Aligning the NWEA RIT Scale with the California Standards Test (CST)

April 2004

John Cronin

A technical report from the Northwest Evaluation Association

NWEA
Northwest Evaluation Association
Partnering to help all kids learn

NWEA
Northwest Evaluation Association
Partnering to help all kids learn

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

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

Aligning the NWEA RIT Scale with the California Standards Test (CST)

April, 2004
John Cronin, Ph.D. – Northwest Evaluation Association

Each year, California students participate in testing as part of the state's assessment program. Students in grades 2 through 8 take tests that assess reading/writing skills 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 California Department of Education has developed scales that are used to assign students to one of five performance levels on the state's assessments. These are, from the lowest cut score to the highest: *far below basic, below basic, basic, proficient,* **and** *advanced***. For purposes of NCLB, the** *proficient* **level is considered the level that represents satisfactory performance.**

Many students who attend school in California also take paper or computerized-adaptive tests developed in cooperation with the Northwest Evaluation Association (NWEA). These 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 level. They are also used to monitor academic growth of students and cohorts. Districts choose whether to include these assessments in their local assessment programs. They are not state mandated.

The versions of NWEA tests in use in California have been specifically aligned to match the content of local and California state curriculum standards. Because of this, we believe there is a good match in content between the NWEA tests and the curriculum standards being used in California.

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 the curriculum alignment. The current study is an expansion of a preliminary study of alignment of the California Standards Tests (CST) that was performed using data from one California school system in June 2003. It is one of an ongoing series of studies that are being conducted to identify the relationships between NWEA tests and state-mandated assessments. Studies of assessments in sixteen states have now been completed.

The primary questions addressed in this study are:

- **To what extent do the same subject scores for the NWEA test correlate to the content-similar subjects on the CST?**
- **What fall and spring RIT scores correspond to various performance levels on the CST tests?**
- **How well can** *proficient* **performance on the California assessments be predicted from fall and spring RIT scores?**

Method

Participating School Systems

An e-mail solicitation was sent in January, 2004 to all California school systems who had two or more seasons of experience with NWEA testing prior to spring 2003 in order to secure participants for the study. Based on the response from this solicitation, fall 2002 and spring 2003 CST and NWEA student assessment records in reading, language usage and mathematics were collected from six school systems. These were the Capistrano Unified, Escondido Union, Gilroy Unified, Lake Elsinore Unified, and Visalia Unified school systems. Hawthorne School District supplied CST and NWEA data for their spring 2003 testing season.

Data Preparation

For purposes of studying NWEA test alignment with the CST, 2nd through 8th grade student test records from fall 2002 and spring 2003 NWEA assessments were matched with the 2003 CST assessments using district assigned student ID numbers. Because NWEA offers assessments in both reading and language usage, the NWEA records were separately matched to the California CST English Language Assessment. Matched records were then screened to remove invalid scores. Table I shows the number of matched student records included in the analysis.

Table 1
Reading and Mathematics Tests Included by Grade

	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Fall Reading	4983	8503	8922	8928	9192	9138	8257
Spring Reading	10348	10582	10871	10694	10610	10637	9688
Fall Language	3278	8486	8839	8902	9099	9242	8349
Spring Language	9402	9376	9711	9686	9723	9927	8948
Fall Mathematics	5096	8644	9023	9042	9157	9086	8087
Spring Mathematics	10686	10726	11032	10822	10840	10999	9971

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

Linking CST scores to the RIT scales. **Fall and spring scores on the RIT scale were linked separately to the appropriate scale on the CST. Three methods of estimating cut scores for CST levels were used. The most straightforward was simple linear regression (CST_{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 (CST_{pred}=a(RIT^2) + b(RIT) + c). For each of these methods, the RIT score was determined by substituting the appropriate CST score for CST_{pred} and solving the equation for RIT.**

A fixed-parameter Rasch model was also used to estimate RIT cut scores. In this method, the CST 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 CST 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 CST. We make the estimates from this level in order to maintain consistency with prior studies of state test alignment, which make comparisons based on the NCLB reported performance level. This allows us to make accurate comparisons of our alignment with different state tests.**

The predictions of CST 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 CST 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 CST levels based on observed RIT scores. The predictions of CST level performance were compared to observed performance in 5 X 5 contingency tables. Once again a prediction index score was generated to provide an estimate of accuracy.

