A Study of the Ongoing Alignment of the NWEA RIT Scale with the North Dakota State Assessment (NDSA) Achievement Tests

John Cronin, Ph.D and Branin Bowe

November 2005

NWEA
Northwest Evaluation Association

Partnering to help all kids learn

NWEA

Northwest Evaluation Association

Partnering to help all kids learn

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 assessments from the North Dakota State Assessment (NDSA) Achievement Tests

John Cronin, Ph.D. and Branin Bowe

January, 2005

Each year, North Dakota students participate in testing as part of the state's assessment program. In fall of 2004, students in grades 3 through 8 and grade 11 took North Dakota State Assessment (NDSA) tests in reading and mathematics. 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 North Dakota Education Department has developed scales that are used to assign students to one of four performance levels on these tests.

Some students who attend school in North Dakota also take tests developed in cooperation with the Northwest Evaluation Association (NWEA). The content of these tests are aligned with the North Dakota 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. Thus we undertook a study to estimate scores on the RIT scale that would be equivalent to performance levels on the NDSA using three methods of estimation. We then compared the relative accuracy with which each methodology predicted results in order to derive these cut score estimates. The primary questions addressed in this study were:

- What RIT scores correspond to various performance levels on the NDSA tests?

- How well can performance on the North Dakota assessments be predicted from RIT scores when NWEA assessments are administered in the same testing season and when NWEA assessments are administered during the prior spring?

Method

Participants

State assessments in North Dakota are administered each fall. NWEA student assessment records in reading and mathematics were collected for the fall 2004 term and for the prior spring. Three school systems, Bismarck, Fargo, and Minot supplied data for both terms. Bismarck and Fargo supplied data for the prior spring term.

Our study included 9127 fall test records and 4369 spring test records from students enrolled in North Dakota school system. Student records were included when a student had both a valid NWEA scale score

and a valid NDSA score in the equivalent subject for at least one season. Table 1 shows the number of records included for mathematics.   The number included for reading is approximately the same.

Table 1 – Study Participants in Mathematics

	Fall 2004 Count			Spring 2004 Count		Total Fall	Total Spring
Grade	Bismarck	Fargo	Minot	Bismarck	Fargo		
3	686	723	163			1572	
4	641	705	180	612	285	1526	897
5	634	676	148	481	251	1458	732
6	697	689	192	670	279	1578	949
7	643	669	152	610	274	1464	884
8	664	710	155	621	265	1529	886
Total	3965	4172	990	3015	1354	9127	4369

Data Preparation

For purposes of studying NWEA test alignment with the NDSA, 3rd through 8th grade student test records from fall 2004 and the prior spring (2004) assessments were matched with the 2004 NDSA assessment by matching the district assigned student ID numbers for testing with the name and ID assigned for the state assessment. Matched records were then screened to remove invalid score. Students who received accommodations on the state test were also removed, in order to assure that both sets of tests were administered under similar conditions.   Minot students were not tested in spring of 2004, thus no students from this district were included. Third grade students in Fargo and Bismarck were not tested in spring of 2nd grade, thus no prior spring data is available for grade 3.

This the largest pool of students that NWEA has included in a state alignment study to date. We had enough student records at each grade to adequately cover the breadth of the scale and perform a robust analysis near the proficiency point for each NWEA tested subject. The number of records available for fall NWEA testing in second grade was considerably smaller than spring, mainly because many school systems do not administer fall NWEA tests to second grade students.

Because local curricula may vary in its alignment with either NWEA or state assessments, we recommend that schools validate our estimates by cross-checking their own students' performance against our projected cut scores.

Analyses

Pearson correlations. The initial analyses focused on the relationships among the NWEA and North Dakota assessment scores at each grade to determine how closely the scores on the NWEA test correlated with same subject scores on the NDSA. Simple bivariate correlation coefficients were computed among these scores.

Linking NDSA scores to the RIT scales. Fall and prior spring scores on the RIT scale were linked separately to the scale for the matching subject of the NDSA. Three methods of estimating cut scores for NDSA levels were used. The most straightforward was simple linear regression ($NDSA_{pred} = a(RIT) + c$). Since we sometimes observe departures from a linear relationship on the lower and upper ends of state

test scales, a second order regression model was also used ($NDSA_{pred}=a(RIT^2) + b(RIT) + c$). For each of these methods, the RIT score was determined by substituting the appropriate NDSA score for $NDSA_{pred}$ and solving the equation for RIT.

A fixed-parameter Rasch model was also used to estimate RIT cut scores. In this method, the NDSA performance level was treated as a test item. The assumption is that the performance level 'item' should contain all the information about the difficulty of the test. Student abilities (RIT scores) were the 'fixed parameter' used to anchor the difficulty estimate of the 'status' item to the RIT scale. The resulting 'difficulty estimate' was taken as the RIT cut score for this method. This is referred to as the Rasch Status on Standard (or simply Rasch SOS) method.

Predicting NDSA performance levels from RIT scores. Fall and Spring RIT scores were first used to predict whether students were likely to achieve performance at or above the proficient performance level on the NDSA.   The predictions of NDSA performance were compared to observed performance in 2 X 2 contingency tables. A prediction index score was generated to measure the ratio of Type I error to accurate prediction of proficiency status. This score is expressed as

1-(Number of Type I errors/Number of correct predictions)

Higher prediction index numbers generally show more accurate prediction with lower levels of Type I error. Type I error occurs when NWEA assessments predict that a student will achieve above a passing level of performance when the student actually achieves a failing score. This index was generated for the linear, second order, and Rasch SOS methodologies. In general, the highest prediction index score was used to select the RIT cut score to be adapted as the official RIT score we would associate with achieving the passing standard on the corresponding NDSA assessment for the particular grade level and subject area. We do make exceptions to this rule when the estimated score produces high accuracy rates but inordinately large numbers of Type II errors. This condition indicates a greatly overestimated cut score, so we select a method that produces a more balanced Type I to Type II error ratio in these instances.

In addition, we evaluated the accuracy of predictions of NDSA levels based on observed RIT scores. The predictions of NDSA level performance were compared to observed performance in 4 X 4 contingency tables. Once again a prediction index score was generated to provide an estimate of accuracy.

