A Study of the Ongoing Alignment of the NWEA RIT Scale with the Washington Assessment of Student Learning (WASL)

John Cronin, Ph.D

December 2004

Northwest Evaluation Association
5885 SW Meadows Road, Suite 200
Lake Oswego, OR 97035-3526

www.nwea.org
Tel 503-624-1951
Fax 503-639-7873

A Study of the Ongoing Alignment of the NWEA RIT Scale with the Washington Assessment of Student Learning (WASL)

John Cronin, Ph.D.

December, 2004

Each year, Washington students participate in testing as part of the state's assessment program. Students in grades 4, 7, and 10 take the Washington Assessment of Student Learning (WASL) in reading, mathematics, writing, and science. These tests serve as an important measure of student achievement for the state's accountability system. Results from these assessments are used to make state-level decisions concerning education, to meet *Adequate Yearly Progress* (AYP) reporting requirements of the *No Child Left Behind Act* (NCLB), and to inform schools and school districts of their performance. The Washington Office of the Superintendent of Public Instruction has developed scales that are used to assign students to one of four performance levels on these tests.

Many students who attend school in Washington also take tests developed in cooperation with the Northwest Evaluation Association (NWEA). The content of these tests are aligned with the Washington standards and they report student performance on a single, cross-grade scale, which NWEA calls the RIT scale. This scale was developed using Rasch scaling methodologies. RIT-based tests are used to inform a variety of educational decisions at the district, school, and classroom level. They are also used to monitor the academic growth of students and cohorts. Districts choose whether to include these assessments in their local assessment programs. They are not state mandated.

In order to use the two testing systems to support each other, an alignment of the scores from the state and RIT-based tests is as important as curriculum alignment. A three year study between 1998 and 2000 (using 1997 to 1999 data) first established estimated RIT scores that aligned with the equivalent cut points on the WASL scale (Hauser, 2000; Hauser, 1998). Because changes in the WASL performance levels were implemented last spring, we undertook a study to determine how those changes affected our estimated cut scores from the prior study. We also re-estimated the relative accuracy with which the NWEA assessments continued to predict WASL results. The primary questions addressed in this study are:

- What RIT scores correspond to various performance levels on the WASL tests?
- How do these RIT scores differ from the 1999 estimates of performance levels?
- How well can performance on the Washington assessments be predicted from RIT scores when NWEA assessments are administered in the same time frame?

Method

Our study included over 12,700 test records from students enrolled in 12 Washington school systems. Student records were included when a student had both a valid NWEA scale score and a valid WASL score in the equivalent subject.

The methodology used to complete this validation study was identical to that used in almost all of the state studies that we have completed in recent years (see Kingsbury et al, 2003). To conserve space, we refer readers to this study, "The State of State Standards", which is available on our website, for more detail about the methods we use to conduct scale alignment studies.

Results

Descriptive Statistics

Table 1 reviews descriptive statistics for the WASL and NWEA assessments. The median RIT scores for this sample in reading and mathematics are slightly above the median for the NWEA norm population with the exception of grade 10 mathematics, in which the median score of the sample population 234 was about 17 points below the national median for the grade.

Normal distributions around a nationally-normed mean are desirable but not necessarily essential when conducting alignment studies. It is more important that the sample provide reasonable numbers of students who perform at all levels on the test scales than normal distribution so that the statistical methods applied have an adequately large sample to derive good estimates of performance levels that are at the higher and lower ends of a test scale. In this case we had excellent representation of students who performed at all performance levels. This was even true in grade 10 mathematics, despite the relatively low performance of the group.

Table 1 – Means, Standard Deviations, and Medians for WASL and NWEA assessments

WASL Reading			
Grade	4	7	10
N	5633	6355	1331
Mean	409.98	404.16	404.57
Median	411	407	407
Std. Dev.	22.12	34.96	33.48
NWEA Reading			
Grade	4	7	10
N	5633	6355	1331
Mean	206.63	220.69	220.47
Median	208	222	223
Std. Dev.	13.87	14.23	17.52
WASL Mathematics			
Grade	4	7	10
N	5477	6135	1157
Mean	405.21	392.95	382.34
Median	408	394	382
Std. Dev.	39.50	43.78	47.38
NWEA Mathematics			
Grade	4	7	10
N	5477	6135	1157
Mean	212.80	230.97	232.21
Median	213	232	234
Std. Dev.	15.19	18.50	20.18

Pearson correlations

Table 2 shows the results of this analysis for each grade. Concurrent validity was tested by examining same subject Pearson correlations between the NWEA and WASL assessments. Same subject correlations were high, although not as high as they have been for most of our other state studies. The coefficients ranged between .77 to .87, numbers that suggest the tests were generally measuring the same constructs. Discriminant validity was tested by examining same subject Pearson correlations next to

correlations for the alternate subject (math against reading). The same subject correlations were higher than correlations against the alternate subject in all subjects and grades with the exception of grade 10 mathematics, which showed higher correlations between the WASL reading and WASL mathematics scores than between the two sets of mathematics scores. It was interesting to note that correlations between the WASL reading and mathematics assessments were considerably higher than those between the NWEA reading and mathematics assessments. While the reasons behind this are not entirely clear, the higher level of correlation between the two WASL tests may be an indicator that the WASL mathematics test may have a higher reading demand than is required by the NWEA mathematics assessments. This may help explain why the overall correlations, especially in mathematics, are lower than we have seen in most of our other studies.