Content Validity

Formal comparisons of the content of NWEA and California tests were not conducted for purposes of this study. The standards used to construct the NWEA Assessments were the same as those used for the California assessments. Both NWEA assessments and the California assessments include multiple-choice items. The CST also includes short answer and extended 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 reasonably accurate predictions of performance levels.

Results

Descriptive Statistics

Tables 2 through 4 review descriptive statistics for the CST and NWEA assessments. The median RIT scores for this sample are generally near or slightly above the NWEA norm in language usage and mathematics. They are slightly below the NWEA norm in reading. Relative to the CST, average scores are generally near to or above the norm in both English/Language Arts and mathematics.

Alignment studies require data that adequately represents the range of the scales being measured. In this case, we concluded from the descriptive statistics that the sample reflected a reasonably representative population. In addition, the population of students performing near the standards was large and should produce robust predictions of performance near the proficiency standard. We were concerned about the number of students who might perform at the *far below basic* **level of performance, since there seemed to be relatively small numbers of these students in the sample population. No other state that we have studied assigns a similar designation.**

Table 2
Means, Standard Deviations, and Medians for the CST and NWEA assessments - Reading

	ELA matched to fall			Fall NWEA Reading			ELA matched to spring			Spring NWEA Reading		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Grade 2	354.67	357	52.87	176.50	178	16.27	340.01	341	54.28	184.20	185	16.55
Grade 3	335.51	335	60.54	187.32	189	16.68	330.70	331	60.49	195.58	198	16.47
Grade 4	345.81	346	50.14	197.02	199	16.57	342.61	340	49.58	202.80	205	16.27
Grade 5	336.84	337	47.10	204.38	206	16.67	334.22	334	46.23	208.93	211	16.39
Grade 6	340.07	338	51.77	208.84	211	16.37	335.80	335	51.61	212.62	215	16.79
Grade 7	338.71	339	51.32	214.06	216	15.92	334.56	333	51.51	216.67	219	16.83
Grade 8	331.95	333	49.46	217.47	219	16.26	327.19	327	49.79	220.44	223	16.92

Table 3
Means, Standard Deviations, and Medians for the CST and NWEA assessments – Language Usage

	ELA matched to fall			Fall NWEA Language			ELA matched to spring			Spring NWEA Language		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Grade 2	347.36	349	53.93	177.43	178	14.59	341.44	341	55.29	189.12	191	15.99
Grade 3	335.13	335	60.70	190.60	193	16.10	332.99	331	61.25	199.53	201	16.28
Grade 4	345.87	346	50.13	200.71	203	15.68	344.57	343	50.37	206.32	209	15.60
Grade 5	336.95	337	47.04	206.98	209	15.41	335.80	334	47.09	211.91	214	15.02
Grade 6	339.36	338	51.82	211.36	214	15.15	337.89	338	51.95	214.97	217	14.94
Grade 7	338.44	339	51.44	215.49	218	14.31	336.84	336	51.76	218.51	220	14.44
Grade 8	331.65	333	49.44	218.18	220	14.31	330.30	330	49.69	220.99	223	14.53

Table 4
Means, Standard Deviations, and Medians for the CST and NWEA assessments - Mathematics

	CST Math matched to fall			Fall NWEA Math			CST Math matched to spring			Spring NWEA Math		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Grade 2	338.45	386	75.17	177.02	178	10.38	339.61	341	54.81	188.03	189	13.41
Grade 3	335.74	352	73.23	189.01	190	13.21	330.43	331	60.74	200.64	202	14.11
Grade 4	349.21	348	66.43	200.89	203	13.51	342.52	340	49.71	209.33	210	15.12
Grade 5	335.81	324	74.57	209.53	210	15.09	334.30	334	46.44	217.63	218	16.77
Grade 6	337.26	329	62.51	215.96	217	16.78	335.61	335	51.73	222.05	223	18.73
Grade 7	330.73	323	57.48	223.25	224	17.84	334.30	333	51.59	227.89	229	20.02
Grade 8	329.80	326	60.58	228.79	230	18.82	326.90	327	49.69	232.82	234	21.01

