A Study of the Ongoing Alignment of the NWEA RIT Scale with the New Mexico Standards Based Assessments (NMSBA)

John Cronin, Ph.D and Branin Bowe

December 2005

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 New Mexico Standards Based Assessments (NMSBA) Achievement Tests

John Cronin, Ph.D. and Branin Bowe

December, 2005

Each year, New Mexico students participate in testing as part of the state's assessment program. In the spring of 2005, students in grades 3 through 9 participated in New Mexico Standards Based Assessments (NMSBA) tests in language arts 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 New Mexico Public 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 New Mexico also take tests developed in cooperation with the Northwest Evaluation Association (NWEA). The content of these tests is aligned with the New Mexico standards and the tests 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 levels. 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 NMSBA 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 NMSBA tests?
- How well can performance on the New Mexico 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 New Mexico are administered each spring. NWEA student assessment records in reading and mathematics were collected for the spring 2005 term and for the prior fall. Seven school systems supplied data for both terms.

Our study included more than 17,000 students in mathematics, about 14,000 students in reading, and approximately 9,000 students in language usage who are enrolled in New Mexico school systems. Student records were included when a student had both a valid NWEA scale score and a valid NMSBA score in the equivalent subject for at least one season. Tables 1 through 6 show the number of student records included for each subject and season.

Table 1 – Study Participants in Reading - Spring

District/Grade	3	4	5	6	7	8	9	Grand Total
CUBA	25	21	25	26	42	38		177
DULCE	32	44	48	41	41	52	50	308
JEMEZ MOUNTAIN	16	33	38	34	42	24	20	207
LOS ALAMOS	222	228	241	248	30	11		980
MORIARTY	231	247	264	296	277	274	294	1883
RIO RANCHO	795	808	784	879	809	865	7	4947
SANTA FE	847	825	913	960	736	745	571	5597
GRAND TOTAL	2168	2206	2313	2484	1977	2009	942	14099

Table 2 – Study Participants in Reading – Prior Fall

District/Grade	3	4	5	6	7	8	9	Grand Total
CUBA	25	21	25	26	42	38		177
DULCE	32	44	48	41	41	52	50	308
JEMEZ MOUNTAIN	16	33	38	34	42	24	20	207
LOS ALAMOS	222	228	241	248	30	11		980
MORIARTY	231	247	264	296	277	274	294	1883
RIO RANCHO	795	808	784	879	809	865	7	4947
SANTA FE	832	806	893	938	711	659	509	5348
GRAND TOTAL	2153	2187	2293	2462	1952	1923	880	13850

Table 3 – Study Participants in Language Usage - Spring

District/Grade	3	4	5	6	7	8	9	Grand Total
JEMEZ MOUNTAIN	16	33	38	34	42	24	20	207
LOS ALAMOS	222	228	241	248	30	11		980
MORIARTY	231	247	264	296	277	274	294	1883
RIO RANCHO	185	202	203	243	244			1077
SANTA FE	380	382	468	520	493	451		2694
GRAND TOTAL	1034	1092	1214	1341	1086	760	314	6841

Table 4 – Study Participants in Language Usage – Prior Fall

District/Grade	3	4	5	6	7	8	9	Grand Total
CUBA	24	20	25	26	42	38		175
LOS ALAMOS	222	228	241	245	30	10		976
MORIARTY	231	247	262	294	276	273	292	1875
RIO RANCHO	105	116	116	240	239			816
SANTA FE	811	806	889	941	710	727	522	5406
GRAND TOTAL	1393	1417	1533	1746	1297	1048	814	9248

Table 5 – Study Participants in Mathematics - Spring

District/Grade	3	4	5	6	7	8	9	Grand Total
CUBA	26	20	19	23	41	38		167
DULCE		13	17	31	38	50	53	202
FARMINGTON	612	552	620	650	659	761	96	3950
JEMEZ MOUNTAIN	16	33	38	33	41	25	20	206
LOS ALAMOS	225	231	243	143	103	1	2	948
MORIARTY	227	244	256	271	276	274	283	1831
RIO RANCHO	796	810	787	868	789	894	11	4955
SANTA FE	821	775	864	905	684	707	525	5281
GRAND TOTAL	2723	2678	2844	2924	2631	2750	990	17540

Table 6 – Study Participants in Mathematics – Prior Fall

District/Grade	3	4	5	6	7	8	9	Grand Total
CUBA	26	20	19	23	41	38		167
FARMINGTON	612	552	620	650	659	761	96	3950
LOS ALAMOS	225	231	243	143	103	1	2	948
MORIARTY	227	244	256	271	276	274	283	1831
RIO RANCHO	796	810	787	868	789	894	11	4955
SANTA FE	821	775	864	905	684	707	525	5281
GRAND TOTAL	2707	2632	2789	2860	2552	2675	917	17132

Data Preparation

For purposes of studying NWEA test alignment with the NMSBA, third through ninth grade student test records from spring 2005 and the prior fall (2004) assessments were matched with the 2005 NMSBA 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 scores. 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.

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 New Mexico assessment scores at each grade to determine how closely the scores on the NWEA test correlated with same subject scores on the NMSBA. Simple bivariate correlation coefficients were computed among these scores.

Linking NMSBA scores to the RIT scales. Spring and prior fall scores on the RIT scale were linked separately to the scale for the matching subject of the NMSBA. Three methods of estimating cut scores for NMSBA levels were used. The most straightforward was simple linear regression ($NMSBA_{pred}$ =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 ($NMSBA_{pred}=a(RIT^2) + b(RIT) +$ c). For each of these methods, the RIT score was determined by substituting the appropriate NMSBA score for $NMSBA_{pred}$ and solving the equation for RIT.

A fixed-parameter Rasch model was also used to estimate RIT cut scores. In this method, the NMSBA 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 NMSBA performance levels from RIT scores. Spring and prior fall RIT scores were first used to predict whether students were likely to achieve performance at or above the proficient performance level on the NMSBA. The predictions of NMSBA 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 NMSBA 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 NMSBA levels based on observed RIT scores. The predictions of NMSBA 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 New Mexico. We did not conduct additional comparisons of the content of NWEA and New Mexico tests as part of this study. Nevertheless, the standards used to construct the NWEA assessments were the same as those used for the New Mexico assessments. Both NWEA assessments and the NMSBA include multiple-choice items. The NMSBA 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 NMSBA 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 7 – Means, Standard Deviations, and Medians for NMSBA and NWEA Assessments

NMSBA Reading							
Grade	3	4	5	6	7	8	9
N	2168	2206	2313	2484	1977	2009	942
Mean	632.65	649.12	669.46	671.16	682.90	690.18	685.62
Median	636	651	673	674	687	691	688
Std Dev	28.51	31.24	32.98	28.71	32.13	27.036	34.76
NWEA Reading – Spring 2005							
Grade	3	4	5	6	7	8	9
N	2168	2206	2313	2484	1977	2009	942
Mean	198.43	204.78	210.17	215.57	216.37	221.12	217.54
Median	200	207	212	217	218	222	219
Std Dev	14.19	13.74	13.41	13.92	13.85	12.56	15.80
NWEA Reading – Fall 2004							
Grade	3	4	5	6	7	8	9
N	2105	2110	2207	2387	1869	1847	810
Mean	189.70	198.35	205.62	211.46	212.83	218.04	217.19
Median	192	200	207	213	215	219	218
Std Dev	14.86	14.40	13.60	14.16	13.89	12.33	13.86
NWEA Language Usage – Spring 2005							
Grade	3	4	5	6	7	8	9
N	1057	1110	1235	1363	1127	797	358
Mean	201.41	208.26	212.39	215.48	214.42	216.72	217.88
Median	202	210	214	217	216	218	219
Std Dev	12.49	12.64	11.78	12.20	11.59	11.06	13.50
NWEA Language Usage – Fall 2004							
Grade	3	4	5	6	7	8	9
N	1394	1419	1533	1768	1326	1048	814
Mean	191.03	200.55	207.59	211.67	211.39	214.47	217.82
Median	192	202	209	213	213	215.5	218
Std Dev	13.62	12.97	12.09	12.40	12.05	10.92	10.76
NMSBA Mathematics							
Grade	3	4	5	6	7	8	9
N	2723	2678	2844	2924	2631	2750	990
Mean	612.13	634.54	647.51	662.03	671.80	692.01	701.28
Median	611	634	646	660	669	691	698
Std Dev	31.05	33.46	28.72	32.72	31.38	32.73	32.03
NWEA Mathematics – Spring 2005							
Grade	3	4	5	6	7	8	9
N	2723	2678	2844	2924	2631	2750	990
Mean	201.72	210.10	217.14	223.07	226.68	232.98	227.68
Median	202	210	217	224	228	234	228
Std Dev	10.92	11.98	12.44	14.21	14.79	14.63	15.14
NWEA Mathematics – Fall 2004							
Grade	3	4	5	6	7	8	9
N	2707	2632	2789	2860	2552	2675	917
Mean	191.83	202.38	210.26	217.11	221.29	227.76	226.66
Median	192	203	210	218	222	229	226
Std Dev	20.78	11.08	11.22	12.78	13.55	13.53	13.70

Pearson correlations

Tables 8 through 10 show the results of this analysis for each grade. Concurrent validity was tested by examining same subject Pearson correlations between the NWEA and NMSBA assessments. When the NWEA and NMSBA tests were administered during the same season (Spring 2005) correlations ranged from .75 to .82 between NWEA reading and the NMSBA language arts. Same season correlations for the NWEA language usage and NMSBA language arts ranged between .73 and .79. Same season correlations between NWEA and NMSBA mathematics tests ranged between .79 and .86. In all cases these correlations were strong enough to suggest that the tests were measuring similar constructs.