Table 2 – Inter-test Correlations for WASL and NWEA assessments by Subject

Grade 4				
	WASL		NWEA	
	Reading	Mathematics	Reading	Mathematics
WASL Reading	1	.74	**.77**	.69
WASL Mathematics	.76	1	.71	**.80**
NWEA Reading	**.77**	.71	1	.76
NWEA Mathematics	.68	**.78**	.77	1
Grade 7				
	WASL		NWEA	
	Reading	Mathematics	Reading	Mathematics
WASL Reading	1	.78	**.77**	.73
WASL Mathematics	.79	1	.77	**.88**
NWEA Reading	**.78**	.78	1	.80
NWEA Mathematics	.64	**.77**	.70	1
Grade 10				
	WASL		NWEA	
	Reading	Mathematics	Reading	Mathematics
WASL Reading	1	.80	**.76**	.67
WASL Mathematics		1	.72	**.78**
NWEA Reading			1	.68
NWEA Mathematics				1

* Shaded cells show Pearson correlations for the reading analysis data set . Unshaded cells show correlations for the mathematics analysis data set. Same subject correlations are shown in boldface.

In general, relationships between NWEA and WASL reading scores tended to be curvilinear while math scores exhibited strong linear relationships. Figures 1 and 2 show the contrast between grade 7 reading and mathematics as examples. Figure 1 shows evidence of a floor effect, meaning that the NWEA assessment seems to have more capacity to measure low performance than the WASL assessment in this subject. This may occur because both the paper and computer-adaptive NWEA assessments are designed to adjust the difficulty of items to reflect the performance of the student taking the test. Because state examinations are typically designed to generate estimates of performance using the grade level standards and content, they may be limited in their ability to deliver items that accurately measure students in the lower ends of the performance range. In the case of reading, very few grade 7 students showed performance on the WASL below scale score 350, while these same students achieved RIT scores that ranged between 150 and 210 on the RIT scale. The same effect is not evident in grade 7 mathematics, with scores closely tracking through all ranges of both measurement scales.

Figure 1 – Grade 7 Reading WASL score plotted against Reading RIT score

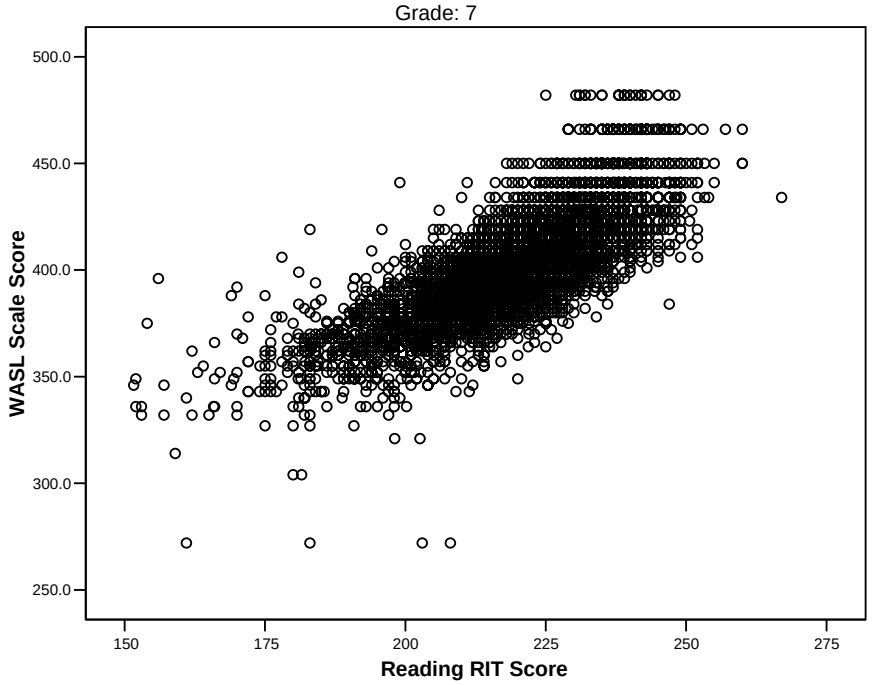
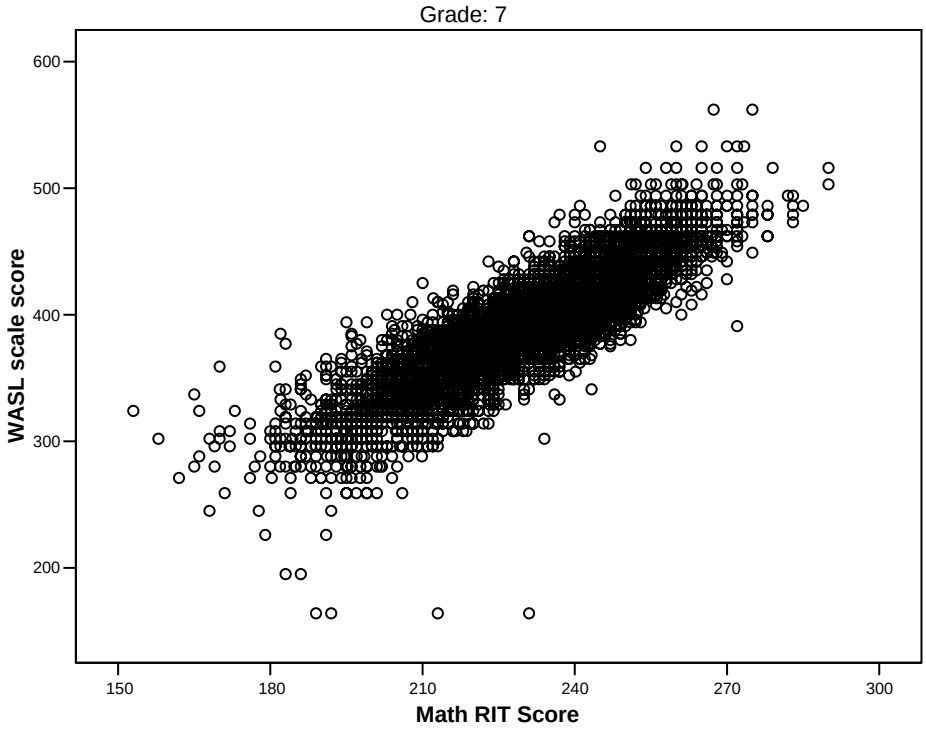