Content Validity

The NWEA technical manual describes the processes used by our test designers to assure the content and complex thinking evaluated on NWEA assessments is aligned with the standards taught in North Dakota. We did not conduct additional comparisons of the content of NWEA and North Dakota tests as part of this study.   Nevertheless, the standards used to construct the NWEA Assessments were the same as those used for the North Dakota assessments.   Both NWEA assessments and the NDSA include multiple-choice items. The NDSA also includes some constructed response questions. Results from our previous studies indicate that the addition of items in alternate formats generally does not, by itself, materially affect the ability of the NWEA test to generate accurate predictions of performance levels.

Results

Descriptive Statistics

Table 2 reviews descriptive statistics for the NDSA and NWEA assessments. The median fall RIT scores for this sample in reading are slightly above the median for the NWEA norm population, with the fall scores ranging between 2 and 3 points above the norm. In mathematics, fall scores of the sample were near the median, ranging between about 2 points below and 2 points above the norm population.

Table 2 – Means, Standard Deviations, and Medians for NDSA and NWEA assessments

NDSA Reading						
Grade	3	4	5	6	7	8
N	1843	1738	1682	1791	1719	1749
Mean	631.78	651.82	666.04	676.54	687.80	688.12
Median	631	653	667	678	688	688
Std Dev	26.99	26.75	28.85	27.34	26.76	29.06
NWEA Reading – Fall 2004						
Grade	3	4	5	6	7	8
N	1843	1738	1682	1791	1719	1749
Mean	193.55	203.34	210.15	214.98	219.06	221.47
Median	195	204	211	215	220	222
Std Dev	12.11	10.28	10.03	10.57	10.48	11.28
NWEA Reading – Spring 2004						
Grade	3	4	5	6	7	8
N		897	732	949	884	886
Mean		201.82	208.89	214.50	219.24	220.52
Median		203	210	215	219	221
Std Dev		10.37	10.10	9.57	10.33	10.88
NDSA Mathematics						
Grade	3	4	5	6	7	8
N	1838	1740	1683	1788	1722	1751
Mean	609.89	637.23	653.54	674.56	685.86	700.87
Median	605	637	653	673	686	700
Std Dev	35.27	29.34	28.90	29.37	29.05	32.76
NWEA Mathematics – Fall 2004						
Grade	3	4	5	6	7	8
N	1838	1740	1683	1788	1722	1751
Mean	194.52	205.43	213.98	220.87	226.05	231.65
Median	195	205	214	222	227	232
Std Dev	10.22	9.75	10.08	11.36	11.85	13.20
NWEA Mathematics – Spring 2004						
Grade	3	4	5	6	7	8
N		897	736	944	899	913
Mean		203.45	212.51	221.06	225.91	229.90
Median		204	213	222	226	230
Std Dev		9.93	9.94	10.85	11.38	12.81

Pearson correlations

Tables 3 and 4 show the results of this analysis for each grade. Concurrent validity was tested by examining same subject Pearson correlations between the NWEA and NDSA assessments. Same subject correlations between the NWEA and NDSA tests were stronger in the later grades than in the early grades. They ranged between .69 (grade 3 fall) and .77 (grade 7 fall and spring) in reading and between .68 (grade 3 fall) and .81 (grade 8 fall) in mathematics. We also examined same subject correlations between the two NWEA administrations. These ranged between .74 (grade 5) and .77 (grades 6 and 7) in reading and between .79 (grade 4) and .84 (grade 7) in mathematics.

These coefficients are low relative to other state studies we have conducted. For example, our most recently completed study was conducted in Arizona. Correlations between fall and spring administrations of the NWEA assessment ranged between .84 and .86 in reading and .83 and .89 in mathematics across the grades tested. Correlations between the NWEA assessment and the Arizona state assessment (AIMS) were also considerably higher.

We are not certain what caused these differences from other studies. Because correlations among all the assessments, both those between the two NWEA assessments and the correlation with NDSA, were lower than usual, it is possible that differences in conditions related to testing might have been a stronger influence than differences in the design of the NWEA and NDSA assessments themselves. One factor supporting this hypothesis may be that the study participants had been using NWEA assessments for a shorter time than is normal for participants in our state studies. All three were in their second year of administration. Our past history has been that members improve in their consistency in test administration and in maintaining good testing conditions as they gain greater experience with the testing process.

Table 3 – Reading Inter-test Correlations for NDSA and NWEA assessments

Reading			
Grade 3			
	NDSA Reading	NWEA Fall 04	
NDSA Reading	1	.70	
NWEA Fall 04	.70	1	
Grade 4			
	NDSA Reading	NWEA Fall 04	NWEA Spring 04
NDSA Reading	1	.69	.68
NWEA Fall 04	.69	1	.76
NWEA Spring 04	.68	.76	1
Grade 5			
	NDSA Reading	NWEA Fall 04	NWEA Spring 04
NDSA Reading	1	.73	.71
NWEA Fall 04	.73	1	.74
NWEA Spring 04	.71	.74	1
Grade 6			
	NDSA Reading	NWEA Fall 04	NWEA Spring 04
NDSA Reading	1	.72	.73
NWEA Fall 04	.72	1	.77
NWEA Spring 04	.73	.77	1
Grade 7			
	NDSA Reading	NWEA Fall 04	NWEA Spring 04
NDSA Reading	1	.77	.77
NWEA Fall 04	.77	1	.77
NWEA Spring 04	.77	.77	1
Grade 8			
	NDSA Reading	NWEA Fall 04	NWEA Spring 04
NDSA Reading	1	.75	.72
NWEA Fall 04	.75	1	.75
NWEA Spring 04	.72	.75	1

Table 4 – Mathematics Inter-test Correlations for NDSA and NWEA assessments

Mathematics			
Grade 3			
	NDSA Mathematics	NWEA Fall 04	
NDSA Mathematics	1	.68	
NWEA Fall 04	.68	1	
Grade 4			
	NDSA Mathematics	NWEA Fall 04	NWEA Spring 04
NDSA Mathematics	1	.69	.69
NWEA Fall 04	.69	1	.79
NWEA Spring 04	.69	.79	1
Grade 5			
	NDSA Mathematics	NWEA Fall 04	NWEA Spring 04
NDSA Mathematics	1	.73	.73
NWEA Fall 04	.73	1	.80
NWEA Spring 04	.73	.80	1
Grade 6			
	NDSA Mathematics	NWEA Fall 04	NWEA Spring 04
NDSA Mathematics	1	.76	.73
NWEA Fall 04	.76	1	.83
NWEA Spring 04	.73	.83	1
Grade 7			
	NDSA Mathematics	NWEA Fall 04	NWEA Spring 04
NDSA Mathematics	1	.78	.78
NWEA Fall 04	.78	1	.84
NWEA Spring 04	.78	.84	1
Grade 8			
	NDSA Mathematics	NWEA Fall 04	NWEA Spring 04
NDSA Mathematics	1	.81	.78
NWEA Fall 04	.81	1	.83
NWEA Spring 04	.78	.83	1