When the NWEA test was administered in the season prior to NMSBA, correlations ranged between .72 and .80 in reading, .72 and .80 in language usage, and .76 and .84 in mathematics. Once again these correlations are strong enough to suggest that the tests were measuring similar constructs. Overall, the strength of the Pearson coefficients generated by this study were typical of those generated in other alignment studies that we have conducted.

Table 8 – Reading Inter-test Correlations for NMSBA and NWEA Assessments

Grade 3			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.82	.78
NWEA Spring 2005	.82	1	.81
NWEA Fall 2004	.78	.81	1
Grade 4			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.78	.77
NWEA Spring 2005	.78	1	.82
NWEA Fall 2004	.77	.82	1
Grade 5			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.80	.78
NWEA Spring 2005	.80	1	.82
NWEA Fall 2004	.78	.82	1
Grade 6			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.79	.80
NWEA Spring 2005	.79	1	.84
NWEA Fall 2004	.80	.84	1
Grade 7			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.82	.80
NWEA Spring 2005	.82	1	.82
NWEA Fall 2004	.80	.82	1
Grade 8			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.75	.73
NWEA Spring 2005	.76	1	.80
NWEA Fall 2004	.73	.80	1
Grade 9			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.76	.72
NWEA Spring 2005	.76	1	.76
NWEA Fall 2004	.72	.76	1

Table 9 – Language Usage Inter-test Correlations for NMSBA and NWEA Assessments

Grade 3			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.77	.80
NWEA Spring 2005	.77	1	.81
NWEA Fall 2004	.80	.81	1
Grade 4			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.79	.77
NWEA Spring 2005	.79	1	.80
NWEA Fall 2004	.77	.80	1
Grade 5			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.79	.77
NWEA Spring 2005	.78	1	.81
NWEA Fall 2004	.77	.84	1
Grade 6			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.78	.77
NWEA Spring 2005	.78	1	.84
NWEA Fall 2004	.77	.84	1
Grade 7			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.76	.74
NWEA Spring 2005	.76	1	.81
NWEA Fall 2004	.74	.81	1
Grade 8			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.74	.73
NWEA Spring 2005	.74	1	.81
NWEA Fall 2004	.73	.81	1
Grade 9			
	NMSBA Language Arts	NWEA Spring 2005	NWEA Fall 2004
NMSBA Language Arts	1	.73	.72
NWEA Spring 2005	.73	1	.76
NWEA Fall 2004	.72	.76	1

Table 10 – Mathematics Inter-test Correlations for NMSBA and NWEA Assessments

Grade 3			
	NMSBA Mathematics	NWEA Spring 2005	NWEA Fall 2004
NMSBA Mathematics	1	.79	.76
NWEA Spring 2005	.79	1	.78
NWEA Fall 2004	.76	.78	1
Grade 4			
	NMSBA Mathematics	NWEA Spring 2005	NWEA Fall 2004
NMSBA Mathematics	1	.84	.81
NWEA Spring 2005	.84	1	.82
NWEA Fall 2004	.81	.82	1
Grade 5			
	NMSBA Mathematics	NWEA Spring 2005	NWEA Fall 2004
NMSBA Mathematics	1	.84	.81
NWEA Spring 2005	.84	1	.82
NWEA Fall 2004	.81	.82	1
Grade 6			
	NMSBA Mathematics	NWEA Spring 2005	NWEA Fall 2004
NMSBA Mathematics	1	.86	.84
NWEA Spring 2005	.86	1	.85
NWEA Fall 2004	.84	.85	1
Grade 7			
	NMSBA Mathematics	NWEA Spring 2005	NWEA Fall 2004
NMSBA Mathematics	1	.86	.83
NWEA Spring 2005	.86	1	.86
NWEA Fall 2004	.83	.86	1
Grade 8			
	NMSBA Mathematics	NWEA Spring 2005	NWEA Fall 2004
NMSBA Mathematics	1	.86	.84
NWEA Spring 2005	.86	1	.87
NWEA Fall 2004	.84	.87	1
Grade 9			
	NMSBA Mathematics	NWEA Spring 2005	NWEA Fall 2004
NMSBA Mathematics	1	.80	.80
NWEA Spring 2005	.80	1	.82
NWEA Fall 2004	.80	.82	1

A review of scatterplots showed that the NMSBA and NWEA tests generally maintained a linear relationship with one another, although there was some evidence of a curvilinear relationship and floor effect emerging in the upper grades. Figure 1 illustrates the most common pattern with grade 3 mathematics. It is obvious from the diagram that the relationship is linear. At the bottom end of the

scale the relationship breaks down a bit (evidenced by a slight burst effect). This typically occurs because students who have an off-day on one test will often perform better when taking the second test under more motivating conditions.

Figure 2 shows an example of floor effect. In this case students performing below 650 on the state test, produce RIT scores that cluster from RIT 175 to RIT 225 (one student even produced a score of RIT 250). This effect is often seen when one assessment has greater range at the low end than its companion. That would be expected with state assessments since state tests are designed to focus measurement on the grade level standards, and are not designed with the intention of producing highly accurate measurement for the lowest students. NWEA assessments are designed to align with the New Mexico 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 3 Mathematics NMSBA score plotted against Spring Mathematics RIT score

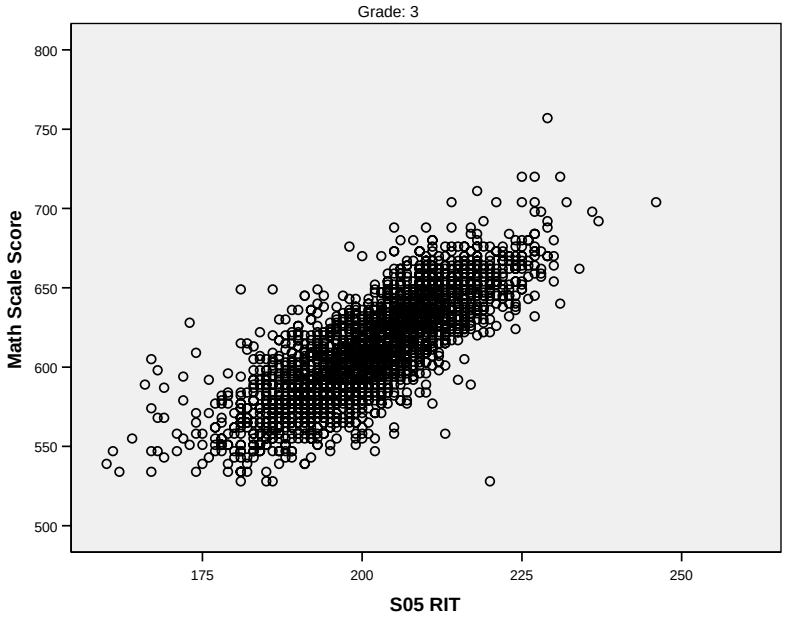
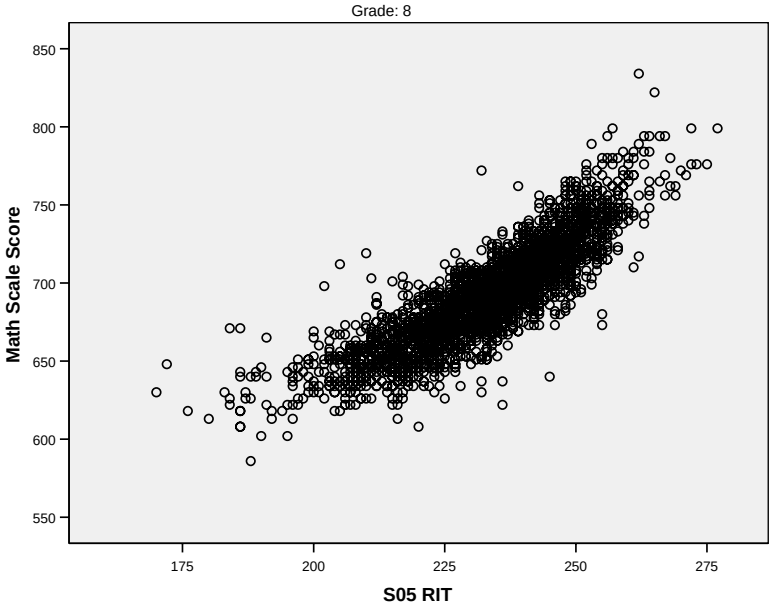


Figure 2 – Grade 8 Mathematics NMSBA score plotted against Spring Mathematics RIT score



Linking NMSBA performance level cut scores to the RIT scales

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 NMSBA. 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 11 shows several estimations of the spring and prior fall RIT scores that correspond to the cut scores for the various performance levels on the NMSBA scales. The estimates were generally quite close, with no set of estimates for a single grade differing by more than 4 RIT points.

