WASL NAEP Performance Data Grade 4 Reading</b>				
	Total public school population	NAEP sample population	Matched reading group	
Scale Scores	WASL <sup>a</sup>	NAEP <sup>b</sup>	WASL	NAEP
N	75,177	3,633	3,049	
Mean	405.8	221.2	407.0	222.3
SD	20.61	34.9	19.4	34.5
Performance Levels				
4 Advanced	24.0%	7.0%	25.0%	8.8%
3 Proficient	42.7%	26.0%	45.5%	26.3%
2 Basic	25.3%	34.0%	24.7%	35.0%
1 Below Basic	6.4%	33.0%	4.8%	30.0%
<b>WASL NAEP Performance Data Grade 4 Mathematics</b>				
	Total public school population	NAEP sample population	Matched math group	
Scale Scores	WASL <sup>a</sup>	NAEP <sup>b</sup>	WASL	NAEP
N	75,364	3,839	3,092	
Mean	400.8	238.4	402.9	240.0
SD	34.43	27.4	32.6	26.6
Performance Levels				
4 Advanced	26.7%	5.0%	22.4%	5.8%
3 Proficient	28.5%	31.0%	24.4%	33.4%
2 Basic	24.0%	45.0%	20.0%	45.5%
1 Below Basic	19.4%	19.0%	13.7%	15.3%

<sup>a</sup> Available: <http://reportcard.k12.wa.us><sup>b</sup> Available: <http://nces.ed.gov/nationsreportcard>

Table 3 shows that 66.7% of the total population of fourth grade students were at proficient or advanced performance level (Level 3 or 4) on the reading WASL, the expectation for meeting or exceeding standard (Office of Superintendent of Public Instruction, 2003a). In the matched NAEP sample, a higher percentage of students, 70.5%, met or exceeded the WASL reading standard. Mathematics showed opposite results. In 2003, 55.2% of the total population of fourth grade students met or exceeded standard on the

mathematics WASL; 46.8% of the matched NAEP sample met or exceeded the WASL standard.

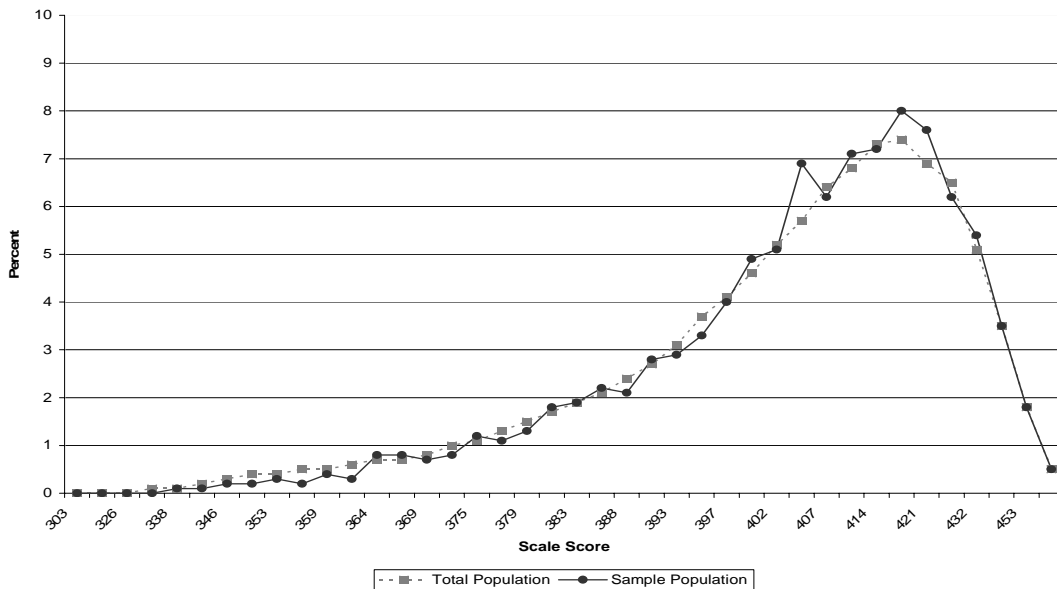
As discussed previously, compared to the WASL, a lower percentage of students in special education programs and English language learners participated in the NAEP; this may be a contributory reason that the mean NAEP scores for the matched reading group were slightly higher than for the larger population of NAEP test takers.