Pearson correlations

Table 5 shows the results of this analysis for each grade. Concurrent validity was tested by examining same subject Pearson correlations between the NWEA and the CST. Same subject correlations were very high. In reading and language arts, all coefficients between the CST and NWEA tests were above .81, with the single exception of the fall grade 2 reading and language tests (r=.76 for reading and r=.77 for language). In mathematics correlation coefficient generally ranged between .74 and .85. Once again the fall grade 2 coefficient for was substantially lower than those for the other tests (r=.67). In the upper grades, reading assessments correlated slightly more closely with the ELA portion of the CST, while language usage correlated slightly more closely at the lower grades.

The results suggest that the NWEA tests were generally measuring the same constructs as the CST. We expected spring NWEA tests to correlate more closely with the CST than the tests administered in the prior fall. This was the case in all grades except grade 8. The lower grade 2 correlations were not surprising. Many 2nd graders in the NWEA test population are taking multiple-choice tests for the first time in fall of second grade and standardized tests on the whole do not show the same consistency with second graders as they do in other grades.

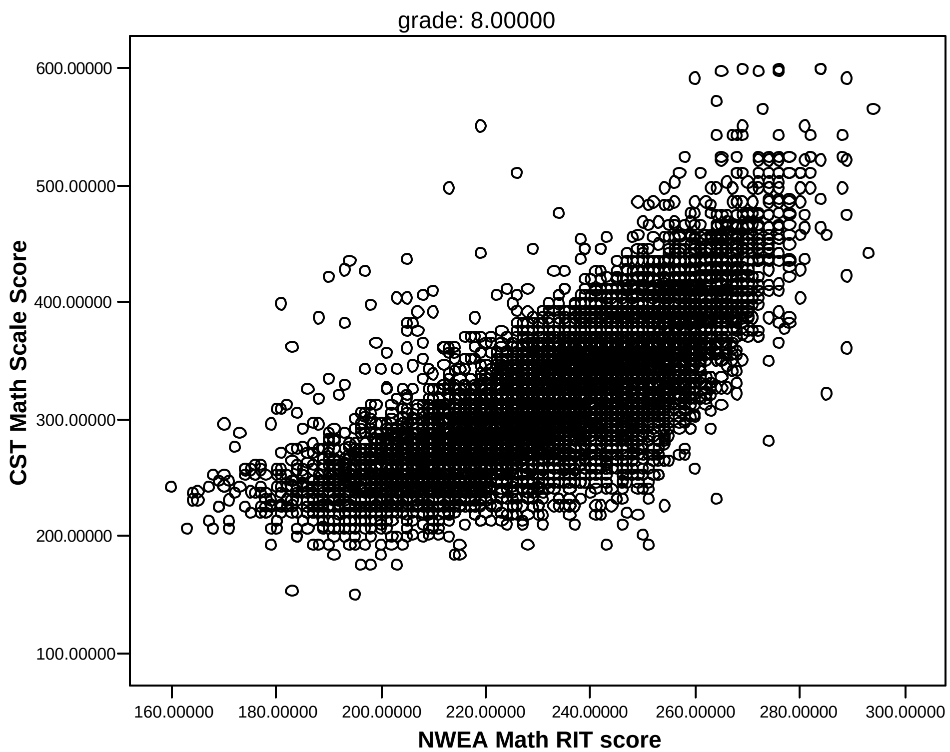
Discriminant validity was tested by examining same subject Pearson correlations next to correlations for the alternate subject on the state assessment. In particular, we tested the NWEA and CST math tests against the California ELA Standards Test. We tested the NWEA reading and language usage tests and the Calfornia ELA tests against the CST Math. In all instances the same subject correlations were higher than correlations against the alternate subject, leading us to conclude that these assessments were more likely to be testing similar constructs than dissimilar.