Figure 2 – Grade 7 Math WASL score plotted against Math RIT



score

Linking WASL performance level cut scores to the RIT scale

The primary purpose of this study was to generate new estimates of the RIT scale scores that most closely correspond to the cut scores for different performance levels on the WASL. This information allows schools to identify students who may need additional support to reach state standards. It can also help schools identify students who are performing well enough that they are ready to tackle work beyond what the state standards require.

Table 3 shows several estimations of the Spring 2003 RIT score that correspond to the cut scores for the various performance levels on the WASL scales. As a rule the three methodologies came to similar estimates of cut scores for each of the performance levels, although the Rasch SOS methodology did produce somewhat higher estimates of the RIT score required to meet the basic standard at some grades.

Table 3 – Estimated points on the RIT scale equating to the minimum scores (rounded) for performance levels on the WASL

	Grade 4											
	Linear Regression				Second-order Regression				Rasch Status-on-Standard			
	BB	B	P	A	BB	B	P	A	BB	B	P	A
Reading	<178	178	198	218	<177	177	199	218	<182	182	199	216
Mathematics	<197	197	210	224	<198	198	210	224	<199	199	210	221
	Grade 7											
	Linear Regression				Second-order Regression				Rasch Status-on-Standard			
	BB	B	P	A	BB	B	P	A	BB	B	P	A
Reading	<198	198	218	232	<198	198	218	231	<201	201	219	229
Mathematics	<222	222	234	250	<223	223	235	250	<222	222	234	248
	Grade 10											
	Linear Regression				Second-order Regression				Rasch Status-on-Standard			
	BB	B	P	A	BB	B	P	A	BB	B	P	A
Reading	<201	201	218	227	<202	202	220	228	<205	205	219	225
Mathematics	<229	229	242	256	<230	230	242	253	<221	221	248	251

Establishing RIT score estimates for WASL performance levels.

Once the cut scores were estimated from the three methods, we evaluated each set of possible cut scores to determine how accurately it predicted students' actual performance on the corresponding WASL assessment. The most accurate method of prediction was generally used to derive the best estimate of RIT cut scores that equate to the different WASL performance levels.

The following methods were used to establish the most accurate method for each performance level:

- **Below Basic and Basic.** We selected the method that correctly identified the largest portion of students who scored in the *below basic* category on WASL.

- **Proficient.** We calculated a *prediction index* statistic for the proposed cut score. This is calculated as 1 – (correct predictions/type I errors). A test with a high prediction index statistic typically reflects both a high rate of accuracy and a low rate of Type I errors. We generally selected the method that produced the highest prediction index number.

- **Advanced.** We selected the method that correctly identified the largest proportion of students who scored in the *advanced* category on the WASL.

Tables 4 and 5 show the recommended RIT cut scores for each of the WASL performance levels. In general, Rasch SOS methods were most reliable for establishing predictive cut scores for the highest and lowest performance levels, while all methods were similarly effective for predicting performance at the proficient level.

Table 4 – Recommended RIT cut scores for WASL performance levels - Reading

	Below Basic			Basic	Proficient			Advanced		
Grade	Score	Method	% of students ID	Score	Score	Method	Prediction Index*	Score	Method	% of students ID
3	<170			170	186			207		
4	<182	R	53.59%	182	199	S, R	.906 (86%)	216	R	63.03%
5	<189			189	206			221		
6	<195			195	213			225		
7	<201	R	59.52%	201	219	R	.868 (81%)	229	R	69.36%
8	<203			203	220			229		
9	<204			204	220			229		
10	<205	R	63.69%	205	220	S	.881 (81%)	225	R	79.36%

(L= Linear Regression, S=Second Order Regression, R=Rasch SOS method)