Table 11 – Estimated points on the RIT scale equating to the minimum scores (rounded) for performance levels on the NMSBA

Reading Spring – NMSBA Language Arts									
	Linear Regression			Second-order Regression			Rasch Status on Standard		
	Nearing Proficient	Proficient	Advanced	Nearing Proficient	Proficient	Advanced	Nearing Proficient	Proficient	Advanced
Grade 3	175	192	222	175	192	221	172	190	223
Grade 4	177	200	224	176	201	223	173	199	225
Grade 5	180	204	223	180	204	224	177	203	222
Grade 6	184	214	239	184	214	239	179	213	240
Grade 7	190	214	238	189	215	237	186	213	239
Grade 8	194	218	251	194	218	251	190	217	254
Grade 9	204	220	260	205	221	254	202	219	265
Reading Prior Fall – NMSBA Language Arts									
	Linear Regression			Second-order Regression			Rasch Status on Standard		
	Nearing Proficient	Proficient	Advanced	Nearing Proficient	Proficient	Advanced	Nearing Proficient	Proficient	Advanced
Grade 3	163	182	216	163	182	215	162	180	218
Grade 4	169	193	218	167	193	217	164	192	218
Grade 5	172	199	219	172	199	219	167	197	218
Grade 6	179	210	236	179	210	235	174	208	235
Grade 7	185	210	234	185	211	233	182	209	236
Grade 8	190	215	248	190	215	247	185	213	253
Grade 9	203	218	258	203	219	252	200	218	255
Language Usage Spring – NMSBA Language Arts									
	Linear Regression			Second-order Regression			Rasch Status on Standard		
	Nearing Proficient	Proficient	Advanced	Nearing Proficient	Proficient	Advanced	Nearing Proficient	Proficient	Advanced
Grade 3	179	195	223	178	195	222	175	192	225
Grade 4	182	203	225	180	204	224	172	202	225
Grade 5	185	207	224	185	207	224	179	206	223
Grade 6	189	215	237	189	216	235	187	214	238
Grade 7	194	216	238	193	217	236	192	214	243
Grade 8	197	219	247	197	219	245	194	218	243
Grade 9	205	220	256	206	221	254	204	219	300

Language Usage Prior Fall – NMSBA Language Arts									
	Linear Regression			Second-order Regression			Rasch Status on Standard		
	Nearing Proficient	Proficient	Advanced	Nearing Proficient	Proficient	Advanced	Nearing Proficient	Proficient	Advanced
Grade 3	169	186	215	169	185	216	167	182	217
Grade 4	176	196	219	175	196	219	170	195	219
Grade 5	179	203	221	179	203	221	174	201	219
Grade 6	184	211	234	183	212	232	178	210	234
Grade 7	189	213	235	188	213	234	187	211	244
Grade 8	192	215	245	192	215	245	189	212	300
Grade 9	206	219	250	207	219	244	202	218	243
Mathematics Spring									
	Linear Regression			Second-order Regression			Rasch Status on Standard		
	Nearing Proficient	Proficient	Advanced	Nearing Proficient	Proficient	Advanced	Nearing Proficient	Proficient	Advanced
Grade 3	178	201	224	175	202	222	171	201	224
Grade 4	189	211	228	187	211	228	181	210	228
Grade 5	198	223	242	196	223	240	193	222	241
Grade 6	201	231	249	197	232	245	204	231	247
Grade 7	213	238	259	213	239	254	211	237	259
Grade 8	216	240	263	216	241	258	214	240	262
Grade 9	208	237	270	207	238	259	204	238	264
Mathematics Prior Fall									
	Linear Regression			Second-order Regression			Rasch Status on Standard		
	Nearing Proficient	Proficient	Advanced	Nearing Proficient	Proficient	Advanced	Nearing Proficient	Proficient	Advanced
Grade 3	167	191	214	165	192	213	160	191	215
Grade 4	182	203	220	179	204	219	175	202	219
Grade 5	192	215	234	190	216	232	186	214	232
Grade 6	196	224	241	193	225	238	199	224	238
Grade 7	207	232	252	208	232	247	206	230	254
Grade 8	211	234	256	210	235	252	209	235	255
Grade 9	209	235	265	207	236	255	205	236	259

Establishing RIT score estimates for NMSBA 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 NMSBA assessment. The most accurate method of prediction was generally used to derive the best estimate of RIT cut scores that equate to the different NMSBA 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 NMSBA test. Next we assessed the accuracy with which the RIT score predicted the 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 New Mexico assessment that corresponds to their grade level.

Tables 12 through 14 summarize the results. In reading, when using the most accurate method, the accuracy of pass/fail prediction ranged between about 76% and 85% for spring data and between about 76% and 84% for prior fall data. Prediction accuracy for grade 9 in reading was substantively lower than prediction for the other grades. In language usage, the accuracy of pass/fail prediction ranged from about 75% to 84% for spring data and between about 75% and 82% for prior fall data. Finally, the accuracy of pass/fail prediction for mathematics ranged from about 82% to 89% for spring data and between 79% and 87% for prior fall data.

The relatively low rate of prediction (76%) for grade 9 reading is a concern and the reasons behind it are not immediately obvious. One possible reason is that grade 9 is not an NCLB reported grade, thus the stakes for grade 9 testing are somewhat different than they are for grades 3 through 8. However, the accuracy rate of pass/fail prediction for grade 9 mathematics was quite high (86%), which makes it unlikely that the lower stakes associated with the grade 9 test would explain this difference. A second possibility would be a shift in emphasis on the grade 9 test. For example, if the grade 9 test places more emphasis on interpretation of literature or writing and less on domains related directly to reading, this would cause a decline in the accuracy of prediction. Unfortunately, we are not in a position to know if that is the reason for the difference.

That said the level of accuracy reported for most grades in this study should be more than adequate to permit the use of NWEA assessments as a tool to identify students who might be at risk relative to passing the state test.

Table 12 – Evaluation of Projected RIT cut scores for NMSBA *proficient* level - Reading

	Spring				Prior Fall			
Grade 3	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	**192**	**84.3%**	**9.5%**	**.888**	**182**	**88.3%**	**9.9%**	**.881**
Second Order	**192**	**84.3%**	**9.5%**	**.888**	**182**	**83.3%**	**9.9%**	**.881**
Rasch	190	84.6%	11.5%	.864	180	82.7%	12.0%	.855
Grade 4	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	200	82.1%	11.3%	.862	**193**	**82.3%**	**10.7%**	**.870**
Second Order	**201**	**82.3%**	**9.9%**	**.880**	**193**	**82.3%**	**10.7%**	**.870**
Rasch	199	81.9%	12.8%	.844	192	82.3%	11.5%	.860
Grade 5	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	**204**	**84.6%**	**9.5%**	**.888**	**199**	**83.7%**	**9.7%**	**.884**
Second Order	**204**	**84.6%**	**9.5%**	**.888**	**199**	**83.7%**	**9.7%**	**.884**
Rasch	203	84.3%	10.9%	.871	197	83.5%	11.7%	.860
Grade 6	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	**214**	**80.5%**	**10.3%**	**.872**	**210**	**81.3%**	**9.7%**	**.880**
Second Order	**214**	**80.5%**	**10.3%**	**.872**	**210**	**81.3%**	**9.7%**	**.880**
Rasch	213	80.6%	11.6%	.856	208	80.0%	13.2%	.836
Grade 7	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	214	82.3%	9.9%	.880	210	83.6%	10.2%	.878
Second Order	**215**	**82.3%**	**8.6%**	**.896**	**211**	**83.0%**	**9.0%**	**.891**
Rasch	213	82.5%	11.1%	.865	209	83.1%	11.7%	.859
Grade 8	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	**218**	**79.8%**	**12.1%**	**.849**	**215**	**79.6%**	**11.5%**	**.855**
Second Order	**218**	**79.8%**	**12.1%**	**.849**	**215**	**79.6%**	**11.5%**	**.855**
Rasch	217	80.4%	13.0%	.838	213	79.1%	14.5%	.817
Grade 9	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	220	74.9%	13.3%	.823	218	76.5%	12.2%	.840
Second Order	**221**	**75.8%**	**11.6%**	**.847**	**219**	**76.3%**	**10.6%**	**.861**
Rasch	219	75.2%	14.6%	.805	218	76.5%	12.2%	.840