Table 5
Pearson Correlations for CST and NWEA assessments by Subject

Grade 2								
Assessment	Assessment							
	CST ELA	NWEA Reading		NWEA Language		CST Math	NWEA Math	
		Fall	Spring	Fall	Spring		Fall	Spring
CST ELA	1.000	.761	.810	.770	.827	.760	.688	.750
CST Math	.760	.616	.669	.616	.698	1.000	.670	.752
Grade 3								
Assessment	Assessment							
	CST ELA	NWEA Reading		NWEA Language		CST Math	NWEA Math	
		Fall	Spring	Fall	Spring		Fall	Spring
CST ELA	1.000	.812	.837	.821	.845	.798	.745	.778
CST Math	.728	.682	.728	.705	.751	1.00	.756	.818
Grade 4								
Assessment	Assessment							
	CST ELA	NWEA Reading		NWEA Language		CST Math	NWEA Math	
		Fall	Spring	Fall	Spring		Fall	Spring
CST ELA	1.000	.828	.833	.822	.811	.782	.759	.788
CST Math	.782	.700	.715	.715	.710	1.000	.788	.833
Grade 5								
Assessment	Assessment							
	CST ELA	NWEA Reading		NWEA Language		CST Math	NWEA Math	
		Fall	Spring	Fall	Spring		Fall	Spring
CST ELA	1.000	.826	.817	.811	.812	.762	.767	.775
CST Math	.762	.700	.701	.710	.718	1.00	.811	.845
Grade 6								
Assessment	Assessment							
	CST ELA	NWEA Reading		NWEA Language		CST Math	NWEA Math	
		Fall	Spring	Fall	Spring		Fall	Spring
CST ELA	1.000	.841	.834	.818	.814	.798	.784	.792
CST Math	.798	.730	.729	.724	.725	1.000	.839	.855
Grade 7								
Assessment	Assessment							
	CST ELA	NWEA Reading		NWEA Language		CST Math	NWEA Math	
		Fall	Spring	Fall	Spring		Fall	Spring
CST ELA	1.000	.832	.831	.807	.807	.781	.787	.784
CST Math	.781	.708	.706	.708	.710	1.000	.851	.851
Grade 8								
Assessment	Assessment							
	CST ELA	NWEA Reading		NWEA Language		CST Math	NWEA Math	
		Fall	Spring	Fall	Spring		Fall	Spring
CST ELA	1.000	.815	.800	.792	.783	.707	.767	.746
CST Math	.707	.658	.666	.672	.657	1.000	.784	.772

Analysis of scatterplots suggested that relationships between most NWEA tests and their CST counterpart were strongly curvilinear with a pronounced floor effect at some grades. Figure 1 provides an example from the 8th grade reading sample that illustrates both the scale relationships and the evidence of some breakdown in correlation near the bottom of the CST Scale. Note how the correlation between the two tests flattens for students performing below 300 on the CST. Note also that large numbers of students achieving below 300 on the CST test achieve a widerange of scores (between 160 and 220 RIT) on the corresponding NWEA exam. One possible explanation for this is that the NWEA test, because it is adaptive as opposed to single form, has the capacity to adjust the difficulty to the test to enable more accurate measurement at the low end of performance.

Figure 1 – Scatterplot depicting Grade 8 NWEA math RIT against the Grade 8 CST math scale score



Linking CST performance level cut scores to the RIT scale

The primary purpose of this study was to estimate the fall and spring RIT scale scores that most closely correspond to the cut scores for the different performance levels on the CST. 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.

Tables 6 and 7 shows several estimations of the Fall and Spring RIT scores that correspond to the cut scores for the various performance levels on the CST scales. As a rule the three methodologies came to very similar estimates of the cut score for each of the performance levels. Estimates of the two lowest (far below basic and below basic) and highest (advanced) cut score varied more, in part because far fewer students perform at these levels and in part because of the non-linear nature of the relationship. In some grades, calibration of the below and far below basic estimates was inconsistent. For example, second order regression estimated a far below basic/below basic cut score for fall of grade 4 in language usage and grade 6 in mathematics (see table 7) that was lower than the respective prior year's estimates. In some cases this

may have occurred because the estimated fall cut scores the lowest level of the CST were close to the lowest valid scores on the NWEA scale.