A review of scatterplots showed that outliers may also have exerted some influence on the correlation coefficients at some grades. In order to preserve the integrity of our validity analysis, we do not remove outliers when calculating correlation coefficients nor do we remove them when assessing predictive accuracy. We do remove outliers from linear regression to assure that the actual predictive cut scores are not skewed by atypical performances. Figure 1 shows an example from grade 5 reading in which an extraordinarily high NDSA outlier score depresses a correlation coefficient. Figure 2 shows an example from grade 6 reading in which three extremely low NDSA scores depress the correlation between the NDSA and NWEA assessments at that grade. In this instance, the low NDSA score suggests that the three students depicted may have simply given up or guessed their way through the state assessment.

In general the relations between NWEA and NDSA tests were best described as slightly curvilinear, with some visible evidence of floor effect at the low end of the scale. Figures 1 and 2 provide good illustrations of both. They show that scores that mild floor effect. Figure 2 shows for example, that 6th grade students who scored near 600 on the state test achieved NWEA scores that ranged anywhere between 160 RIT and 210 RIT. Normally this condition exists when one test is able to measure the low end of the performance continuum more accurately.

This may be attributable to the design of the state test. Because the state test is written to focus on the standards for the grade (as NCLB requires), low performing students may not have been offered many items that they could answer correctly. If most items on a test are too hard for low performing students, the test simply becomes an exercise in guessing and will produce a score that does not accurately represent what content may have been learned. NWEA assessments are designed to align with the North Dakota state standards, but their adaptive nature ensures they offer low performing students items that accurately represent both what has been learned and what hasn't. This design assures more accurate results that are reflected in a standard error of measure that stays relatively constant across the entire scale, while state test designs generally produce higher standard errors of measure near the scale's extremes.

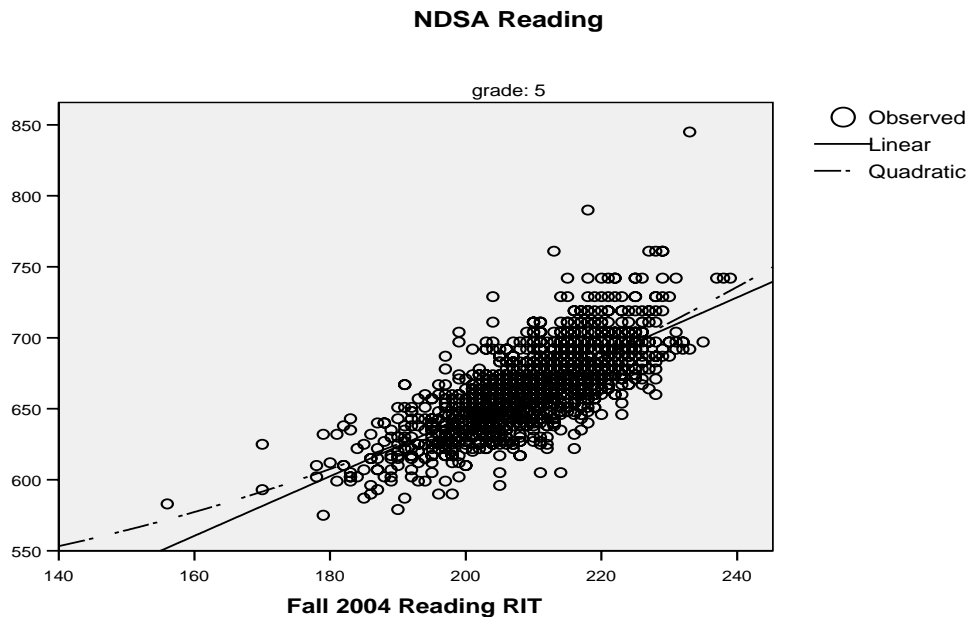## Figure 1 – Grade 5 Reading NDSA score plotted against Reading RIT score

NDSA Reading

grade: 5

Figure 2 – Grade 6 Reading NDSA score plotted against Reading RIT score

NDSA Reading
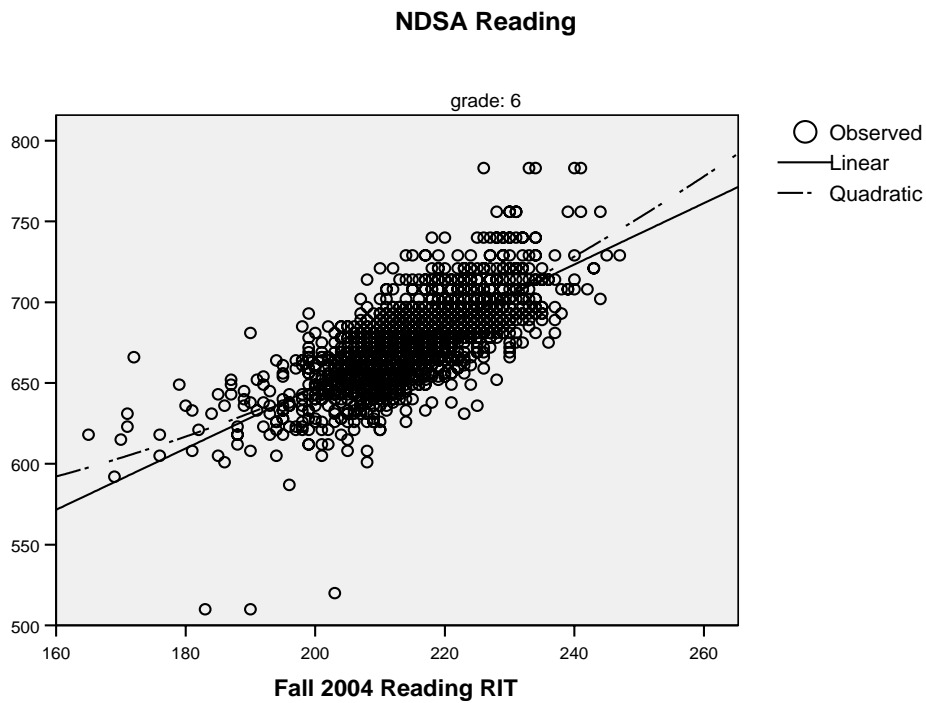
Linking NDSA performance level cut scores to the RIT scale

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 NDSA. 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 5 shows several estimations of the fall and prior spring RIT scores that correspond to the cut scores for the various performance levels on the NDSA scales. The estimates were generally quite close, with no set of estimates for a single grade differing by more than 3 RIT points.