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

Table 13 – Evaluation of Projected RIT cut scores for NMSBA *proficient* level – Language Usage

	Spring				Prior Fall			
Grade 3	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	**195**	**82.2%**	**9.5%**	**.884**	**186**	**80.3%**	**10.2%**	**.873**
Second Order	**195**	**82.2%**	**9.5%**	**.884**	185	80.0%	11.4%	.857
Rasch	192	82.6%	12.7%	.846	182	81.2%	13.7%	.831
Grade 4	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	203	82.8%	10.9%	.868	**196**	**82.1%**	**10.3%**	**.875**
Second Order	**204**	**83.6%**	**9.1%**	**.891**	**196**	**82.1%**	**10.3%**	**.875**
Rasch	202	81.7%	12.6%	.846	195	81.7%	11.9%	.854
Grade 5	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	**207**	**81.8%**	**11.4%**	**.861**	**203**	**81.1%**	**10.8%**	**.867**
Second Order	**207**	**81.8%**	**11.4%**	**.861**	**203**	**81.1%**	**10.8%**	**.867**
Rasch	206	82.2%	12.4%	.849	201	81.4%	13.6%	.833
Grade 6	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	215	78.8%	13.0%	.835	211	79.7%	12.0%	.850
Second Order	**216**	**78.0%**	**11.6%**	**.851**	**212**	**79.8%**	**10.2%**	**.872**
Rasch	214	78.5%	14.8%	.811	210	78.8%	14.1%	.820
Grade 7	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	216	79.1%	10.3%	.869	**213**	**77.7%**	**10.6%**	**.864**
Second Order	**217**	**78.8%**	**8.5%**	**.892**	**213**	**77.7%**	**10.6%**	**.864**
Rasch	214	76.9%	15.1%	.803	211	77.7%	13.9%	.821
Grade 8	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	**219**	**74.7%**	**10.9%**	**.854**	**215**	**75.2%**	**11.7%**	**.844**
Second Order	**219**	**74.7%**	**10.9%**	**.854**	**215**	**75.2%**	**11.7%**	**.844**
Rasch	218	75.8%	12.6%	.833	212	73.9%	17.8%	.758
Grade 9	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	220	78.5%	10.9%	.861	**219**	**76.0%**	**11.5%**	**.848**
Second Order	**221**	**76.9%**	**10.3%**	**.867**	**219**	**76.0%**	**11.5%**	**.848**
Rasch	219	78.2%	13.5%	.828	218	79.1%	14.0%	.822

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

Table 14 – Evaluation of Projected RIT cut scores for NMSBA *proficient* level – Mathematics

	Spring				Prior Fall			
Grade 3	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	201	81.1%	11.2%	.862	191	78.6%	12.4%	.842
Second Order	**202**	**81.5%**	**9.3%**	**.886**	**192**	**78.6%**	**10.7%**	**.864**
Rasch	201	81.1%	11.2%	.862	191	78.6%	12.4%	.842
Grade 4	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	**211**	**81.9%**	**8.4%**	**.897**	203	80.9%	9.6%	.881
Second Order	**211**	**81.9%**	**8.4%**	**.897**	**204**	**81.0%**	**7.7%**	**.905**
Rasch	210	82.2%	10.2%	.876	202	80.2%	12.0%	.851
Grade 5	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	**223**	**85.1%**	**6.5%**	**.923**	215	83.0%	7.7%	.908
Second Order	**223**	**85.1%**	**6.5%**	**.923**	**216**	**83.3%**	**5.7%**	**.932**
Rasch	222	84.9%	8.1%	.905	214	83.0%	9.0%	.892
Grade 6	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	231	88.0%	6.0%	.932	224	86.9%	6.8%	.922
Second Order	**232**	**88.2%**	**4.3%**	**.951**	**225**	**87.0%**	**5.3%**	**.939**
Rasch	231	88.0%	6.0%	.932	224	86.9%	6.8%	.922
Grade 7	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	238	89.5%	3.8%	.958	**232**	**86.6%**	**5.0%**	**.943**
Second Order	**239**	**89.1%**	**3.0%**	**.966**	**232**	**86.6%**	**5.0%**	**.943**
Rasch	237	89.7%	4.8%	.947	230	86.3%	7.7%	.911
Grade 8	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	240	87.4%	7.3%	.917	234	85.3%	9.0%	.895
Second Order	**241**	**87.5%**	**5.9%**	**.933**	**235**	**85.8%**	**7.0%**	**.918**
Rasch	240	87.4%	7.3%	.917	**235**	**85.8%**	**7.0%**	**.918**
Grade 9	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	237	85.9%	6.6%	.924	235	85.9%	8.4%	.900
Second Order	**238**	**85.9%**	**5.5%**	**.936**	**236**	**86.2%**	**5.8%**	**.933**
Rasch	**238**	**85.9%**	**5.5%**	**.936**	**236**	**86.2%**	**5.8%**	**.933**

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* levels and to define the cut score for the *advanced* level. The following methods were used to establish these:

- **Nearing Proficiency/Beginning Step.** We selected the method that correctly identified the largest proportion of students who scored at the *beginning step* level.
- **Advanced.** We selected the method that correctly identified the largest proportion of students who scored in the *advanced* category on the NMSBA. 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 15 and 16.

Table 15 – Evaluation of Projected RIT cut scores for NMSBA *beginning step, nearing proficiency, and advanced* performance levels – Reading to NMSBA Language Arts

Grade	Method	Spring Reading					Prior Fall Reading				
		Beginning Step/Nearing Proficiency		Advanced		Prediction Index	Beginning Step/Nearing Proficiency		Advanced		Prediction Index
		Cut Score	% Beg Found	Cut Score	% Adv Found		Cut Score	% Beg Found	Cut Score	% Adv Found	
3	Linear	**175**	**46.7%**	222	21.1%	.798	**163**	**40.0%**	216	13.3%	.778
	Second Order	**175**	**46.7%**	**221**	**24.7%**	.795	**163**	**40.0%**	**215**	**16.4%**	.768
	Rasch	172	38.6%	223	16.9%	.768	162	35.1%	218	9.1%	.754
4	Linear	**177**	**45.7%**	224	26.2%	.749	**169**	**42.0%**	218	29.1%	.757
	Second Order	176	43.3%	**223**	**31.8%**	.758	167	40.2%	**217**	**33.8%**	.747
	Rasch	173	30.7%	225	19.7%	.716	164	30.4%	218	29.1%	.735
5	Linear	**180**	**44.2%**	**223**	**50.5%**	.749	**172**	**34.6%**	**219**	**45.5%**	.747
	Second Order	**180**	**44.2%**	224	45.4%	.762	**172**	**34.6%**	**219**	**45.5%**	.747
	Rasch	177	32.7%	222	55.5%	.699	167	22.2%	218	50.6%	.692
6	Linear	**184**	**39.6%**	**239**	**18.5%**	.800	**179**	**46.1%**	236	15.4%	.825
	Second Order	**184**	**39.6%**	**239**	**18.5%**	.800	**179**	**46.1%**	**235**	**18.8%**	.819
	Rasch	179	27.7%	240	15.6%	.776	174	30.3%	**235**	**18.8%**	.757
7	Linear	**190**	**48.1%**	238	27.6%	.785	**185**	**43.3%**	234	25.4%	.783
	Second Order	189	45.1%	**237**	**30.9%**	.795	**185**	**43.3%**	**233**	**28.7%**	.785
	Rasch	186	35.3%	239	22.1%	.764	182	35.8%	236	13.8%	.764
8	Linear	**194**	**38.6%**	**251**	**1.6%**	.793	**190**	**27.7%**	248	0.0%	.796
	Second Order	**194**	**38.6%**	**251**	**1.6%**	.793	**190**	**27.7%**	**247**	**0.0%**	.796
	Rasch	190	32.7%	254	0.0%	.779	185	20.5%	253	0.0%	.753
9	Linear	204	52.5%	260	0.0%	.639	**203**	**41.4%**	258	0.0%	.649
	Second Order	**205**	**55.1%**	**254**	**0.0%**	.672	**203**	**41.4%**	**252**	**5.6%**	.669
	Rasch	202	49.5%	265	0.0%	.615	200	31.6%	255	5.6%	.616

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

Table 16 – Evaluation of Projected RIT cut scores for NMSBA *beginning step, nearing proficiency, and advanced* performance levels – Language Usage to NMSBA Language Arts