Table 6
Estimated points on the RIT scale for SPRING that equate to the minimum scores (rounded) for performance levels on the CST

	Linear Regression				Second-order Regression				Rasch Status-on-Standard			
Reading	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv
Grade 2	157	170	188	208	154	170	189	206	159	173	188	202
Grade 3	173	186	202	219	172	188	203	216	176	189	202	214
Grade 4	174	186	206	222	166	188	208	220	174	191	208	218
Grade 5	183	194	216	235	179	197	217	229	185	200	215	228
Grade 6	188	199	218	235	188	203	220	232	190	204	219	230
Grade 7	190	204	223	242	188	207	225	238	193	208	223	235
Grade 8	196	209	230	248	194	212	230	242	201	214	229	240
	Linear Regression				Second-order Regression				Rasch Status-on-Standard			
Language Usage	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv
Grade 2	163	175	192	210	161	175	193	209	164	178	193	205
Grade 3	177	189	205	221	175	191	206	218	176	193	205	217
Grade 4	179	190	208	224	171	192	210	222	177	196	210	220
Grade 5	188	198	218	235	185	201	218	230	191	204	218	228
Grade 6	192	202	219	235	191	205	221	231	195	207	220	229
Grade 7	195	206	223	240	192	209	225	237	199	210	223	234
Grade 8	198	210	228	245	198	214	230	241	203	215	227	237
	Linear Regression				Second-order Regression				Rasch Status-on-Standard			
Mathematics	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv
Grade 2	158	173	185	199	155	173	185	199	162	177	185	196
Grade 3	175	189	202	216	173	190	202	215	176	193	202	212
Grade 4	182	197	212	225	180	198	212	225	184	201	211	223
Grade 5	197	209	224	245	194	211	224	241	198	213	224	239
Grade 6	194	211	231	252	189	214	231	248	192	215	229	245
Grade 7	197	217	239	265	188	219	239	259	200	221	238	257
Grade 8	202	223	246	273	197	225	246	267	208	227	244	264

Table 7
Estimated points on the RIT scale for the FALL PRIOR to CST testing that equate to the minimum scores (rounded) for performance levels on the CST

	Linear Regression				Second-order Regression				Rasch Status-on-Standard			
Reading	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv
Grade 2	141	155	175	196	137	154	175	195	149	157	176	191
Grade 3	162	175	192	210	160	178	194	208	163	179	193	205
Grade 4	166	179	199	216	155	181	201	214	163	184	201	211
Grade 5	177	189	210	229	172	191	211	224	179	194	210	223
Grade 6	183	194	213	229	181	197	215	227	185	199	214	225
Grade 7	187	200	218	237	184	203	220	234	190	204	218	231
Grade 8	192	205	225	242	189	207	225	237	196	210	224	236
	Linear Regression				Second-order Regression				Rasch Status-on-Standard			
Language Usage	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv
Grade 2	149	160	178	196	149	161	179	197	156	163	178	193
Grade 3	166	179	195	212	164	182	197	211	167	183	196	207
Grade 4	171	183	202	218	161	186	205	217	169	189	204	215
Grade 5	181	192	212	230	178	195	214	225	184	198	212	224
Grade 6	186	197	215	230	187	201	217	228	190	203	217	226
Grade 7	190	202	219	236	187	205	221	233	195	207	220	230
Grade 8	195	206	224	240	195	210	226	237	200	212	224	234
	Linear Regression				Second-order Regression				Rasch Status-on-Standard			
Mathematics	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv	Below	Basic	Prof	Adv
Grade 2	146	159	170	182	138	158	170	183	153	164	172	180
Grade 3	162	176	188	203	151	177	190	203	163	180	189	200
Grade 4	174	188	201	215	168	189	202	213	176	193	203	212
Grade 5	189	201	213	233	188	203	215	230	191	206	215	227
Grade 6	189	204	220	241	183	207	223	238	188	208	221	235
Grade 7	197	212	231	254	193	215	233	250	197	215	231	248
Grade 8	200	217	237	262	196	218	237	257	204	221	236	256

Predicting CST pass-fail status from RIT scores

Once the spring and fall cut scores were estimated from the three methods, we evaluated each possible cut score to determine how accurately it predicted students' actual performance on the corresponding CST assessment. The most accurate method of prediction was generally used to derive the best estimate of RIT cut scores that equate to the different CST performance levels. Once again a *prediction index* **statistic (described on page 3) scored the accuracy of prediction.**

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 CST 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 passing score for a student who actually fails 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 California assessment that corresponds to their grade level.**

Tables 8, 9, and 10 summarize the results. When using spring RIT scores, all methods accurately predicted proficiency status with average rate of 84% or better in English/Language Arts and 83% for mathematics. When using fall RIT scores the accuracy rate dropped only slightly, with all methods accurately predicting pass/fail status with an accuracy rate greater than 83% for English/Language Arts and 82% for mathematics. Second-order regression methods were consistently more accurate at predicting proficiency status than the other methods.