Table 5 – Estimated points on the RIT scale equating to the minimum scores (rounded) for performance levels on the NDSA

	Reading Fall								
	Linear Regression			Second-order Regression			Rasch Status on Standard		
	Partially Proficient	Proficient	Advanced	Partially Proficient	Proficient	Advanced	Partially Proficient	Proficient	Advanced
Grade 3	161	181	206	147	181	205	150	181	208
Grade 4	179	192	213	175	192	213	176	190	216
Grade 5	188	200	222	187	200	221	185	200	225
Grade 6	191	204	226	190	204	225	187	202	228
Grade 7	197	208	230	195	209	229	191	208	231
Grade 8	201	212	236	201	213	235	198	212	237
	Reading Prior Spring								
	Linear Regression			Second-order Regression			Rasch Status on Standard		
	Partially Proficient	Proficient	Advanced	Partially Proficient	Proficient	Advanced	Partially Proficient	Proficient	Advanced
Grade 3									
Grade 4	175	189	212	164	190	211	173	187	215
Grade 5	185	199	222	183	199	221	181	197	224
Grade 6	193	204	224	193	205	224	189	203	227
Grade 7	197	208	230	198	209	230	187	207	233
Grade 8	198	210	235	197	211	234	193	210	237
	Mathematics Fall								
	Linear Regression			Second-order Regression			Rasch Status on Standard		
	Partially Proficient	Proficient	Advanced	Partially Proficient	Proficient	Advanced	Partially Proficient	Proficient	Advanced
Grade 3	167	178	213	164	177	211	157	176	207
Grade 4	180	192	222	183	193	222	177	191	219
Grade 5	193	204	232	193	204	231	183	203	226
Grade 6	196	210	232	196	210	232	193	209	231
Grade 7	202	218	238	202	219	237	197	216	239
Grade 8	209	225	239	208	226	239	204	225	246
	Mathematics Prior Spring								
	Linear Regression			Second-order Regression			Rasch Status on Standard		
	Partially Proficient	Proficient	Advanced	Partially Proficient	Proficient	Advanced	Partially Proficient	Proficient	Advanced
Grade 3									
Grade 4	177	189	220	179	190	219	170	188	217
Grade 5	191	202	231	192	202	230	182	201	225
Grade 6	195	209	232	197	209	231	190	208	231
Grade 7	202	217	236	203	218	236	194	215	238
Grade 8	204	222	237	199	223	237	196	222	246

Establishing RIT score estimates for NDSA 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 NDSA assessment. The most accurate method of prediction was generally used to derive the best estimate of RIT cut scores that equate to the different NDSA performance levels.

For this study, we first assessed the accuracy of the RIT scale in correctly predicting whether students are likely to reach the *proficient* level on the corresponding NDSA test. Next we assessed the accuracy with which the RIT predicted proper performance level assignment on this test. Use of the prediction index statistic helped assure that the method chosen produced a high ratio of accurate passing predictions relative to Type I errors. Type I errors occur when the RIT scale predicts a *proficient* score for a student who actually does not pass the assessment. These types of errors raise particular concern because they fail to identify students who might need additional support and resources in order to achieve their targets. A high prediction index number indicates that the test maximizes accuracy of prediction while minimizing Type I errors.

In these kinds of studies we want to emphasize that prediction is not used to foretell an inevitable future for the student, rather it is used to help schools plan for instruction and offer appropriate interventions to children who need additional support to be successful. For purposes of the *No Child Left Behind Act,* schools are judged on their ability to move children to the *proficient* level and beyond. RIT scores can provide teachers with advance notice about students who may not reach these goals on the North Dakota assessment that corresponds to their grade level.

Tables 6 and 7 summarize the results. In reading, the accuracy of pass/fail prediction ranged between about 83% and 86%, depending on grade and season tested. In mathematics the accuracy of pass/fail prediction ranged between about 84% and 88%. This level of accuracy should be adequate to permit the NWEA assessments use as a tool to identify students who might be at risk relative to passing the state test.

Table 6 – Evaluation of Projected RIT cut scores for NDSA *proficient* level - Reading

	Fall				Prior Spring			
Grade 3	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	181	84.5%	11.7%	.861				
Second Order	181	84.5%	11.7%	.861				
Rasch	181	84.5%	11.7%	.861				
Grade 4	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	192	86.2%	9.9%	.885	189	86.1%	9.9%	.885
Second Order	192	86.2%	9.9%	.885	190	86.3%	9.5%	.890
Rasch	190	85.7%	11.7%	.863	187	85.2%	11.8%	.884
Grade 5	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	200	86.2%	11.0%	.873	199	85.9%	9.7%	.890
Second Order	200	86.2%	11.0%	.873	199	85.9%	10.0%	.884
Rasch	200	86.2%	11.0%	.873	197	85.9%	12.4%	.891
Grade 6	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	204	83.9%	13.0%	.845	204	86.3%	10.7%	.875
Second Order	204	83.9%	13.0%	.845	205	86.8%	9.5%	.891
Rasch	202	83.3%	14.6%	.824	203	86.3%	11.6%	.866
Grade 7	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	208	85.7%	11.1%	.871	208	85.0%	12.4%	.854
Second Order	209	85.9%	10.0%	.884	209	84.7%	11.8%	.861
Rasch	208	85.7%	11.1%	.871	207	85.0%	13.2%	.844
Grade 8	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	212	83.1%	12.9%	.845	210	82.8%	13.9%	.832
Second Order	213	83.1%	11.6%	.884	211	83.2%	13.0%	.844
Rasch	212	83.1%	12.9%	.845	210	82.8%	13.9%	.832

Method used to select the cut score for this grade is in bold

Table 7– Evaluation of Projected RIT cut scores for NDSA *proficient* level - Mathematics