* percent correctly predicted is in parentheses

Table 5 – Recommended RIT cut scores for WASL performance levels – Mathematics

	Below Basic			Basic	Proficient			Advanced		
Grade	Score	Method	% of students ID	Score	Score	Method	Prediction Index*	Score	Method	% of students ID
3	<186			186	199			212		
4	<199	R	64.61%	199	210	L,S,R	.891 (83%)	221	R	70.40%
5	<208			208	219			231		
6	<217			217	226			240		
7	<223	S	79.60%	223	235	S	.926 (86%)	248	R	73.53%
8	<227			227	238			249		
9	<229			229	240			250		
10	<231	L	77.27%	231	242	L,S	.931 (86%)	251	R	76.67%

We evaluate the relative accuracy of state alignment studies by comparing the prediction index statistics generated by these studies for accuracy in assessing proficiency status and performance level. Table 6 summarizes the accuracy of proficiency status prediction for this study relative to other state alignment studies and Table 7 summarizes the accuracy of performance level prediction. The results show that the prediction index statistics for proficiency status prediction are low when compared to other state studies and slightly lower than those generated by the original Washington study. Nevertheless, the Washington index statistics showed rates of correct prediction for proficiency that were above 80% and ratios of correct prediction to Type I error that ranged from about 5 to 1 to nearly 12 to 1.

Table 6 – Prediction Indices (Based on Proficiency Status) for Previous NWEA State Alignment Studies

State	Reading	State	Language	State	Math
Texas	.967*	Texas	.968*	Texas	.969*
Minnesota	.944*	South Carolina Exit	.938*	Wyoming	.961
South Carolina Exit	.940*	California	.913*	Colorado '01	.957
Pennsylvania	.935*	Indiana '01	.907*	Illinois	.946*
Wyoming	.931	Colorado '03	.903*	Colorado '03	.943*
Colorado '03	.931*	Indiana '03	.894*	South Carolina '03	.943*
Illinois	.928*	South Carolina '04	.889*	Minnesota	.936*
California	.925*	Arizona	.874*	South Carolina Exit	.933*
Arizona	.912*			Pennsylvania	.926*
Colorado '01	.910*			Washington '99	.920
Nevada	.902*			Arizona	.919*
South Carolina '03	.902*			South Carolina '04	.914*
Indiana '01	.902*			Washington '04	.912*
Indiana '03	.900*			California	.910*
Washington '99	.893			Indiana '01	.899*
Washington '04	.886*			Nevada	.866*
South Carolina '04	.884*			Indiana '03	.860*

Table 7 – Prediction index scores by performance level assignment for previous NWEA state alignment Studies

State	Reading	State	Math
Texas	.868	Texas	.900
Indiana	.860*	Illinois	.888*
Colorado	.840	Colorado	.808
Illinois	.804*	Indiana	.804*
Nevada	.776*	Pennsylvania	.769*
Pennsylvania	.770*	South Carolina '03	.764*
South Carolina '03	.757*	Arizona	.726*
Arizona	.756*	Nevada	.742*
South Carolina '04	.717*	South Carolina '04	.741*
Washington '04	**.667**	**Washington '04**	**.721**
South Carolina Exit	.649*	South Carolina Exit	.705*
Minnesota	.627*	Minnesota	.611*
California	.600*	California	.565*

Using RIT scores to estimate student probability of achieving passing performance on the WASL

Although the predicted RIT cut scores can help teachers and students establish targets for NWEA assessments that can help assure success on the state test, teachers should be aware that students performing near the proficient cut score on the RIT scale have only about a 50% probability of passing the WASL. The information in Tables 8 and 9 provide educators with more precise data related to students' probabilities of achieving proficiency.

These tables show the proportion of students at each 5 point RIT level who earned scores at or above the *proficient* level on their respective WASL assessment. Using reading as an example, we find that about 30% of the Grade 4 students who achieved a reading RIT score between 190 and 194 went on to achieve a passing score on the WASL assessment. A reading teacher would know that only about one in three of these students is likely to achieve a proficient score on the WASL unless they work harder, receive more focused instruction, or have access to additional resources.

On the other hand, about 95% of students who scored between RITs of 210 and 214 achieved proficiency on the Washington assessment. Teachers should feel free to focus their efforts with these students on content and skills that go beyond the minimum expectations for performance.

Figures 3 and 4 are graphic depictions of the data in the tables.

Table 8 – Proportion of students passing the WASL reading based on same spring RIT reading score

RIT	Grade 4	Grade 7	Grade 10
165	0.00%		
170	5.00%		
175	11.11%		
180	6.17%		
185	18.81%	0.00%	4.35%
190	29.92%	0.79%	4.55%
195	47.23%	2.86%	13.04%
200	67.15%	7.27%	16.95%
205	86.52%	12.68%	22.08%
210	95.47%	26.45%	28.83%
215	98.18%	46.84%	44.93%
220	99.12%	65.49%	67.25%
225	100.00%	83.45%	79.01%
230		93.29%	87.93%
235		96.07%	96.73%
240		99.64%	98.84%
245		99.29%	100.00%
250		100.00%	

Table 9 – Proportion of students passing the WASL mathematics based on same spring RIT mathematics score