For purposes of considering possible sample bias, a frequency table was then created to compare performance of the total number of valid scores on the 2003 WASL fourth grade reading test ( $N = 75,177$ ) to the performance of the matched reading group ( $n = 3,049$ ).

Figure 1 shows the results of the scale score comparison.

Figure 1

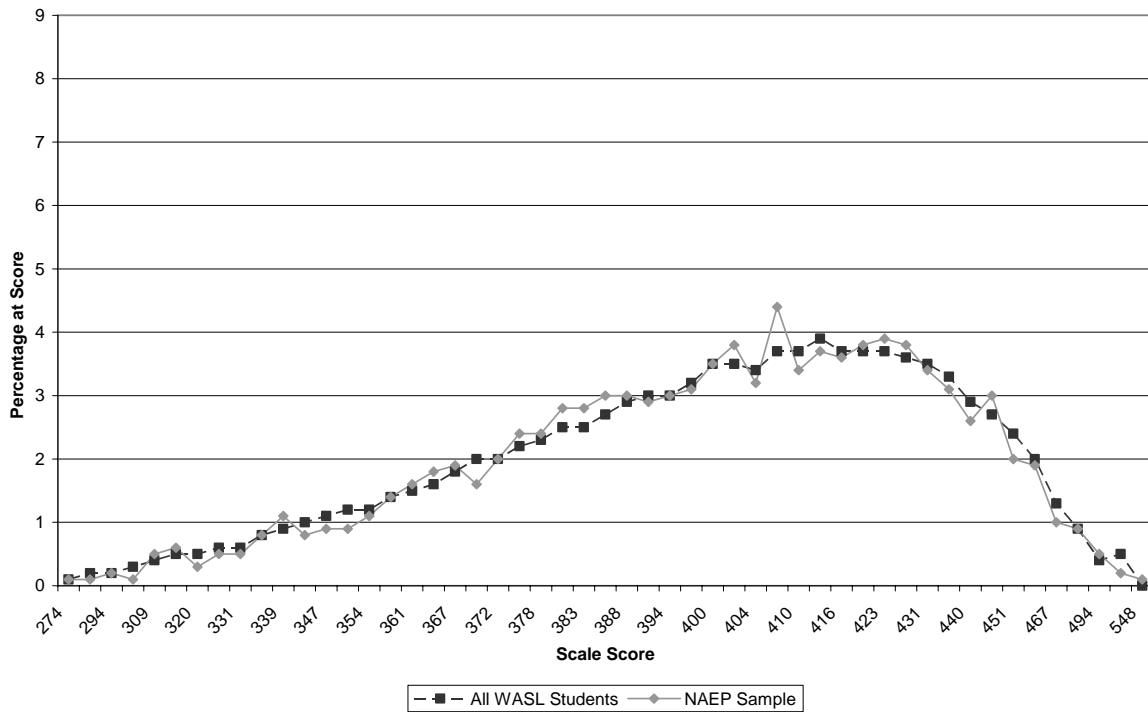
Percent Distribution of WASL Reading Scale Scores for all 2003 Fourth Grade Students and Matched NAEP/WASL Group



A frequency table was also created to compare performance of the total number of valid scores on the 2003 WASL fourth grade mathematics test ( $N = 75,364$ ) to the performance of the matched mathematics group ( $n = 3,092$ ). Figure 2 shows the results of the mathematics scale score comparison.