	Fall				Prior Spring			
Grade 3	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	**178**	**88.4%**	**9.9%**	**.888**				
Second Order	177	88.4%	10.2%	.884				
Rasch	176	88.2%	10.8%	.878				
Grade 4	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	192	88.4%	9.9%	.889	189	88.2%	10.0%	.886
Second Order	**193**	**88.2%**	**9.3%**	**.894**	**190**	**88.2%**	**9.5%**	**.893**
Rasch	191	88.5%	10.4%	.883	188	87.8%	10.8%	.877
Grade 5	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	**204**	**86.2%**	**10.2%**	**.882**	**202**	**86.1%**	**11.4%**	**.868**
Second Order	**204**	**86.2%**	**10.2%**	**.882**	**202**	**86.1%**	**10.2%**	**.868**
Rasch	203	85.8%	11.7%	.864	201	86.7%	11.8%	.864
Grade 6	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	**210**	**84.9%**	**10.4%**	**.878**	**209**	**86.4%**	**10.2%**	**.882**
Second Order	**210**	**84.9%**	**10.4%**	**.878**	**209**	**86.4%**	**10.2%**	**.882**
Rasch	209	85.3%	11.3%	.868	208	86.4%	11.2%	.870
Grade 7	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	218	84.9%	9.5%	.888	217	86.5%	9.0%	.896
Second Order	**219**	**84.9%**	**8.2%**	**.903**	**218**	**85.9%**	**7.8%**	**.909**
Rasch	216	84.1%	12.5%	.852	215	86.4%	10.9%	.874
Grade 8	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	225	84.8%	10.4%	.878	222	83.4%	11.0%	.869
Second Order	**226**	**84.8%**	**8.7%**	**.898**	**223**	**84.1%**	**9.3%**	**.889**
Rasch	225	84.1%	10.3%	.878	222	83.4%	11.0%	.869

Method used to select the cut score for this grade is in bold

Next we selected cut scores to differentiate the *partially proficient* and *novice* level and to define the cut score for the *advanced* level. The following methods were used to establish these:

- **Partially Proficient/Novice.** We selected the method that correctly identified the largest proportion of students who scored at the *novice* level.

- **Advanced.** We selected the method that correctly identified the largest proportion of students who scored in the *advanced* category on the NDSA. Because the population distribution of this sample created a greater risk of errors of overprediction, we used the methodology that produced the lowest proportion of type I errors.

The results of this are summarized in Tables 8 and 9.

Table 8 – Evaluation of Projected RIT cut scores for NDSA *novice, partially proficient, and advanced* performance levels - Reading

Grade	Method	Fall Reading					Prior Spring Reading				
		Novice/Part Prof		Advanced		Prediction Index	Novice/Part Prof		Advanced		Prediction Index
		Cut Score	% Nov Found	Cut Score	% Adv Found		Cut Score	% Nov Found	Cut Score	% Adv Found	
3	Linear	**161**	**25.0%**	206	44.6%	.721					
	Second Order	147	20.0%	**205**	**52.1%**	.689					
	Rasch	150	0.0%	208	35.0%	.756					
4	Linear	**179**	**20.0%**	**213**	**51.9%**	.709	**175**	**10.4%**	212	49.2%	.723
	Second Order	175	13.8%	**213**	**51.9%**	.708	164	2.1%	**211**	**56.0%**	.696
	Rasch	176	13.8%	216	30.9%	.736	173	10.4%	215	33.0%	.753
5	Linear	**188**	**34.3%**	222	36.3%	.742	**185**	**8.1%**	222	28.3%	.735
	Second Order	187	28.6%	**221**	**43.9%**	.723	183	8.1%	**221**	**32.6%**	.734
	Rasch	185	20.0%	225	22.5%	.777	181	8.1%	224	21.0%	.734
6	Linear	**191**	**24.4%**	226	49.4%	.702	**193**	**27.3%**	**224**	**46.3%**	.696
	Second Order	190	21.8%	**225**	**54.9%**	.690	**193**	**27.3%**	**224**	**46.3%**	.716
	Rasch	187	16.7%	228	38.2%	.704	189	26.2%	227	26.2%	.727
7	Linear	**197**	**21.2%**	230	54.2%	.735	197	35.1%	**230**	**52.3%**	.720
	Second Order	195	18.2%	**229**	**58.3%**	.722	**198**	**35.1%**	**230**	**52.3%**	.729
	Rasch	191	13.6%	231	48.1%	.763	187	16.2%	233	34.1%	.732
8	Linear	**201**	**32.0%**	236	39.0%	.742	**198**	**24.6%**	235	32.7%	.696
	Second Order	**201**	**32.0%**	**235**	**42.5%**	.749	197	23.1%	**234**	**41.3%**	.695
	Rasch	198	24.6%	237	33.6%	.748	193	12.3%	237	26.0%	.702

Method used to select the cut score for this grade is in bold

Table 9 – Evaluation of Projected RIT cut scores for NDSA *novice, partially proficient, and advanced* **performance levels - Mathematics**

Grade	Method	Fall Mathematics					Prior Spring Mathematics				
		Novice/Part Prof		Advanced		Prediction Index	Novice/Part Prof		Advanced		Prediction Index
		Cut Score	% Nov Found	Cut Score	% Adv Found		Cut Score	% Nov Found	Cut Score	% Adv Found	
3	Linear	**167**	**3.0%**	213	12.4%	.835					
	Second Order	164	0.0%	211	16.9%	.826					
	Rasch	157	0.0%	**207**	**37.3%**	.794					
4	Linear	180	5.9%	222	17.5%	.798	177	5.9%	220	16.8%	.789
	Second Order	**183**	**13.7%**	222	17.5%	.807	**179**	**11.8%**	219	11.8%	.794
	Rasch	177	3.9%	**219**	**24.9%**	.773	170	0.0%	**217**	**29.7%**	.768
5	Linear	**193**	**27.0%**	232	14.5%	.801	191	28.6%	231	13.8%	.790
	Second Order	**193**	**27.0%**	231	16.0%	.799	**192**	**32.1%**	230	32.1%	.785
	Rasch	183	4.8%	**226**	**34.9%**	.721	182	0.0%	**225**	**37.1%**	.722
6	Linear	**196**	**28.0%**	232	48.4%	.721	195	18.4%	232	46.2%	.724
	Second Order	**196**	**28.0%**	232	48.4%	.721	**197**	**23.7%**	231	51.8%	.703
	Rasch	193	20.0%	**231**	**53.9%**	.685	190	7.9%	**231**	51.8%	.675
7	Linear	**202**	**27.6%**	238	57.9%	.751	202	26.2%	**236**	**64.6%**	.747
	Second Order	**202**	**27.6%**	237	63.3%	.759	**203**	**28.6%**	236	64.6%	.760
	Rasch	197	11.8%	**239**	**50.0%**	.717	194	7.1%	238	51.3%	.732
8	Linear	**209**	**47.2%**	239	86.3%	.565	**204**	**26.1%**	237	85.3%	.540
	Second Order	208	41.7%	**239**	**86.3%**	.588	199	13.0%	**237**	85.3%	.562
	Rasch	204	21.3%	246	58.5%	.715	196	13.0%	246	42.9%	.745