Table 8
Accuracy of reading RIT scores in predicting CST proficiency status – ELA

	Fall				Spring			
Grade 2	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	175	79.34%	11.51%	.855	188	82.11%	10.02%	.878
Second Order	175	79.34%	11.51%	.855	189	82.46%	7.67%	.907
Rasch	176	79.52%	10.44%	.869	188	82.11%	10.02%	.878
Grade 3	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	192	82.93%	10.61%	.872	202	85.16%	8.65%	.898
Second Order	194	83.65%	7.57%	.909	203	85.17%	6.04%	.929
Rasch	193	83.39%	9.24%	.889	202	85.16%	8.65%	.898
Grade 4	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	199	83.29%	11.73%	.859	206	84.60%	10.49%	.876
Second Order	201	84.24%	8.66%	.897	208	85.20%	7.55%	.911
Rasch	201	84.24%	8.66%	.897	208	85.20%	7.55%	.911
Grade 5	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	210	84.93%	7.70%	.909	216	84.12%	7.74%	.908
Second Order	211	85.17%	6.22%	.927	217	83.83%	6.33%	.921*
Rasch	210	84.93%	7.70%	.909	215	84.01%	9.36%	.889
Grade 6	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	213	85.65%	8.61%	.899	218	86.03%	8.62%	.900
Second Order	215	85.37%	6.13%	.928	220	86.23%	4.90%	.943
Rasch	214	85.62%	7.34%	.914	219	86.51%	7.02%	.919
Grade 7	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	218	84.80%	9.16%	.892	223	85.44%	7.81%	.909
Second Order	220	85.53%	5.77%	.933	225	85.14%	5.03%	.941
Rasch	218	84.80%	9.16%	.892	223	85.44%	7.81%	.909
Grade 8	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	225	85.65%	6.71%	.922	230	85.41%	7.01%	.918
Second Order	225	85.65%	6.71%	.922	230	85.41%	7.01%	.918
Rasch	224	85.48%	8.04%	.906	229	85.57%	8.02%	.907

Table 9
Accuracy of language usage RIT scores in predicting CST proficiency status – ELA

	Fall				Spring			
Grade 2	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	178	81.15%	10.22%	.874	192	82.02%	11.67%	.858
Second Order	179	81.42%	8.91%	.891	193	82.33%	10.23%	.876
Rasch	178	81.15%	10.22%	.874	193	82.33%	10.23%	.876
Grade 3	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	195	83.00%	10.72%	.871	205	85.24%	8.53%	.900
Second Order	197	83.60%	7.97%	.905	206	85.31%	7.13%	.916
Rasch	196	83.40%	9.30%	.888	205	85.24%	8.53%	.900
Grade 4	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	202	82.59%	12.37%	.850	208	83.42%	12.52%	.850
Second Order	205	83.43%	7.53%	.910	210	84.37%	9.29%	.890
Rasch	204	83.13%	9.43%	.887	210	84.37%	9.29%	.890
Grade 5	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	212	83.96%	9.28%	.889	218	83.70%	8.72%	.896
Second Order	214	83.86%	6.83%	.919	218	83.70%	8.72%	.896
Rasch	212	83.96%	9.28%	.889	218	83.70%	8.72%	.896
Grade 6	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	215	83.62%	11.06%	.868	219	83.03%	11.19%	.865
Second Order	217	84.24%	7.90%	.906	221	84.13%	7.42%	.912
Rasch	217	84.24%	7.90%	.906	220	83.57%	9.35%	.888
Grade 7	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	219	83.41%	10.35%	.876	223	83.15%	9.76%	.883
Second Order	221	83.48%	7.10%	.915	225	82.87%	6.69%	.919
Rasch	220	83.51%	8.62%	.897	223	83.15%	9.76%	.883
Grade 8	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	224	83.47%	9.21%	.890	228	83.50%	7.77%	.907
Second Order	226	83.71%	5.77%	.931	230	83.12%	4.80%	.942
Rasch	224	83.47%	9.21%	.890	227	83.45%	9.37%	.888