Figure 2

Percent Distribution of WASL Mathematics Scale Scores for all 2003 Fourth Grade Students and Matched NAEP/WASL Group



The WASL performance of students in the matched groups closely matched the WASL performance for the total fourth grade student population with exceptions near the cut points for Level 3, Proficient (400), and Level 4, Advanced (421). More students in the matched groups scored at these levels than in the total student population. The similarities in

performance results between the groups suggested that the NAEP sample was representative of Washington's fourth grade population and that this area of bias concern was not a major issue.

Next, linear regression comparing NAEP to WASL fourth grade scores indicated that students who scored higher on the NAEP tended to score higher on the WASL. Table 4 presents the results of the regression analysis.

Table 4

Summary of Regression for 2003 Grade 4 NAEP / WASL Results ( $N = 3,049, 3092$ )

Variable	<i>M</i>	<i>SD</i>	$r^2$	<i>B</i>	<i>SE B</i>	<i>r</i> (Beta)	<i>SEE</i>	<i>F</i>	<i>p</i> value
NAEP reading score <sup>a</sup>	222.3	34.5	.363	.338	1.83	.602	15.49	1733	.000
WASL reading score <sup>b</sup>	407	19.4							
NAEP math score <sup>a</sup>	240.0	26.6	.468	.839	3.89	.684	23.78	2717	.000
WASL math score <sup>b</sup>	403	32.6							

<sup>a</sup> NAEP scale is 0 – 500

<sup>b</sup> WASL scale is 150 - 600

Analysis of the matched reading group with both NAEP and WASL scores ( $N=3,049$ ) resulted in a mean score for the NAEP of 222.3; the WASL mean score was 407 ( $SDs = 34.53$  and  $19.40$  respectively). Results of the analysis indicated that performance on the NAEP and WASL assessments was positively correlated ( $B = .338$ ). Each additional 10 points on the NAEP scale predicted an increase of about 3.4 points on the WASL. A Pearson

correlation indicated that the relationship between performance on the two assessments was moderately strong,  $r = .602$ , and statistically significant,  $p = .000$  (one-tailed). Given the size of the sample, the range of the estimated performance was again small (95 percent confidence interval: [.322, .354]).

Analysis of the matched mathematics group with both NAEP and WASL scores ( $N=3,092$ ) resulted in a mean score for the NAEP of 240.0; the mean WASL score was 403 ( $SDs = 26.57$  and  $32.59$  respectively). Results of the analysis indicated that performance on the NAEP and WASL assessments was positively correlated ( $B = .839$ ) for mathematics as well as reading. Each additional 10 points on the NAEP mathematics scale predicted an increase of about 8.4 points on the WASL. A Pearson correlation indicated that the relationship between performance on the two mathematics assessments was moderately strong,  $r = .684$ , and statistically significant,  $p = .000$  (one-tailed). Given the size of the sample, the range of the estimated performance was small (95 percent confidence interval: [.808, .871]).

As shown in Table 4, performance on the reading and mathematics NAEP and the WASL were positively correlated. The effect size was .363 for reading, indicating that the model explained approximately 36% of the variation in the individual score performance on the WASL reading assessment ( $r^2 = .363$ ). The effect size for mathematics was .468; the model explained approximately 47% of the variation in WASL mathematics performance ( $r^2 = .468$ ). An analysis of variance returned  $F$  values of 1733 and 2717 ( $p = .000$ ) respectively, indicating that the probability of the relationships occurring by chance was very small.

Grade 8 Performance

Analyses of performance on the 2003 grade 8 NAEP were more problematic. The WASL is administered at grade 7, a year earlier for those students who sat for the eighth grade NAEP. Students would be expected to perform better on an assessment administered one year later, by virtue of a year's worth of instruction. Table 5 includes the mean score, standard deviation, and number, as well as percentages at each achievement level, of students who took the reading or mathematics portion of the 2003 NAEP and the 2002 WASL.