Method used to select the cut score for this grade is in bold

As a general rule, NWEA tests far more effectively predicted *advanced* status than *novice* standard and were more effective at predicted both in the upper grades than in grades 3 and 4. Part of this may simply be explained by the fact that the predicted cut scores for *novice/partially proficient* performance are near the low point on our scale in grades 3 and 4 and that few students in this sample performed near those scores. In grade 3 reading, for example, the most accurate of the three cut scores (RIT = 164) was a 4th percentile score. Only 17 of the 1572 students in the grade 3 reading sample achieved at or below this score. In grade 3 mathematics, the most accurate cut score (RIT = 167) reflects the 3rd percentile on NWEA norms and was achieved by only 5 of the 1572 students in this sample. More likely, it was difficult

to predict this level of performance with great accuracy because so few students in the sample actually performed at or near the novice level on their state test.

Table 10 aggregates the information in Tables 6 through 9 into a single table showing the recommended cut scores for each performance level and grade for both the fall and prior spring assessments.

Table 11 – Recommended RIT cut scores for NDSA performance levels

Reading								
Grade	Fall				Prior Spring			
	Novice	Part Prof	**Proficient**	Advanced	Novice	Part Prof	**Proficient**	Advanced
3	<161	161	**181**	205				
4	<179	179	**192**	213	<175	175	**190**	211
5	<188	188	**200**	221	<185	185	**197**	221
6	<191	191	**204**	225	<193	193	**205**	224
7	<197	197	**209**	229	<198	198	**209**	230
8	<201	201	**213**	235	<198	198	**211**	234
Mathematics								
Grade	Fall				Prior Spring			
	Novice	Part Prof	**Proficient**	Advanced	Novice	Part Prof	**Proficient**	Advanced
3	<167	167	**178**	207				
4	<183	183	**193**	219	<179	179	**190**	217
5	<193	193	**204**	226	<192	192	**202**	225
6	<196	196	**210**	231	<197	197	**209**	231
7	<202	202	**219**	239	<203	203	**218**	236
8	<209	209	**226**	239	<204	204	**223**	237

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 for the season in which both the state and NWEA test were administered (in this case fall). Table 12 summarizes the accuracy of proficiency status prediction for this study relative to other state alignment studies and Table 13 summarizes the accuracy of performance level prediction. The results show that the prediction index statistics for proficiency status in both reading and mathematics are low relative to other states studied. In relation to predictions of performance level, the results show that NDSA performed near the middle of the group relative to the other states studied.

Table 12 – Prediction Indices (Based on Proficiency Status) for Previous NWEA State Alignment Studies

State	Reading	State	Language	State	Math
Texas	.967	Texas	.968	Tennessee	.975
Tennessee	.958	South Carolina Exit	.938	Texas	.969
Minnesota	.944	California	.913	Wyoming	.961
South Carolina Exit	.940	Indiana '01	.907	Colorado '01	.957
Pennsylvania	.935	Colorado '03	.903	Illinois	.946
Wyoming	.931	Indiana '03	.894	Colorado '03	.943
Colorado '03	.931	South Carolina '04	.889	South Carolina '03	.943
Illinois	.928	Arizona	.874	Minnesota	.936
California	.925			South Carolina Exit	.933
Arizona '03	.912			Pennsylvania	.926
Colorado '01	.910			Washington '99	.920
Montana	.903			Arizona '03	.919
Nevada	.902			South Carolina '04	.914
South Carolina '03	.902			Washington '04	.912
Indiana '01	.902			California	.910
Indiana '03	.900			Arizona '05	.910
Washington '99	.893			Montana	.899
Arizona '05	.891			Indiana '01	.899
Washington '04	.886			North Dakota	.890
South Carolina '04	.884			Nevada	.866
North Dakota	.868			Indiana '03	.860

Table 13 – 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	Tennessee	.860
Illinois	.804	Colorado	.808
Arizona '05	.781	Indiana	.804
Nevada	.776	Pennsylvania	.769
Pennsylvania	.770	South Carolina '03	.764
South Carolina '03	.757	North Dakota	.751
Arizona '03	.756	Nevada	.742
North Dakota	.745	South Carolina '04	.741
South Carolina '04	.717	Arizona '05	.730
Montana	.670	Arizona '03	.726
Washington	.667	Washington	.721
South Carolina Exit	.649	Montana	.707
Minnesota	.627	South Carolina Exit	.705
California	.600	Minnesota	.611
Tennessee	.591	California	.565

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

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 NDSA. 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 NDSA assessment. Using reading as an example, we find that about 21% of the grade 5 students who achieved a reading RIT score between 190 and 194 went on to achieve a proficient score on the NDSA assessment. A reading teacher would know that only about one in three of these students is likely to achieve a proficient score on the NDSA unless they work harder, receive more focused instruction, or have access to additional resources.

On the other hand, about 90% of students who scored between RITs of 210 and 214 achieved proficiency on the North Dakota assessment at this grade. 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 14 – Proportion of students passing the NDSA reading based on same fall RIT reading score

| | Reading | | | | | |
	3	4	5	6	7	8
150	0.00%					
155	11.11%					
160	22.22%	0.00%				
165	21.43%	40.00%				
170	24.07%	20.00%				
175	33.72%	26.67%				
180	46.15%	21.95%				
185	69.02%	30.88%	0.00%	0.00%		
190	84.58%	48.03%	20.83%	10.53%		
195	92.52%	72.22%	30.10%	38.00%		0.00%
200	98.79%	89.00%	60.22%	38.83%	10.45%	10.42%
205	98.85%	94.33%	80.87%	56.90%	42.11%	25.56%
210	100.00%	98.79%	89.76%	81.31%	67.18%	51.72%
215		100.00%	98.13%	93.15%	86.23%	70.42%
220			100.00%	97.93%	93.26%	81.53%
225				98.80%	99.05%	95.15%
230				100.00%	100.00%	98.04%
235						98.11%
240						100.00%

Table 15 – Proportion of students passing the NDSA reading based on same prior spring RIT reading score