Grade	Method	Spring Language Usage					Prior Fall Language Usage				
		Beginning Step/Nearing Proficiency		Advanced		Prediction Index	Beginning Step/Nearing Proficiency		Advanced		Prediction Index
		Cut Score	% Beg Found	Cut Score	% Adv Found		Cut Score	% Beg Found	Cut Score	% Adv Found	
3	Linear	**179**	**34.6%**	223	21.1%	.790	**169**	**47.2%**	**215**	**18.1%**	.754
	Second Order	178	29.6%	**222**	**25.6%**	.779	**169**	**47.2%**	216	17.1%	.745
	Rasch	175	19.8%	225	14.4%	.747	167	38.7%	217	13.1%	.715
4	Linear	**182**	**47.1%**	225	38.6%	.767	**176**	**48.3%**	**219**	**29.5%**	.755
	Second Order	180	43.1%	**224**	**41.0%**	.789	175	46.1%	**219**	**29.5%**	.753
	Rasch	172	17.6%	225	38.6%	.721	170	23.6%	**219**	**29.5%**	.703
5	Linear	**185**	**44.0%**	224	45.1%	.705	**179**	**25.8%**	221	40.8%	.710
	Second Order	**185**	**44.0%**	224	45.1%	.705	**179**	**25.8%**	221	40.8%	.710
	Rasch	179	26.0%	**223**	**49.8%**	.661	174	18.2%	**219**	**52.8%**	.634
6	Linear	**189**	**46.8%**	237	12.3%	.765	**184**	**39.0%**	**234**	**9.3%**	.782
	Second Order	**189**	**46.8%**	**235**	**20.2%**	.772	183	36.4%	232	17.1%	.792
	Rasch	187	41.9%	238	9.6%	.737	178	29.9%	**234**	**9.3%**	.740
7	Linear	**194**	**38.1%**	238	2.1%	.765	**189**	**38.9%**	235	4.8%	.768
	Second Order	193	35.1%	**236**	**8.3%**	.776	188	38.1%	**234**	**6.5%**	.764
	Rasch	192	32.0%	243	0.0%	.681	187	38.1%	244	0.0%	.724
8	Linear	**197**	**38.8%**	247	0.0%	.762	**192**	**31.5%**	**245**	**0.0%**	.760
	Second Order	**197**	**38.8%**	245	9.1%	.760	**192**	**31.5%**	**245**	**0.0%**	.760
	Rasch	194	32.8%	**243**	**9.1%**	.731	189	23.3%	Unable to estimate		
9	Linear	205	42.6%	256	0.0%	.639	206	42.6%	250	0.0%	.664
	Second Order	**206**	**50.0%**	**254**	**0.0%**	.674	**207**	**43.2%**	244	5.6%	.662
	Rasch	204	39.7%	Unable to estimate			202	26.4%	**243**	**5.6%**	.577

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

Table 17 – Evaluation of Projected RIT cut scores for NMSBA *beginning step, nearing proficiency, and advanced* performance levels – Mathematics

Grade	Method	Spring Mathematics					Prior Fall Mathematics				
		Beginning Step/Nearing Proficiency		Advanced		Prediction Index	Beginning Step/Nearing Proficiency		Advanced		Prediction Index
		Cut Score	% Beg Found	Cut Score	% Adv Found		Cut Score	% Beg Found	Cut Score	% Adv Found	
3	Linear	**178**	**18.2%**	224	28.3%	.796	**167**	**12.2%**	214	20.2%	.771
	Second Order	175	12.1%	**222**	**35.8%**	.812	165	9.2%	**213**	**22.5%**	.787
	Rasch	171	8.1%	224	28.3%	.791	160	6.1%	215	17.3%	.770
4	Linear	**189**	**30.7%**	**228**	**53.8%**	.795	**182**	**34.0%**	220	40.2%	.790
	Second Order	187	24.7%	**228**	**53.8%**	.790	179	24.3%	**219**	**45.5%**	.808
	Rasch	181	7.3%	228	53.8%	.751	175	15.3%	**219**	**45.5%**	.734
5	Linear	**198**	**39.0%**	242	31.3%	.845	**192**	**34.6%**	234	29.4%	.8722
	Second Order	196	30.3%	**240**	**39.9%**	.829	190	26.3%	**232**	**38.7%**	.838
	Rasch	193	20.6%	241	35.0%	.802	186	19.7%	**232**	**38.7%**	.781
6	Linear	201	31.2%	249	28.8%	.764	196	24.3%	241	26.4%	.739
	Second Order	197	18.7%	**245**	**53.8%**	.731	193	16.3%	**238**	**41.0%**	.730
	Rasch	**204**	**42.7%**	247	35.4%	.777	**199**	**42.2%**	**238**	**41.0%**	.763
7	Linear	**213**	**62.3%**	259	16.7%	.841	207	53.5%	252	8.9%	.800
	Second Order	**213**	**62.3%**	**254**	**41.1%**	.838	**208**	**57.7%**	**247**	**34.4%**	.796
	Rasch	211	54.4%	259	16.7%	.805	206	52.5%	254	4.4%	.755
8	Linear	**216**	**60.4%**	263	20.0%	.826	**211**	**49.7%**	256	14.5%	.773
	Second Order	**216**	**60.4%**	**258**	**40.0%**	.835	210	46.0%	**252**	**33.1%**	.781
	Rasch	214	51.5%	262	22.4%	.807	209	44.7%	255	20.0%	.785
9	Linear	**208**	**43.0%**	270	9.1%	.804	**209**	**43.5%**	265	9.1%	.785
	Second Order	207	39.3%	**259**	**27.3%**	.810	207	37.1%	**255**	**22.7%**	.794
	Rasch	204	30.4%	264	18.2%	.795	205	34.8%	259	13.6%	.788

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

When applying the selected methodology, spring NWEA reading assessments identified between about 39% and 55% of the students performing in the *beginning step* classification in language arts, while the language usage assessment identified between 35% and 50% of students performing in that classification.

In mathematics, the spring assessment identified between about 18% and 62% of the students at the *beginning step* level. When using prior fall data, the NWEA reading assessments identified between about 28% and 48% of the students performing at the lowest level, while the language usage assessment identified between 26% and 48% of the students in this category. The prior fall NWEA mathematics assessment identified between 12% and 58% of the students performing at the *beginning step* level.

Some of the prediction accuracy rates for the highest and lowest performance levels were lower than we typically see in studies of this type. This usually occurs when a very small proportion of the sample population performs at these levels on the state assessment. In grades 8 and 9 reading, for example, only 84 of the 2951 students tested achieved *advanced* status on the state test. This sample is inadequate to allow estimation of an absolutely stable cut score for that performance level, and explains some of the low level of prediction. In grade 3 mathematics we found a similar problem; only 99 of the 2723 students tested performed at the *beginning step*.

One artifact of this is that the estimates of *advanced* performance do not always seem to calibrate across grades evenly. Table 18 shows the cut score recommendations for each subject. In reading, the estimate for *advanced* performance was 223 RIT for grades 4 and 5. This estimate jumps to 239 for grade 6 and then slips to 237 for grade 7. This is probably a product of both the small sample available for this level and, between grade 5 and 6, a true difference in the difficulty of the standard set.

We do not believe that the small number of students in the sample performing at level 1 and level 4 were a product of a skewed distribution. Rather we believe few students were sampled in these ranges because the cut scores associated with *beginning step* were generally very low and those associated with *advanced* very high.

Finally, we also generally find some degradation in the accuracy of prediction when the state combines reading with writing and language usage skills into a single domain on the state assessment. Since New Mexico combines the domains, it is possible that this slightly degraded our ability to predict performance levels for Language Arts from the reading or language usage assessments by themselves.