Table 10
Accuracy of mathematics RIT scores in predicting CST proficiency status – mathematics

	Fall				*Spring*			
Grade 2	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	170	78.69%	14.38%	.817	184	80.34%	11.69%	.854
Second Order	170	78.69%	14.38%	.872	185	80.49%	10.36%	.871
Rasch	172	78.57%	11.48%	.854	185	80.49%	10.36%	.871
Grade 3	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	188	78.24%	13.57%	.827	201	82.67%	10.33%	.875
Second Order	190	78.81%	10.08%	.872	202	82.61%	8.87%	.893
Rasch	189	78.75%	11.79%	.850	202	82.61%	8.87%	.893
Grade 4	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	201	80.02%	13.62%	.829	211	83.61%	8.46%	.899
Second Order	202	80.56%	11.70%	.855	212	83.63%	7.11%	.915
Rasch	203	80.51%	10.35%	.871	211	83.61%	8.46%	.899
Grade 5	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	213	83.53%	11.47%	.863	223	86.45%	8.84%	.898
Second Order	215	84.75%	7.85%	.907	224	86.88%	7.25%	.917
Rasch	215	84.75%	7.85%	.907	224	86.88%	7.25%	.917
Grade 6	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	220	85.19%	9.22%	.892	229	87.63%	6.86%	.922
Second Order	223	85.56%	4.77%	.944	231	87.81%	5.01%	.943
Rasch	221	85.72%	7.69%	.910	229	87.63%	6.86%	.922
Grade 7	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	231	86.78%	7.43%	.914	238	87.76%	7.25%	.917
Second Order	233	86.84%	4.99%	.943	239	88.07%	6.22%	.929
Rasch	231	86.78%	7.43%	.914	238	87.76%	7.25%	.917
Grade 8	Cut Score	Accuracy	Type I Error	Prediction Index	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	237	79.93%	9.60%	.880	245	81.89%	9.67%	.882
Second Order	237	79.93%	9.60%	.880	246	81.97%	8.85%	.892
Rasch	236	80.04%	10.71%	.866	244	81.85%	10.47%	.872

Table 11 summarizes the accuracy of proficiency prediction for this study relative to other state alignment studies. Prediction index scores for California are near average in reading and slightly above average for the language usage test (relative to predicting results in English/Language Arts). Prediction index scores for mathematics were lower than the average for prior state alignment studies that we have conducted. The table suggests that little accuracy was lost when we used the fall assessment to predict state assessment proficiency status. Prediction index averages for the fall assessment were only slightly lower than spring.

One factor affecting accuracy of proficiency status prediction in California was the state's testing of second grade students. California is the only state we have studied to date that administers their state assessment in second grade. We expected that the accuracy of prediction for second graders would be somewhat lower than third graders and the results reflected our expectations.

Despite this fact, the rates of correct prediction are easily high enough to provide useful information to educators who are planning instruction to ensure all students perform at a level that meets the standards.

Table 11
Prediction Indices (Based on Proficiency Status)
for Previous NWEA State Alignment Studies

State	Reading	State	Lanaguage	State	Math
Texas	.974	Texas	.968	Texas	.970
Washington	.971	California (spr)	.913	Wyoming	.961
Minnesota	.944	California (fall)	.913	Colorado '01	.957
Pennsylvania	.935	Indiana '01	.907	Washington	.949
Wyoming	.931	Colorado '03	.903	Illinois	.946
Colorado '03	.931	Indiana '03	.894	Colorado '03	.943
Illinois	.928	Arizona	.874	South Carolina	.943
California (spr)	.925			Minnesota	.936
California (fall)	.914			Washington	.936
Arizona	.912			Pennsylvania	.926
Colorado '01	.910			Arizona	.919
Nevada	.902			California (spr)	.910
South Carolina	.902			Indiana '01	.899
Indiana '01	.902			California (fall)	.895
Indiana '03	.900			Nevada	.866
Washington	.886			Indiana '03	.860

* Texas results were generated by a study of over 1,000 per grade from a single school district.