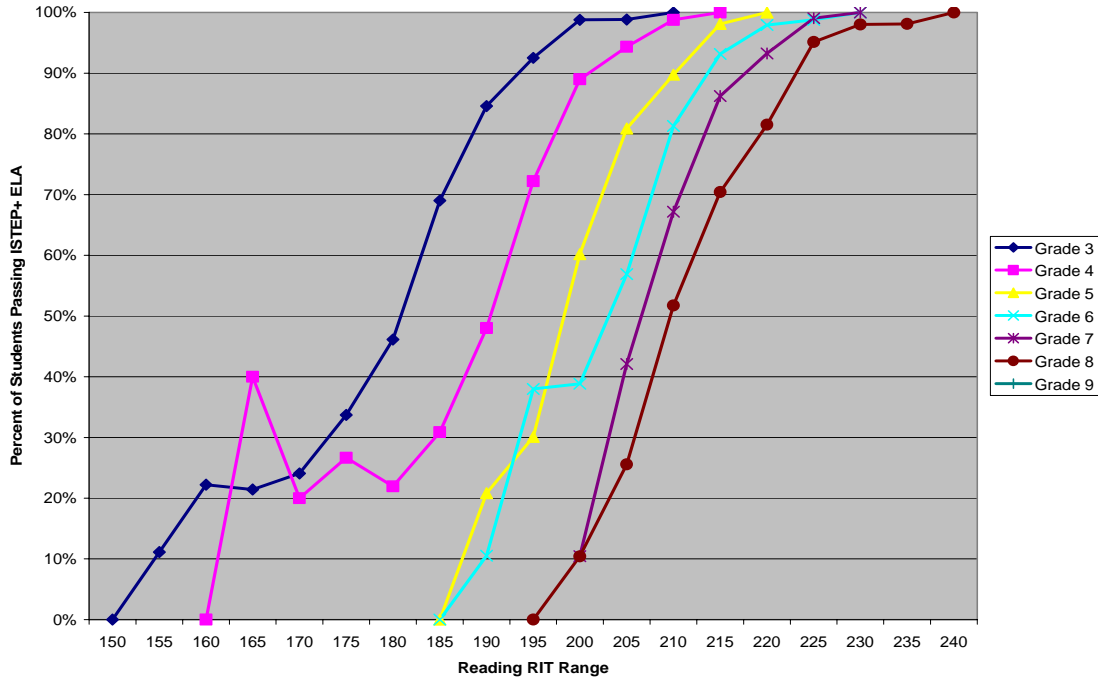| | Reading | | | | | |
	3	4	5	6	7	8
170						
175		30.00%				
180		33.33%				
185		35.85%	15.38%	0.00%	0.00%	
190		62.82%	17.24%	12.50%	12.50%	
195		82.27%	37.68%	16.67%	18.18%	0.00%
200		85.86%	66.67%	33.33%	22.86%	25.93%
205		98.28%	84.00%	64.00%	44.87%	32.81%
210		100.00%	93.63%	84.49%	62.68%	56.30%
215		98.61%	99.19%	95.94%	83.43%	73.83%
220		100.00%	98.08%	95.73%	95.93%	85.28%
225			100.00%	98.99%	98.40%	96.27%
230				100.00%	98.78%	97.98%
235					100.00%	100.00%

Table 16 – Proportion of students passing the NDSA mathematics based on same fall RIT mathematics score

| Grade/RIT Range	Mathematics					
	3	4	5	6	7	8
165						
170	31.03%					
175	43.06%					
180	59.42%	16.67%				
185	78.33%	16.39%	0.00%	0.00%		
190	94.62%	51.92%	4.55%	8.33%	0.00%	
195	96.75%	68.98%	17.81%	9.38%	9.09%	
200	98.71%	87.71%	39.33%	15.71%	8.57%	
205	100.00%	95.82%	67.89%	47.58%	16.92%	0.00%
210		98.38%	83.76%	56.74%	24.53%	3.23%
215		100.00%	96.39%	77.91%	49.71%	15.79%
220			96.12%	91.85%	72.12%	33.33%
225			100.00%	97.54%	86.82%	54.23%
230				98.97%	95.33%	80.58%
235				98.06%	99.42%	94.37%
240				100.00%	98.78%	98.83%
245					100.00%	100.00%
250						

Table 17 - Proportion of students passing the NDSA mathematics based on prior spring RIT mathematics score

| Grade/RIT Range	Mathematics					
	3	4	5	6	7	8
180		5.88%				
185		40.43%				
190		58.97%	0.00%	0.00%		
195		77.70%	10.34%	11.76%		0.00%
200		91.28%	53.93%	17.07%	0.00%	7.14%
205		98.47%	69.92%	43.06%	14.89%	6.25%
210		98.20%	86.99%	62.62%	26.15%	16.33%
215		100.00%	100.00%	83.45%	57.94%	24.05%
220				90.58%	76.22%	52.83%
225				95.92%	92.35%	68.61%
230				100.00%	96.95%	84.52%
235				98.59%	98.17%	97.52%
240				100.00%	100.00%	98.13%
245						100.00%
250						
255						

Figure 3 – Percent of Students Passing Reading NDSA by Fall RIT Performance Range
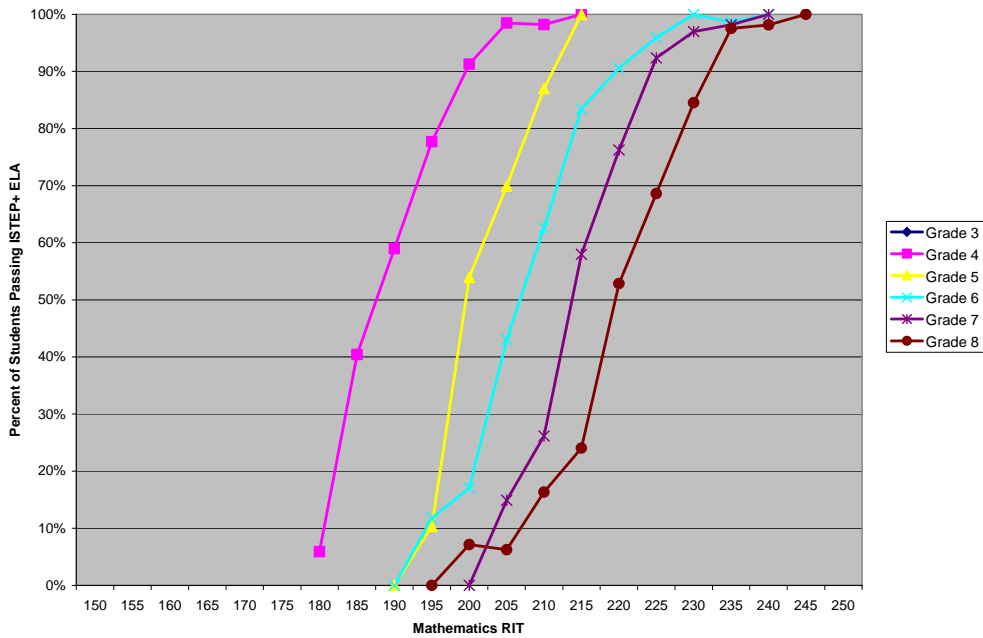