Table 5

Performance of Washington Students Grade 7 WASL Compared to Grade 8 NAEP Results

<b>Performance Data WASL Grade 7 – NAEP Grade 8 Reading</b>				
Scale Scores	Total public school population Grade 7 WASL <sup>a</sup>	NAEP sample population NAEP <sup>b</sup>	Matched reading group Grade 7 WASL	Grade 8 NAEP
N	75,262	2,625	2,075	
Mean	395.3	264.0	396.3	265.1
SD	19.9	34.4	18.5	33.8
Performance Levels				
4 Advanced	14.2%	3%	13.9%	3.2%
3 Proficient	30.4%	30%	32.7%	32.0%
2 Basic	37.9%	43%	40.8%	41.9%
1 Below Basic	13.2%	24%	12.6%	23.0%
<b>Performance Data WASL Grade 7 – NAEP Grade 8 Mathematics</b>				
Scale Scores	Total public school population Grade 7 WASL <sup>a</sup>	NAEP sample population NAEP <sup>b</sup>	Matched math group Grade 7 WASL	Grade 8 NAEP
N	75,354	2,629	2,076	
Mean	374.3	281.5	378.1	283.7
SD	48.4	35.9	48.2	35.1
Performance Levels				
4 Advanced	13.2%	6%	15.6%	7.7%
3 Proficient	17.2%	26%	17.1%	27.4%
2 Basic	17.1%	40%	18.4%	40.4%
1 Below Basic	50.2%	28%	48.9%	24.6%

<sup>a</sup> Available: <http://reportcard.k12.wa.us>

<sup>b</sup> Available: <http://nces.ed.gov/nationsreportcard>

Linear regression comparing 2003 NAEP eighth grade to 2002 WASL seventh grade scores indicated that students who scored higher on the NAEP tended to score higher on the WASL. Table 6 presents the results of the regression analysis.

Table 6

## Summary of Regression for 2003 Grade 8 NAEP / 2002 Grade 7 WASL Results

( $N = 2,075, 2,076$ )

Variable	<i>M</i>	<i>SD</i>	$r^2$	<i>B</i>	<i>SE B</i>	<i>r</i> (Beta)	<i>SEE</i>	<i>F</i>	<i>p</i> value
NAEP reading score	265.1	33.8	.376	.336	2.54	.613	14.63	1249	.000
WASL reading score	396	18.5							
NAEP math score	283.7	35.1	.583	1.05	5.57	.764	31.1	2901	.000
WASL math score	378	48.2							

<sup>a</sup> NAEP scale is 0 – 500

<sup>b</sup> WASL scale is 150 - 600

Analysis of the matched reading group with NAEP and WASL scores ( $N = 2,075$ ) resulted in a mean score for the NAEP of 265.1; the WASL mean score was 396.3 ( $SDs = 33.8$  and  $18.5$  respectively). Results of the analysis indicated that performance on the NAEP and WASL reading assessments was positively correlated ( $B = .336$ ). Each additional 10 points on the NAEP scale predicted an increase of about 3.4 points on the WASL. A Pearson correlation indicated that the relationship between performance on the two assessments was moderately strong,  $r = .613$ , and statistically significant,  $p = .000$  (one-

tailed). Given the size of the sample, the range of the estimated performance was small (95 percent confidence interval: [.318, .355]).

Analysis of the matched mathematics group with NAEP and WASL scores ( $N=2,076$ ) resulted in a mean score for the NAEP of 283.7; the mean WASL score was 378.1 ( $SDs = 35.06$  and  $48.19$  respectively). Results of the analysis indicated that performance on the NAEP and WASL mathematics assessments was also positively correlated ( $B = 1.05$ ). Each additional 10 points on the NAEP mathematics scale predicted an increase of about 10 points on the WASL. A Pearson correlation indicated that the relationship between performance on the two mathematics assessments was strong,  $r = .764$ , and statistically significant,  $p = .000$  (one-tailed). Again, given the size of the sample the range of the estimated performance was small (95 percent confidence interval: [1.012, 1.088]).