Table 18 – Recommended RIT cut scores for NMSBA performance levels

Reading – NMSBA Language Arts								
Grade	Spring				Prior Fall			
	Beginning Step	Nearing Proficiency	**Proficient**	Advanced	Beginning Step	Nearing Proficiency	**Proficient**	Advanced
3	<175	175	**192**	221	<163	163	**182**	215
4	<177	177	**201**	223	<169	169	**193**	217
5	<180	180	**204**	223	<172	172	**199**	219
6	<184	184	**214**	239	<179	179	**210**	235
7	<190	190	**215**	237	<185	185	**211**	233
8	<194	194	**218**	251	<190	190	**215**	247
9	<205	205	**221**	254	<203	203	**219**	252
Language Usage – NMSBA Language Arts								
Grade	Spring				Prior Fall			
	Beginning Step	Nearing Proficiency	**Proficient**	Advanced	Beginning Step	Nearing Proficiency	**Proficient**	Advanced
3	<179	179	**195**	222	<169	169	**186**	215
4	<182	182	**204**	224	<176	176	**196**	219
5	<185	185	**207**	223	<179	179	**203**	219
6	<189	189	**216**	235	<184	184	**212**	234
7	<194	194	**217**	236	<189	189	**213**	234
8	<197	197	**219**	243	<192	192	**215**	245
9	<206	206	**221**	254	<207	207	**219**	243
Mathematics								
Grade	Spring				Prior Fall			
	Beginning Step	Nearing Proficiency	**Proficient**	Advanced	Beginning Step	Nearing Proficiency	**Proficient**	Advanced
3	<178	178	**202**	222	<167	167	**192**	213
4	<189	189	**211**	228	<182	182	**204**	219
5	<198	198	**223**	240	<192	192	**216**	232
6	<204	204	**232**	245	<199	199	**225**	238
7	<213	213	**239**	254	<208	208	**232**	247
8	<216	216	**241**	258	<211	211	**235**	252
9	<208	208	**238**	259	<209	209	**236**	255

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 spring). Table 19 summarizes the accuracy of proficiency status prediction for this study relative to other state alignment studies and Table 20 summarizes the accuracy of performance level prediction. The results show that the prediction index statistics for proficiency status on the NMSBA language arts assessment, using our reading and language usage tests, is low relative to other states studied, while mathematics prediction is

slightly above the median for the group. Interestingly, in spite of the issues cited relative to prediction of the *beginning step* and *advanced* level, the performance level predictions for reading were in the upper third of the group of states studied. The performance index statistics for mathematics were fourth highest among all the studies conducted to date.

Table 19 – 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	New Mexico	.872	South Carolina Exit	.933
Arizona '03	.912			New Mexico	.928
Colorado '01	.910			Pennsylvania	.926
Montana	.903			Washington '99	.920
Nevada	.902			Arizona '03	.919
South Carolina '03	.902			South Carolina '04	.914
Indiana '01	.902			Washington '04	.912
Indiana '03	.900			California	.910
Washington '99	.893			Arizona '05	.910
Arizona '05	.891			Montana	.899
Washington '04	.886			Indiana '01	.899
South Carolina '04	.884			North Dakota	.890
New Mexico	.877			Nevada	.866
North Dakota	.868			Indiana '03	.860

Table 20 – 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	New Mexico	.811
Arizona '05	.781	Colorado	.808
New Mexico	.778	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 NMSBA

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 NMSBA. The information in Tables 21 through 26 provides 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 NMSBA assessment. Using reading as an example (see Table 21), we find that about 17% of the grade 5 students who achieved a reading RIT score between 190 and 194 went on to achieve a proficient score on the NMSBA Language Arts assessment. A reading teacher would know that only about one in six of these students is likely to achieve a proficient score on the NMSBA 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 215 and 219 achieved proficiency on the New Mexico 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 through 8are graphic depictions of the data in the tables.

Table 21 - Proportion of Students Passing the NMSBA Language Arts Assessment Based on Same Spring RIT Reading Score

	Reading						
	3	4	5	6	7	8	9
160	0.0%	0.0%					
165	2.5%	6.7%					
170	5.1%	3.1%					
175	4.0%	5.1%	0.0%				
180	20.3%	9.3%	2.1%				
185	32.3%	17.4%	11.1%	0.0%			
190	55.3%	19.5%	17.2%	4.6%	1.8%		0.0%
195	73.0%	39.0%	21.1%	6.0%	4.6%	0.0%	2.6%
200	87.3%	54.2%	40.7%	11.2%	9.5%	10.3%	8.0%
205	95.4%	79.1%	67.9%	27.5%	20.3%	11.1%	12.5%
210	98.0%	92.8%	84.0%	46.2%	48.0%	28.4%	29.9%
215	100.0%	96.0%	94.4%	64.4%	66.3%	50.7%	45.1%
220		98.9%	96.7%	82.9%	84.3%	64.7%	47.3%
225		100.0%	100.0%	92.7%	95.8%	85.7%	76.8%
230				94.8%	98.3%	94.7%	84.5%
235				99.1%	100.0%	95.6%	93.8%
240				100.0%		98.4%	96.8%
245						95.2%	100.0%
250						100.0%	

Table 22 - Proportion of Students Passing the NMSBA Language Arts Assessment Based on Prior Fall RIT Reading Score

	Reading						
	3	4	5	6	7	8	9
150	3.7%	0.0%					
155	6.0%	9.1%					
160	3.4%	6.7%					
165	21.7%	8.3%	0.0%				
170	25.2%	9.4%	11.1%				
175	38.8%	14.9%	8.6%				
180	52.4%	19.5%	10.0%	0.0%			
185	69.3%	28.8%	14.0%	1.9%	4.4%		
190	86.0%	47.9%	33.8%	10.1%	6.0%	0.0%	
195	90.0%	63.5%	44.1%	14.8%	8.7%	14.3%	0.0%
200	97.5%	83.5%	66.4%	29.1%	16.4%	14.8%	6.3%
205	100.0%	93.3%	85.5%	36.6%	34.1%	26.6%	20.2%
210		97.6%	90.6%	64.3%	61.5%	38.6%	28.3%
215		99.2%	96.4%	80.1%	81.2%	64.5%	49.6%
220		100.0%	100.0%	93.4%	93.7%	78.9%	61.9%
225				95.7%	98.9%	91.3%	73.8%
230				98.5%	98.4%	95.5%	89.7%
235				100.0%	100.0%	94.4%	93.2%
240						100.0%	100.0%

Table 23 – Proportion of Students Passing the NMSBA Language Arts Assessment Based on Same Spring RIT Language Usage Score

	Language Usage						
	3	4	5	6	7	8	9
170	0.0%						
175	7.7%	0.0%					
180	20.0%	9.5%					
185	25.0%	16.7%		0.0%	0.0%		
190	48.7%	24.5%	0.0%	3.7%	7.7%		
195	66.4%	24.5%	17.1%	3.9%	5.7%	0.0%	6.7%
200	82.5%	41.3%	37.4%	10.5%	12.9%	7.9%	5.9%
205	90.3%	69.3%	50.9%	15.6%	14.1%	9.0%	15.2%
210	94.7%	85.7%	77.6%	41.0%	37.6%	25.9%	22.2%
215	98.9%	95.4%	87.6%	63.2%	59.7%	48.4%	31.3%
220	97.5%	97.2%	97.6%	77.1%	78.2%	71.2%	66.7%
225	100.0%	100.0%	98.1%	89.2%	92.3%	81.2%	85.7%
230			100.0%	97.0%	96.6%	94.9%	87.0%
235				100.0%	100.0%	92.9%	90.0%
240						100.0%	88.9%
245							0.0%

Table 24 - Proportion of Students Passing the NMSBA Language Arts Assessment Based on Prior Fall RIT Language Usage Score

	Language Usage						
	3	4	5	6	7	8	9
155	0.0%						
160	3.2%						
165	5.5%	0.0%					
170	16.2%	5.4%					
175	32.3%	16.3%					
180	55.8%	16.1%	0.0%				
185	62.3%	16.3%	4.7%	0.0%	0.0%		
190	75.0%	34.8%	17.5%	3.4%	5.6%	0.0%	
195	85.1%	57.8%	33.3%	10.3%	8.3%	12.1%	
200	94.7%	75.3%	52.1%	20.1%	16.7%	18.8%	6.5%
205	97.5%	86.5%	72.7%	33.2%	33.3%	24.4%	14.3%
210	100.0%	98.3%	84.2%	55.3%	51.7%	46.5%	18.8%
215		98.1%	94.4%	75.5%	70.5%	61.3%	46.1%
220		100.0%	99.3%	88.5%	92.2%	84.7%	68.5%
225			100.0%	98.4%	93.3%	91.2%	79.2%
230				100.0%	94.7%	94.1%	90.3%
235					100.0%	100.0%	96.8%
240							100.0%

Table 25 – Proportion of Students Passing the NMSBA Mathematics Assessment Based on Same Spring RIT Mathematics Score

	Mathematics						
	3	4	5	6	7	8	9
175	0.0%						
180	6.7%						
185	6.1%	0.0%					
190	14.1%	3.6%	0.0%				
195	26.8%	3.2%	0.9%				
200	47.2%	11.5%	0.9%	0.0%			
205	79.6%	33.9%	2.1%	0.9%			
210	94.8%	56.8%	7.7%	0.3%			0.0%
215	98.3%	84.7%	22.2%	2.7%	0.0%		1.1%
220	98.8%	93.5%	46.8%	7.9%	1.0%	0.0%	4.5%
225	100.0%	98.8%	74.5%	24.6%	5.7%	3.6%	7.5%
230		100.0%	88.2%	59.4%	20.3%	9.6%	21.9%
235			97.5%	80.2%	49.8%	24.6%	56.2%
240			98.2%	95.0%	76.8%	58.5%	65.3%
245			100.0%	99.2%	94.3%	83.0%	82.6%
250				97.6%	98.8%	95.3%	94.6%
255				100.0%	100.0%	97.9%	100.0%
260						100.0%	