Figure 4 – Percent of Students Passing Reading NDSA by prior Spring RIT Performance Range
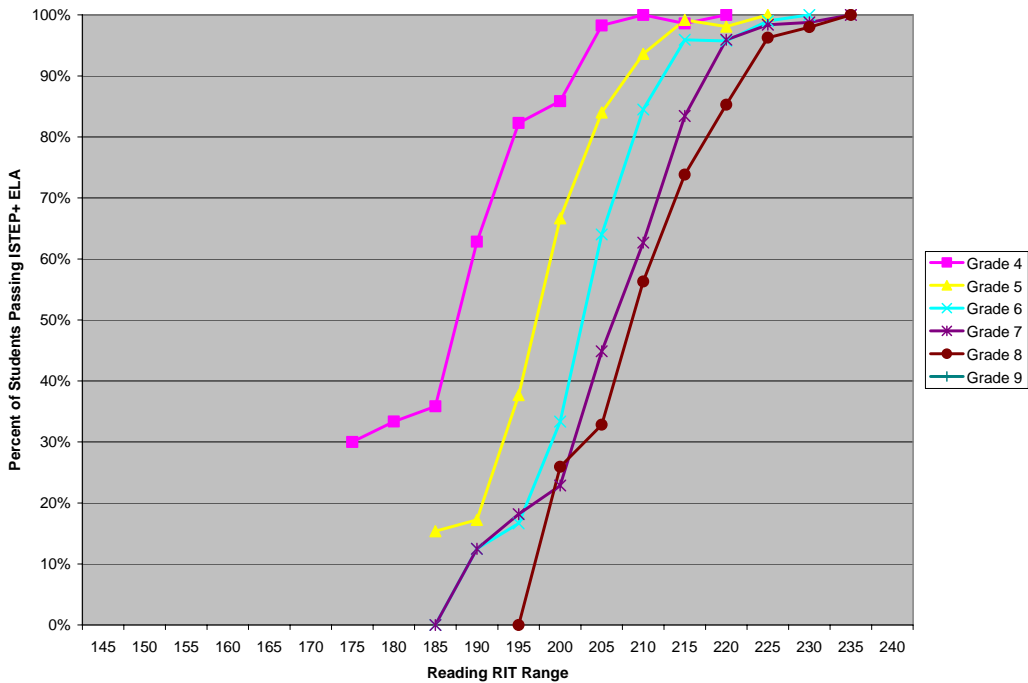
Figure 5 – Percent of Students Passing Mathematics NDSA by Fall RIT Performance Range
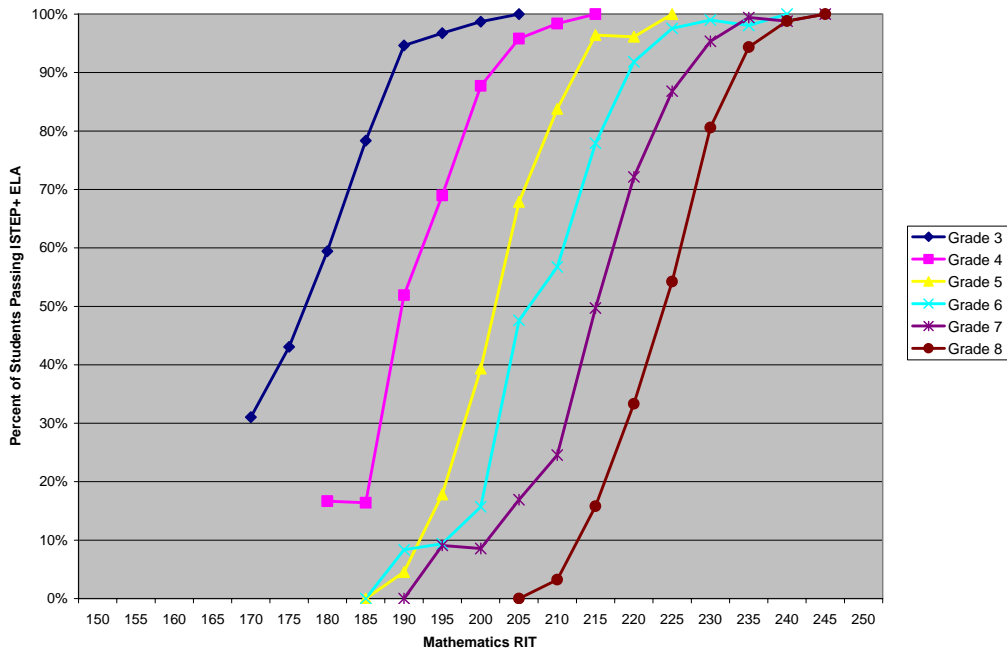
Figure 6 – Percent of Students Passing Mathematics NDSA by Prior Spring RIT Performance Range

Comparing the NDSA standards to other states

Northwest Evaluation Association tests have been aligned with the cut scores state assessments in 22 states.   To get an estimate of the difficulty of the NDSA 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.   Rather than report the results of our overview in this paper, we maintain a copy at the following link, so that you always have access to the most up-to-date results.

http://www.nwea.org/research/national.asp

Summary and Conclusions

This study investigated the relationship between the scales used for the NDSA 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 NDSA performance levels in those subjects. Test records for more than 9,000 students were included in this study.

Three methods generated an estimate of RIT cut scores that could be used to project NDSA performance levels. Rasch SOS methods generally produced the most accurate cut score estimates.   Accuracy of predicting NDSA proficient 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 North Dakota educators use NWEA assessments to better inform, plan and deliver student instruction. Good information, when matched with the professionalism and commitment of our North Dakota 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.