As shown in Table 6, performance on the reading and mathematics NAEP and the WASL were both positively correlated. The effect size was .376 for reading, indicating that the model explained approximately 37% of the variation in the individual score performance on the WASL reading assessment ( $r^2 = .376$ ). The effect size for mathematics was .583; the model explained approximately 58% of the variation in WASL mathematics performance ( $r^2 = .583$ ). An analysis of variance returned  $F$  values of 1249 and 2901 ( $p = .000$ ) respectively, indicating that the probability of the relationships occurring by chance was very small.

### Discussion of Findings

The positive relationships between the NAEP and WASL suggest that the NAEP may be an appropriate measure when used to confirm student achievement on the WASL. The assessments are quite similar in design, structure, and content; however, a thorough linkage

analysis has not been performed. The assessments do represent different intentions; the WASL is designed to provide individual student data for use at the school and classroom level and, at grade 10, is used to establish eligibility for a high school diploma. Additionally, the NCLB Act requires the use of WASL results to report, for grades 3 through 8 and once in high school, Adequate Yearly Progress (AYP) towards reaching the federal goal of 100% of students scoring proficient or higher on state assessments by 2014. Failure to meet AYP goals triggers punitive action and eventually requires restructuring at the school, district, and state levels (Heuschel & Harmon, 2003). Thus, the WASL can be called a high stakes assessment.

The NAEP is intended to provide data to compare the progress of states and the nation over time and to provide the U.S. Department of Education confirmatory information regarding reported AYP. It could be argued that the recent NCLB legislation that provides sanctions for states, districts, and schools and the use of results from the NAEP as a confirmatory measure to validate reported state AYP qualifies the NAEP as a high stakes assessment as well. Thus, the ability of the NAEP to predict performance on the WASL, and the use of NAEP results as confirmatory evidence of WASL progress, is an important matter. Examination of assessment results sought to consider whether NAEP performance could be used as valid evidence to confirm progress on the WASL.

The relationships between reading and mathematics scores on the 2003 fourth and eighth grade NAEP and 2003 fourth and 2002 seventh grade WASL were moderately strong and positively correlated, with betas from .602 to .764. While the designs of the two assessments were quite similar however, a fundamental difference in the items selected for the tests gives insight into the strength of the relationships. The fourth grade NAEP

assessment includes items that range in difficulty from about a second grade level to about an eighth grade level; the eighth grade assessment includes items that range in difficulty from grade 4 to grade 12. The WASL items are aligned to the specific grade level standards and represent a difficulty level that is targeted as closely as possible to grade 4 or 7. Given this difference, it might be surprising to find stronger relationships between the tests. It does appear that the results of the two assessments will trend in the same direction over time, however. Thus, it is reasonable to infer that the NAEP can be used as one piece of confirmatory evidence for WASL progress.

#### Summary

This study of the 2003 NAEP and WASL investigated the relationships between the two assessments and the extent to which scores on the assessments were predictive of one another. The demographics of the group of students chosen by systematic sampling methods to participate in the NAEP were similar to Washington State fourth grade student demographics and their WASL performance. Demographic comparisons between seventh grade WASL takers and eighth grade NAEP indicated the group percentages were less similar. Results of regression analyses indicated that the relationships were moderately strong and positively correlated; suggesting that use of the NAEP results can be one piece of evidence used to confirm WASL achievement.

## Bibliography

- American Educational Research Association. (2000, July). *AERA position statement concerning high-stakes testing in preK-12 education*. Retrieved October 22, 2002, from <http://www.aera.net/about/policy/stakes.htm>
- Anderson, D., Ensign, G. MacQuarrie, D. (2002, December). Setting the WASL performance standards: Implications for the future. Paper presented at the annual fall meeting of the Washington Educational Research Association, Seattle, WA.
- Buckendahl, C. W., Smith, R. W., Impara, J. C., & Plake, B. S. (2000, October). A comparison of the Angoff and bookmark standard setting methods. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL.
- Council of Chief State School Officers. (2002, September). *Models for alignment analysis and assistance to states*. Washington, DC: Author.
- Feuer, M. J., Holland, P. W., Green, B. F., Bertenthal, M. W., & Cadelle-Hemphill, F. (Eds.). (1999). *Uncommon measures: Equivalence and linkage among educational tests*. National Academy Press, Washington, DC.
- Fraenkel, J. R. & Wallen, N. E. (1996). *How to design and evaluate research in education*. (3<sup>rd</sup> ed.). New York: McGraw-Hill.
- Heuschel, M. A. & Harmon, B. (2003, January). *Elementary and Secondary Education Act: Impact on Washington State*. Paper presented at the 10<sup>th</sup> Annual Office of Superintendent of Public Instruction January Conference, Spokane, WA.