RIT	Grade 4	Grade 7	Grade 10
170	0.00%		
175	5.00%		
180	5.62%		
185	5.96%		
190	5.20%		
195	10.14%	0.00%	
200	22.26%	0.48%	0.00%
205	37.40%	1.14%	2.38%
210	62.50%	1.99%	1.54%
215	80.57%	4.45%	2.74%
220	89.52%	10.09%	7.79%
225	98.41%	22.10%	3.81%
230	100.00%	42.71%	18.18%
235	100.00%	61.68%	24.06%
240	99.01%	80.03%	50.00%
245	100.00%	93.01%	69.41%
250		98.21%	84.71%
255		100.00%	96.20%

Figure 3 – Percent of Students Passing Mathematics WASL by RIT Performance Range

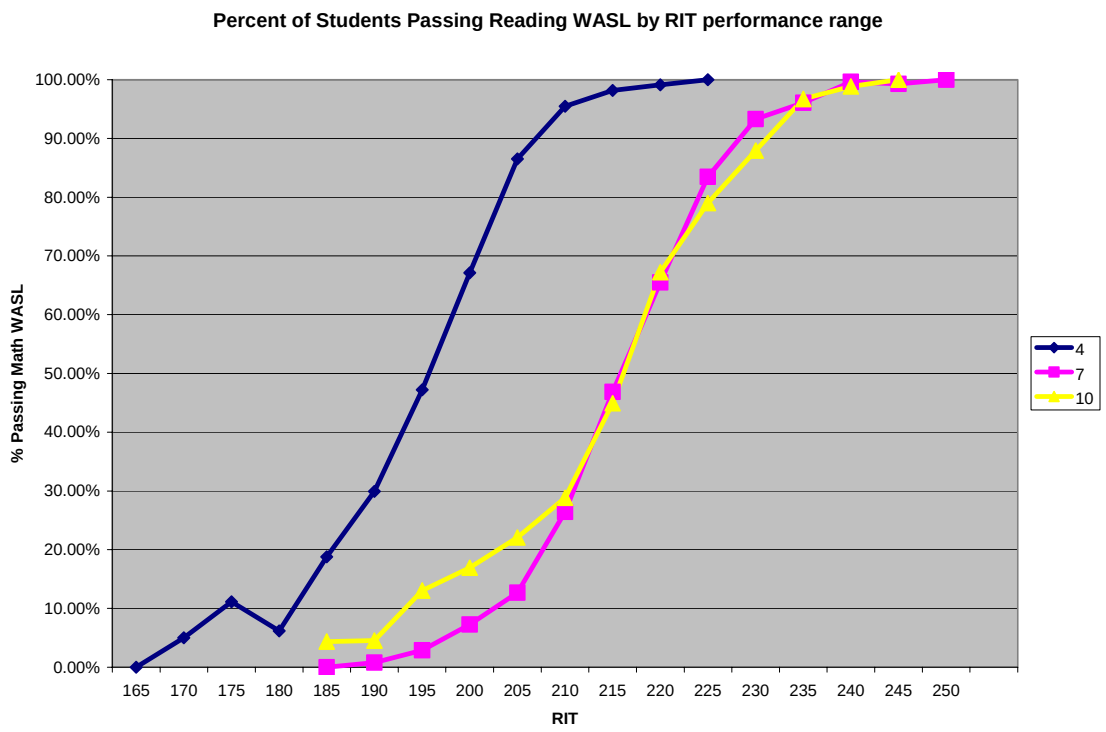
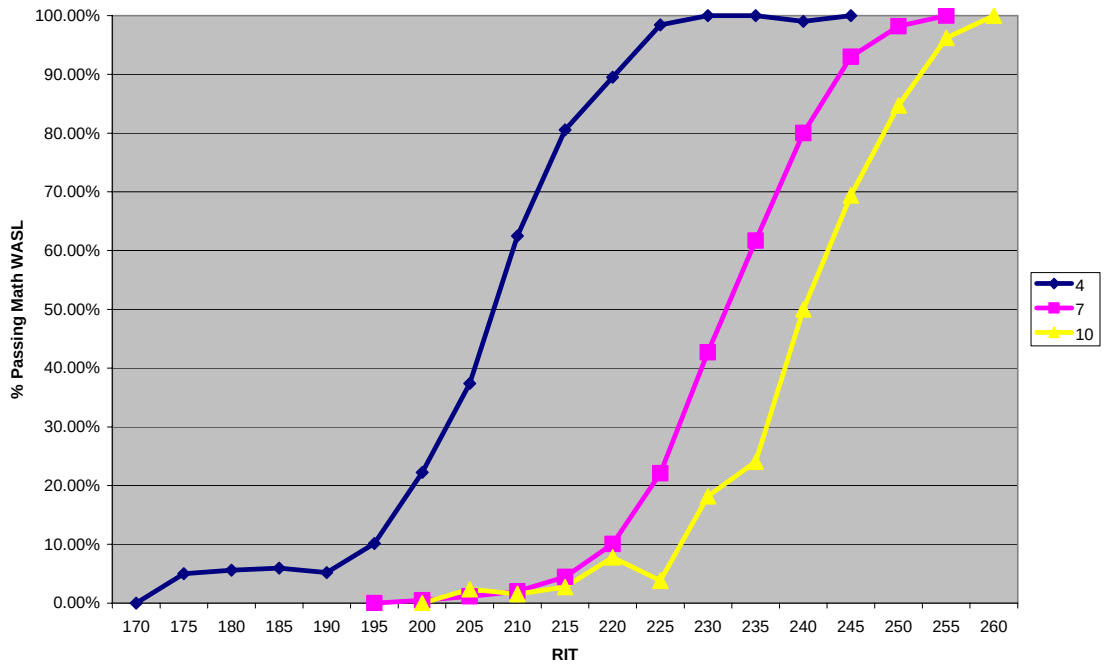