Table 26 - Proportion of Students Passing the NMSBA Language Arts Assessment Based on Prior Fall RIT Mathematics Score

	Mathematics						
	3	4	5	6	7	8	9
160	0.0%						
165	7.4%						
170	6.5%						
175	9.6%	0.0%					
180	16.4%	2.5%	0.0%				
185	29.5%	3.2%	1.8%				
190	54.1%	8.3%	0.9%	0.0%			
195	71.5%	24.8%	0.9%	0.7%			
200	90.2%	42.6%	5.8%	0.9%	0.0%		
205	97.2%	73.3%	14.2%	0.3%	0.5%	0.0%	0.0%
210	98.8%	89.9%	34.2%	2.6%	1.2%	0.5%	1.4%
215	100.0%	97.5%	60.2%	12.3%	2.6%	1.4%	2.3%
220		100.0%	82.3%	36.0%	11.2%	3.7%	3.8%
225			97.8%	71.2%	29.0%	13.2%	11.5%
230			98.5%	87.5%	57.1%	32.0%	31.0%
235			100.0%	97.0%	78.6%	60.8%	50.5%
240				98.3%	91.2%	84.6%	82.9%
245				100.0%	100.0%	97.0%	94.6%
250						100.0%	100.0%

Figure 3 – Percent of Students Passing Language Arts NMSBA by Spring Reading RIT Performance Range

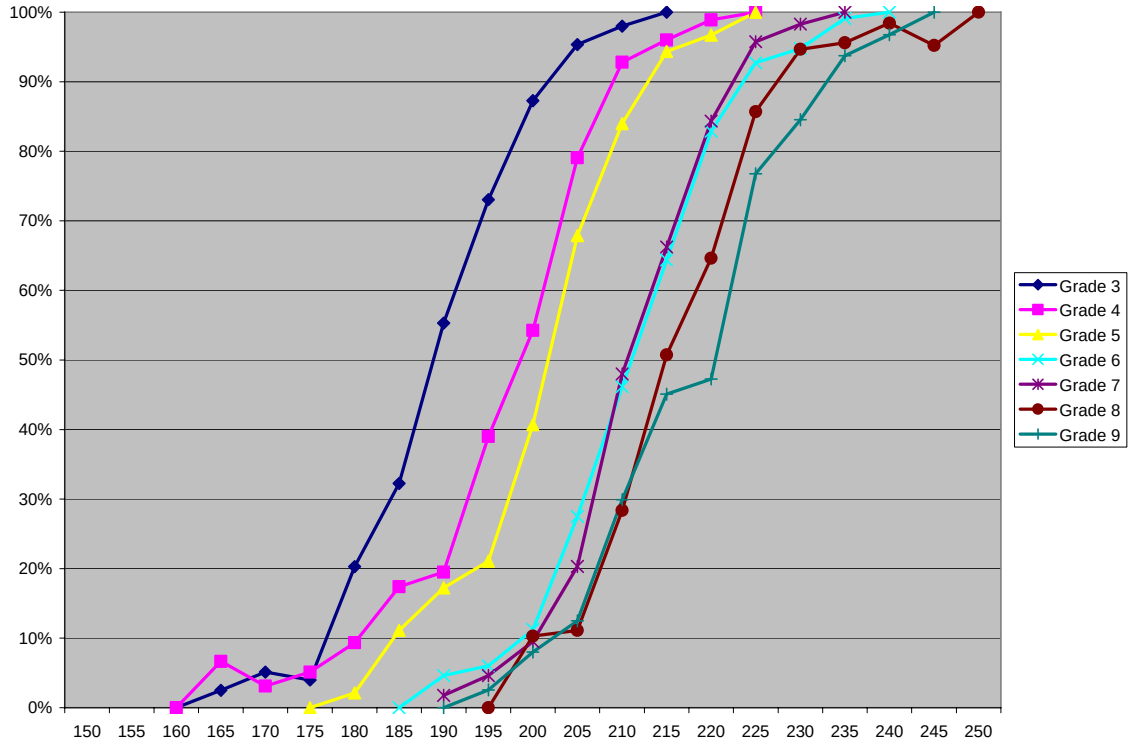


Figure 4 – Percent of Students Passing Language Arts NMSBA by prior Fall RIT Performance Range

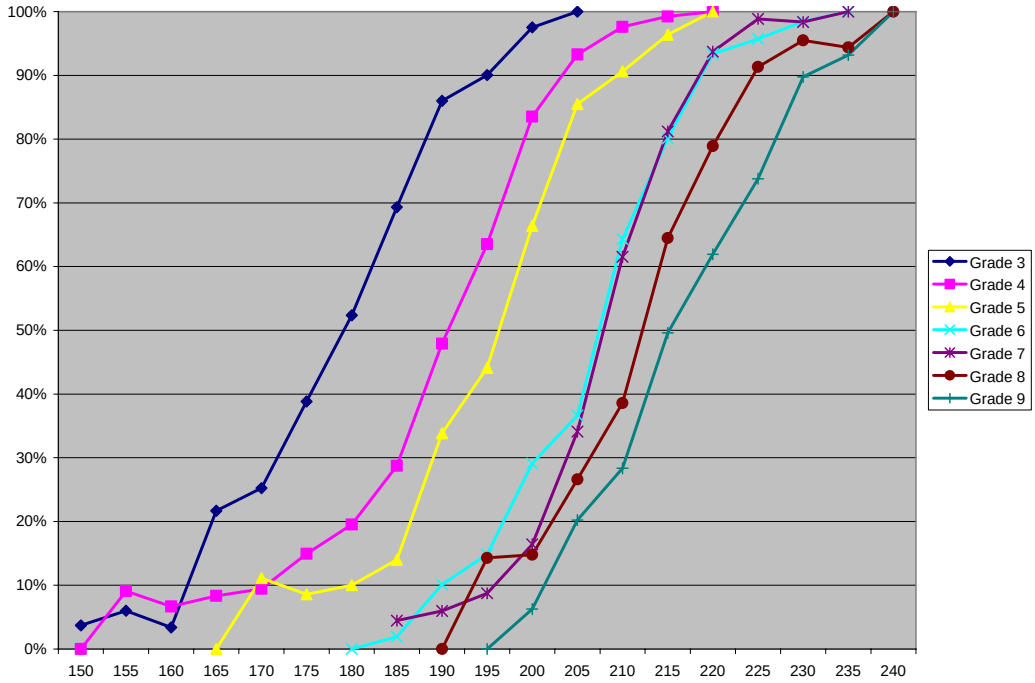


Figure 5 – Percent of Students Passing Language Arts NMSBA by Spring Language Usage RIT Performance Range

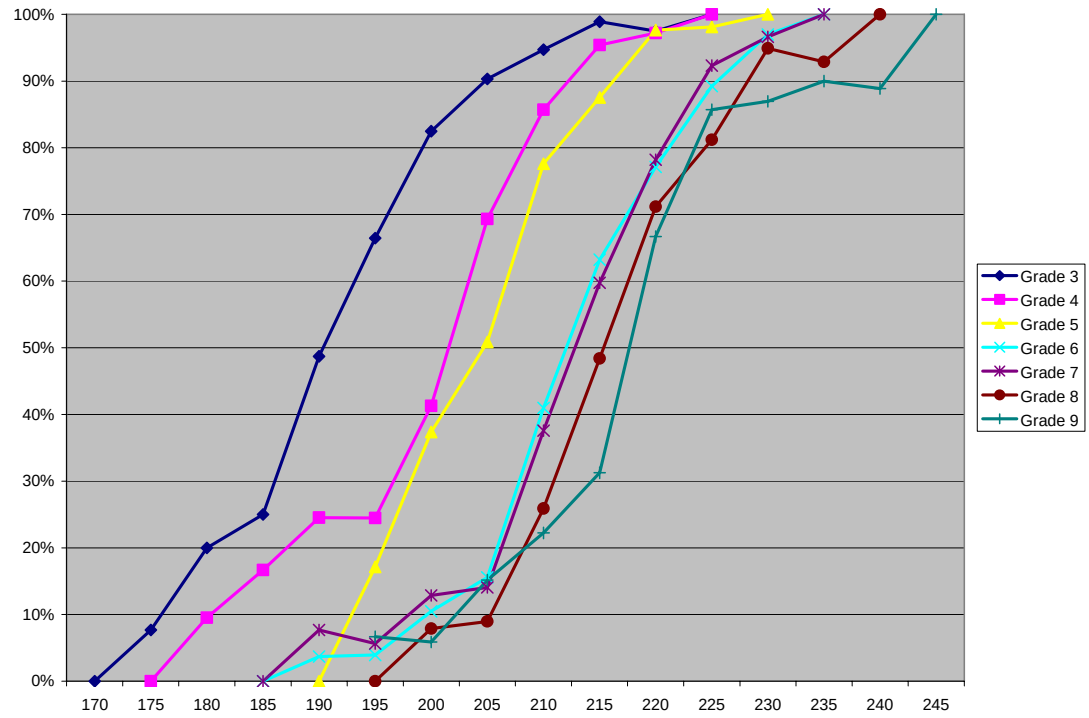


Figure 6 – Percent of Students Passing Language Arts NMSBA by Prior Fall RIT Performance Range