- Kenney, P. A., & Silver, E. A. (1999). *Design features for the content analysis of a state assessment and NAEP*. Paper presented at the annual meeting of the American Educational Research Association in Montreal, Quebec, Canada. (ERIC Document Reproduction Service No. ED446150)
- Linn, R. L. (1994, November). *Assessment-based reform: Challenges to educational measurement*. Paper presented at the annual William H. Angoff Memorial Lecture, Princeton, NJ.
- MacQuarrie, D. (2003). History of the Washington state assessment program. Unpublished manuscript.
- McGhee, D. E. (2002). *The relationship between WASL scores and performance in the first year of college*. Seattle, Washington. University of Washington: Office of Educational Assessment.
- McLaughlin, D. & Bandeira De Mello, V. (2002). *Comparison of National Assessment of Educational Progress (NAEP) and statewide assessment results: Report to Washington on the 1998 reading assessments*. Palo Alto, CA: American Institutes for Research.
- NAEP Validity Studies Panel. (2002, October). *An agenda for NAEP validity research*. Palo Alto, CA: American Institutes for Research.
- National Assessment Governing Board. (n.d.). *How does NAEP work?* Retrieved December 30, 2001, from <http://www.nagb.org.html>
- National Assessment Governing Board. (2002a, September). *Mathematics framework for the 2003 National Assessment of Educational Progress*. Washington, DC: Author.

- National Assessment Governing Board. (2002b, September). *Reading framework for the 2003 National Assessment of Educational Progress*. Washington, DC: Author.
- National Assessment Governing Board. (n.d.). *What is NAEP?* Retrieved December 30, 2001, from <http://www.nagb.org.html>
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- No Child Left Behind Act of 2001, Pub. L. No. 107 – 110, § 602, 411 (2002). Retrieved December 30, 2002 from <http://www.nagb.org/about/plaw.html>
- Office of Superintendent of Public Instruction. (2002a, November). *Essential academic learning requirements: Technical manual*. Olympia, WA: Author.
- Office of Superintendent of Public Instruction. (2002b) *Reauthorized Elementary and Secondary Education Act (ESEA): “No child left behind” (NCLB act)*. Retrieved December 11, 2002 from [www.k12.wa.us](http://www.k12.wa.us)
- Pavelchek, D., Stern, P., & Olson, D. (2002). *Relationship between the WASL and placement tests used by community and technical colleges*. Pullman, Washington: Washington State University, Social & Economic Sciences Research Center.
- Rust, K. (2002, September). *NAEP design: Sampling and data collection*. Paper presented at the orientation meeting of the NAEP State Coordinators. Washington, DC.
- Shephard, L. A. (1991, March) Interview on assessment issues with Lorrie Shephard. Research News and Comment, *Educational Researcher*, 20, 21-23, 27.
- Taylor, C. S. (2002a). *Washington assessment of student learning: Grade 4 2001 technical report*. Olympia, Washington: Riverside Publishing.

Taylor, C. S. (2002b). *Washington assessment of student learning: Grade 7 2001 technical report*. Olympia, Washington: Riverside Publishing.

Thompson-Klenk, L. (2002, November). *Washington assessment of student learning in reading/listening*. Paper presented at the Regional Fall Assessment Workshops, SeaTac, WA.