Figure 4 – Percent of Students Passing Mathematics WASL by RIT Performance Range



Comparing changes in the estimated WASL standards relative to the prior alignment study

Table 10 compares the cut scores found for the current study with those generated by our prior study. The Washington Office of the Superintendent of Public Instruction facilitated a process to re-evaluate the state's performance standards in 2004 and established new cut scores that went into effect during spring testing of that year. This was done, in part, because the lack of historical performance data precluded considering students' prior performance in the original standard workshop. This is an important consideration and the state now has an extensive history of student test performance that can be used to inform standard setting. This information is relevant, for example, to evaluating standards relative to the NCLB requirement that at all students reach proficiency by 2014. The original mathematics standard, which was set above the 70th percentile at all grades, could be argued to be beyond the level of achievement students need to enter some of the state's universities, and was certainly beyond reach for at least some of the state's students. Given the change in circumstances brought about by NCLB, rethinking the standard is not necessarily inappropriate

OSPI expected that standards would probably be lowered by the new process and the new projected cut scores on our scale are lower than the ones generated by the prior study. In all grades and both subjects the difference was between 7 and 9 RIT points lower than the prior estimate. It is apparent from examining the associated percentile scores, that these changes should substantively increase the number of students reaching proficiency on the state assessment.

Table 10 – Estimated RIT cut scores for the Proficient level of performance on the WASL 1999/2004*

	Reading		Mathematics	
	1999	2004	1999	2004
Grade 4	207 (53)	199 (32)	218 (76)	210 (54)
Grade 7	226 (67)	219 (46)	242 (78)	235 (65)
Grade 10	227 (53)	220 (35)	257 (75)	242 (25)

*NWEA percentile score (based on 2002 norms study) is in parentheses

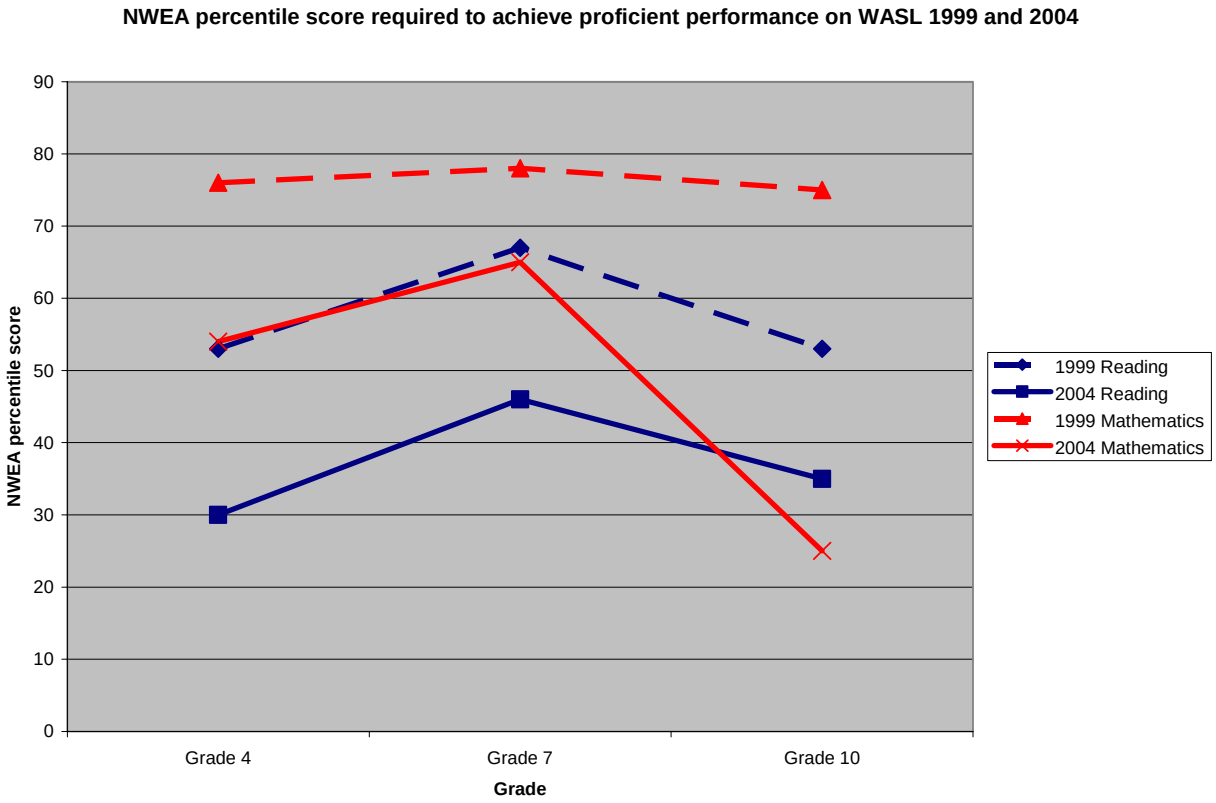
In our prior study, we noted that the Washington mathematics proficiency standard was higher, relative to our norm population, than the reading standard. That remains the case after the 2004 adjustment with the exception of grade 10 mathematics.