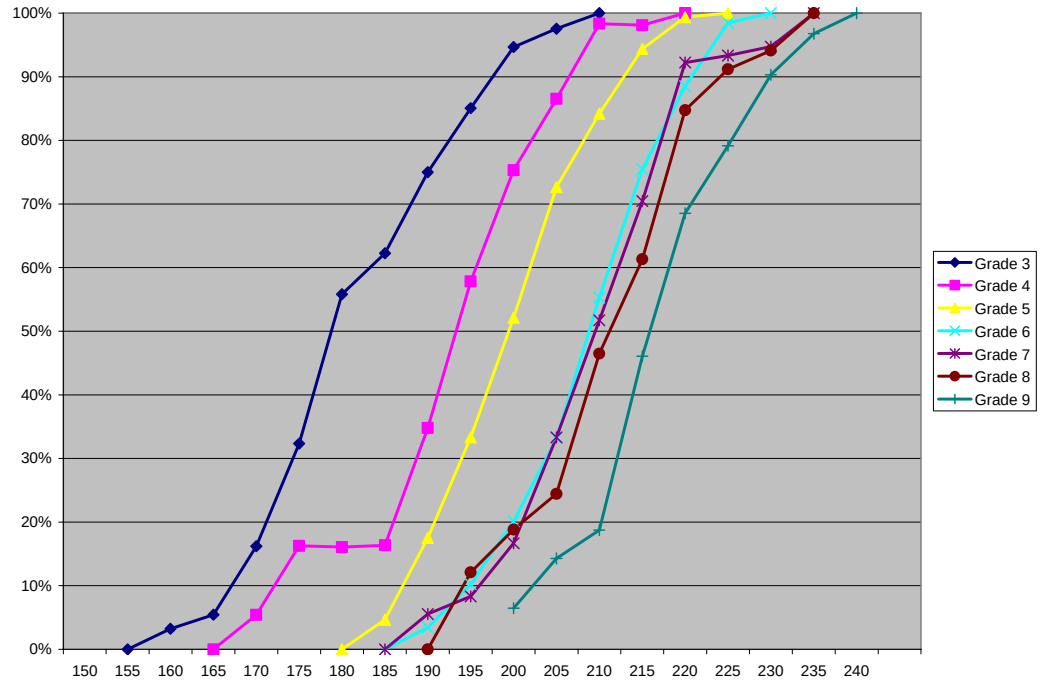


Figure 7 – Percent of Students Passing Mathematics NMSBA by Spring Mathematics RIT Performance Range

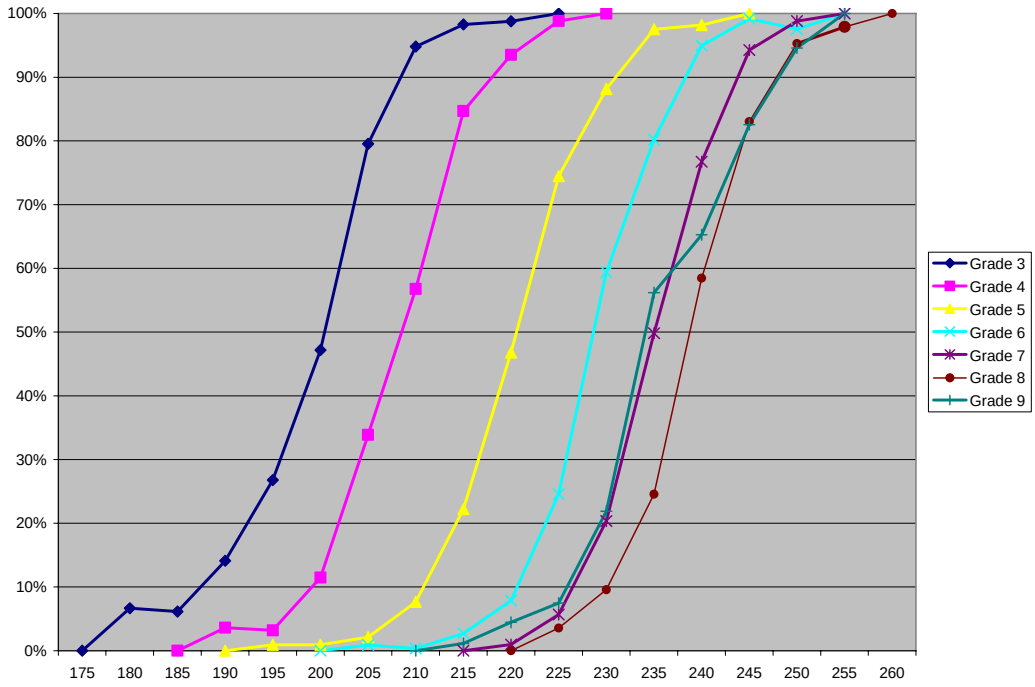
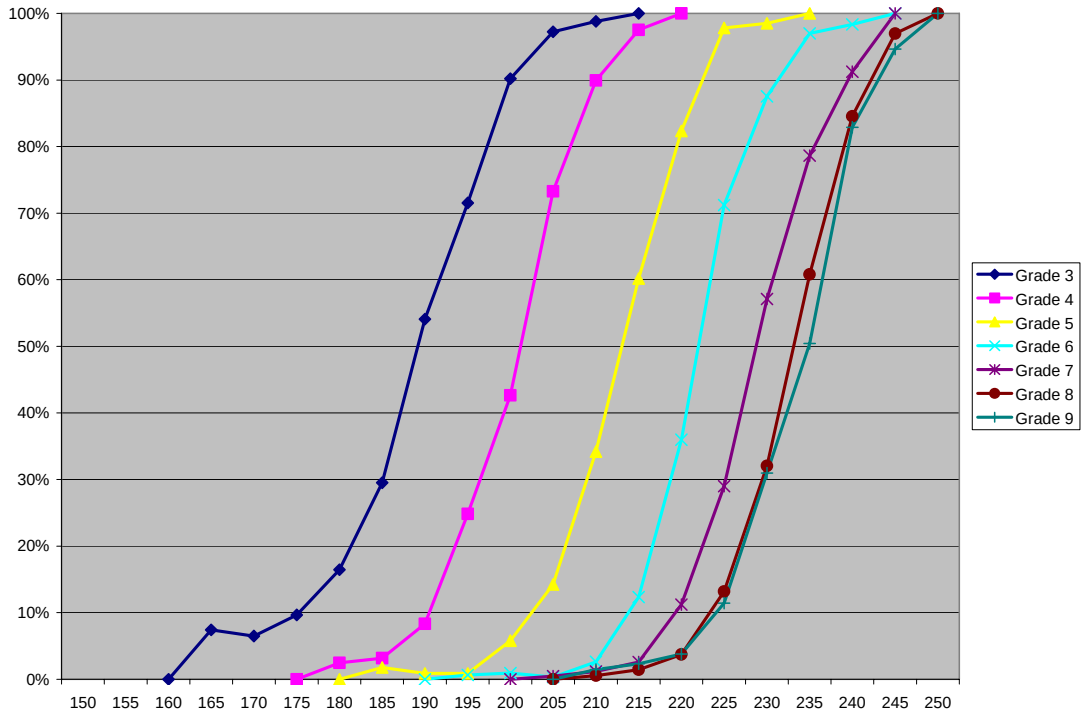


Figure 8 – Percent of Students Passing Mathematics NMSBA by Prior Fall RIT Performance Range



Comparing the NMSBA 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 NMSBA 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.

www.nwea.org/research/national.asp

Summary and Conclusions

This study investigated the relationship between the scales used for the NMSBA 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 NMSBA performance levels in those subjects. Test records for more than 17,000 students were included in this study.

Three methods generated an estimate of RIT cut scores that could be used to project NMSBA performance levels. Rasch SOS methods generally produced the most accurate cut score estimates. Accuracy of predicting NMSBA 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 district 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 these 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 provide useful information to help New Mexico educators use NWEA assessments to better inform, plan, and deliver student instruction. Good information, when matched with the professionalism and commitment of our New Mexico colleagues, will assure that all students have 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.