We also noted in the prior study that the grade 7 proficiency standard was not closely calibrated to expected performance at the other grades, particularly in reading. When we refer to a calibrated standard, we mean that the standard for grade 4 performance should be no easier or difficult to meet than the standard for grade 7 or grade 10 performance. To illustrate, we found that many grade 4 students who achieved at the proficient standard in reading (RIT = 207, 53rd percentile) could maintain that percentile standing and not achieve proficiency in grade 7 (RIT = 226, 67th percentile). This creates problems at grade 4 because students who may need additional support to reach the grade 7 standard will be identified as safe against the standard by the grade 4 test. It also creates issues at grade 7 because aggregate results leave the impression that grade 7 teachers achieve poorer results than grade 4 teachers when, in fact, the standard set is simply more difficult to achieve.

The issues around calibration seem to have been exacerbated by the 2004 adjustments in the proficiency standard. In mathematics, for example, the required relative level of performance increases substantively between grades 4 and 7 (from the 54th to the 65th percentile) and decreases dramatically between grades 7

and 10 (from the 65th percentile to the 25th percentile). The decrease in 10th grade expectations may be partially attributed to the fact that students will be required to pass this assessment to graduate. We know of at least two other states, California and South Carolina, in which the 10th grade standard is generally much lower than the proficiency standard set at other grades. In both, passing the 10th grade test is a prerequisite for graduation.

Figure 5 – NWEA percentile score required to achieve proficient performance on WASL 1999 - 2004



Finally, we found in the prior study that the grade 7 standard was more challenging than the standard set at other grades. That remains the case. Educators should be aware of the effect this is likely to have on passing rates at each grade. Table 11 is intended to help with that process. Table 11 shows the RIT score needed at each grade for students to meet the proficiency cut score for grade 7, which is the most challenging grade. This table allows educators to make true apples to apples comparisons in regard to how grades are doing relative to a calibrated standard. For example, if you want to compare the relative performance of 4th and 7th grade relative to the math standard, it is more effective to compare the number of 4th graders who perform at the 65th percentile (RIT = 205) than it is to use the estimated cut score for 4th grade proficiency (RIT = 198) because the 65th percentile comparison is roughly equivalent to the 7th grade expectations. This table is also more useful for planning interventions when you want to know which students in grades 3 through 6 are not likely to pass the grade 7 standard. A 4th grade student performing in reading at a RIT of 198 for example, has a 50/50 chance of passing the 4th grade state assessment, but this same student would need a RIT of about 205 to be on track to pass the 7th grade reading standard.

Table 11 – RIT score that calibrates (based on percentile) to proficient performance for grade 7 in reading and mathematics

	Reading		Mathematics	
	Estimated cut score for this grade	Cut score calibrated to the grade 7 standard (46th %ile)	Estimated cut score for this grade	Cut score calibrated to the grade 7 standard (65th %ile)
Grade 3	186 (22)	198	199 (46)	204
Grade 4	199 (32)	205	210 (54)	213
Grade 5	206 (35)	211	219 (56)	222
Grade 6	213 (40)	216	226 (58)	228
Grade 7	219 (46)	219	235 (65)	235
Grade 8	220 (37)	223	238 (55)	243
Grade 9	220 (35)	225	240 (42)	249
Grade 10	220 (35)	225	242 (25)	254

Comparing the WASL standards relative to those in place in other states

Northwest Evaluation Association tests have been aligned with the cut scores state assessments in 16 states. To get an estimate of the difficulty of the WASL in relation to other state tests, we evaluated the standard defined as the NCLB passing score and compared it to the cut score representing the same standard in these other states.

The results are summarized in tables 11 and 12. With the 2004 adjustment in cut scores, Washington's standards now typically fall in the lower-middle relative to the other states that we've studied.

In general, we believe standards should be judged on how well they align with the purposes the community has set for establishing performance expectations, not purely on how high or low the "bar" is set. If the purpose of a performance expectation is to assure that all students passing a standard will be ready to attend four year university, then the standard will need to be relatively high. On the other hand, if the purpose of a performance expectation is to assure that all students passing it graduate with the basic reading and math skills needed for entry level employment, the standard will be lower. It is clear from the evidence we've collected so far that proficiency is not yet a concept with a shared definition, because performance standards vary greatly from state to state. It would be fair to say, however, that most states that we have studied who have set standards since implementation of No Child Left Behind has begun have tended to establish standards near or below the 50th percentile on our norms.

Washington's prior standards were in place prior to the enactment of the No Child Left Behind legislation and, although passing the exit standard was a graduation requirement for all students, these standards were implemented without substantive data showing how students were likely to perform relative to them. If these standards are intended to represent a level of performance expected of all students, they seem more realistic than the prior standard and are not necessarily low relative to that kind of expectation. The possible exception is the 10th grade mathematics standard, which is so much lower that it does seem clearly disconnected from the performance standards at other grades.

Table 13 - Cut scores representing "proficient" or "meets standards" level of performance on 16 state assessments

Reading

Grade 3			Grade 4			Grade 5			Grade 6			Grade 7			Grade 8			Grade 9			Grade 10		
State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile
NV	202	58	WY	214	73	SC	218	68	SC	222	64	WA99	226	67	WY	232	74	MT	224	43	OR	236	77
CA	200	51	SC	209	59	NV	215	59	CA	216	46	SC	226	67	SC	230	68	IA	224	43	WA99	227	53
SC	196	42	WA99	207	53	CA	214	56	MT	211	35	CA	221	50	OR	227	58	ID	221	37	ID	224	44
MN	196	42	CA	205	46	PA	212	50	ID	211	35	WA04	219	46	CA	226	54	CO	204	9	MT	224	44
OR	193	35	ID	200	34	AZ	210	45	IN	210	32	MT	218	43	AZ	224	49				IA	223	44
ID	193	35	WA04	199	32	OR	209	42	IA	209	30	IA	216	37	PA	223	46				WA04	220	42
MT	193	35	MT	196	26	IL	207	37	TX	208	28	NV	215	35	IN	219	35				CO	209	35
IL	193	35	IA	196	26	MN	207	37	CO	197	11	ID	215	35	MT	219	35				SC	209	15
IN	192	32	NV	194	22	MT	206	35				TX	210	24	IA	219	35				CA	208	14
IA	191	31	CO	191	18	ID	206	35				CO	206	18	ID	218	32						
AZ	190	29				IA	205	32							IL	218	32						
TX	179	13				TX	204	30							MN	218	32						
CO	179	13				CO	197	18							CO	206	12						

In South Carolina and California the standard reflects the performance level required as a prerequisite to graduation.

Table 14 - Cut scores representing "proficient" or "meets standards" level of performance on 16 state assessments - Mathematics

Grade 3			Grade 4			Grade 5			Grade 6			Grade 7			Grade 8			Grade 9			Grade 10		
State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile
SC	212	84	WY	221	83	SC	230	81	SC	232	73	WA99	242	78	WY	257	89	MT	242	47	WA99	257	73
CA	204	63	SC	219	78	CA	225	71	CA	230	68	SC	241	76	AZ	248	75	IA	241	44	MT	247	40
NV	203	59	WA99	218	76	AZ	220	59	IN	221	47	CA	238	71	SC	247	73	ID	240	42	IA	247	40
IN	201	50	CA	212	59	NV	216	48	ID	219	42	WA04	235	65	CA	240	60	CO	235	32	OR	245	33
MN	200	49	WA04	210	54	PA	216	48	IA	218	40	ID	225	44	PA	237	53				ID	242	25
OR	199	46	ID	205	39	OR	215	46	MT	218	40	MT	224	42	OR	235	50				WA04	242	25
AZ	199	46	IA	205	39	ID	213	41	CO	207	19	IA	222	38	ID	233	46				CO	233	14
MT	197	39	MT	205	39	MT	212	38				TX	221	35	MN	231	42				CA	232	13
IA	197	39	NV	200	26	IA	212	38				NV	220	33	IN	231	42				SC	223	7
ID	196	36				MN	211	36				CO	216	26	IL	230	40						
IL	193	29				IL	210	33							MT	228	36						
						TX	209	31							IA	228	36						
						CO	201	15							CO	225	31						

NWEA
Northwest Evaluation Association
Partnering to help all kids learn

Summary and Conclusions

This study investigated the relationship between the scales used for the WASL assessments and the RIT scales used to report performance on Northwest Evaluation Association tests. The study estimated the changes in reading and mathematics RIT score equivalents for the WASL performance levels in those subjects. Test records for more than 12,000 students were included in this study.

Three methods generated an estimate of RIT cut scores that could be used to project WASL performance levels. Rasch SOS methods generally produced the most accurate cut score estimates. Accuracy of predicting WASL passing performance was well above 80% for all grades and subjects studied when using the best methodology.

Readers should exercise some caution about generalizing these results to their own settings. Curricular or instructional differences unique to your districts may influence the accuracy with which the estimated cut scores reflect actual performance in your setting. With this limitation in mind, we would encourage educators to use this data as one tool to inform standards-based decisions.

The information gathered in this study came from measures employing the NWEA RIT Scale. Because all of the research that we have to date indicates that scores generated from computer-based tests and Achievement Level Test (ALT) scores are virtually interchangeable, readers should feel comfortable applying the results of this study in any setting that uses the RIT scale.

We hope that data from this study provides useful information to help Washington educators use NWEA assessments to better inform, plan and deliver student instruction. Good information, when matched with the professionalism and commitment of our Washington colleagues, will assure that every student has the opportunity to reach their aspirations.

References

Kingsbury, G., Olson, A., Cronin, J., Hauser, C., Houser, R. (2003). The State of State Standards: Research Investigating Proficiency Levels in Fourteen States. Lake Oswego, OR: Northwest Evaluation Association.

Hauser, C. (2000). Linking the Scales From the *Washington Assessment of Student Learning* to the Northwest Evaluation Association RIT Scales. Lake Oswego, WA: Northwest Evaluation Association.

Hauser, C. Linking State and District Measurement Scales Using Item Response Theory: Toward Informing Local Policy and Standard Setting (1998, December). Paper presented at the 14th Annual Washington State Assessment Conference: SeaTac, WA.