Predicting CST Performance Levels from RIT Scores

The CST reports five levels of performance. Four cut scores are set to define these five levels. Analyzing the capacity of RIT scores to predict students' CST performance levels can help educators triangulate information about student performance on their state test, assuring that instructional plans and interventions are adequately reinforced by data. Predictions of performance level are not as accurate as the predictions of proficiency status. This is true in part because tests vary in their ability to measure students at the highest and lowest performance levels. In the case of the California state assessment, predictions of performance level were influenced by the high number of performance levels used for the test (California and Minnesota are the only states we have studied that use five) and the small number of students scoring in the lowest category (*far below basic***) on the state assessment.**

When predicting performance levels, a case is identified as accurate when the performance level assigned by the CST and RIT score are the same. A Type I error occurs when the RIT score assigns a performance level that is higher than the student actually achieved on the state test. For example, if the RIT score projects an *advanced* **performance for the student and the CST result is** *proficient***, we declare the case a Type I error because the RIT score overestimated performance.**

In addition to assessing the rate of correct prediction, we also assessed accuracy by evaluating the success with which the projected RIT cut scores for the highest and lowest performance levels identified students in these two categories. For example, if 1000 grade 3 students performed at the advanced level in a subject and a RIT score identified 600 students as advanced, then we would say the RIT score was successful at finding 60% of the advanced students. For the highest and lowest performance level, we used this methodology to assign the cut score that would best predict the *far below basic* **and** *advanced* **performance levels.**

Tables 12, 13 and 14 summarize these results.

Table 12
Accuracy of the RIT scale in predicting CST performance level – reading

	Fall					Spring				
Grade 2	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	39.9%	14.6%	.633	28.9%	0.0%	53.9%	23.5%	.564	36.9%	46.6%
2nd Order	40.3%	15.3%	.621	30.9%	0.0%	54.3%	22.9%	.579	52.7%	49.9%
Rasch	50.8%	26.6%	.476	65.7%	27.7%	54.3%	25.2%	.536	65.0%	55.0%
Grade 3	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	41.1%	19.0%	.539	24.4%	60.1%	56.1%	23.8%	.575	34.7%	52.5%
2nd Order	41.7%	16.8%	.596	30.6%	55.9%	57.0%	20.9%	.633	52.7%	49.9%
Rasch	53.6%	28.2%	.528	65.3%	55.9%	57.1%	23.4%	.590	63.8%	61.3%
Grade 4	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	44.7%	28.7%	.358	37.5%	43.6%	57.8%	22.7%	.607	40.6%	42.8%
2nd Order	46.1%	28.2%	.387	44.6%	17.0%	60.2%	21.2%	.648	58.7%	28.6%
Rasch	58.6%	22.3%	.620	72.0%	43.0%	59.3%	21.3%	.641	69.5%	47.7%
Grade 5	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	41.9%	36.6%	.127	21.0%	33.2%	59.1%	22.3%	.613	21.8%	42.8%
2nd Order	43.0%	38.0%	.115	47.2%	25.7%	58.2%	17.1%	.706	50.0%	31.7%
Rasch	57.8%	22.1%	.617	63.0%	47.6%	59.3%	21.3%	.641	69.5%	47.7%
Grade 6	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	48.0%	32.4%	.325	47.0%	42.2%	59.1%	22.3%	.623	36.5%	64.2%
2nd Order	49.2%	29.8%	.394	47.2%	36.5%	60.5%	17.2%	.715	57.9%	54.9%
Rasch	60.1%	20.0%	.667	69.0%	57.2%	60.7%	20.9%	.655	70.8%	58.1%
Grade 7	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	46.7%	37.3%	.202	33.4%	29.4%	58.7%	21.8%	.628	30.2%	47.0%
2nd Order	48.9%	34.0%	.305	49.0%	24.9%	60.0%	18.9%	.684	54.0%	43.2%
Rasch	58.8%	23.2%	.606	65.8%	50.9%	59.6%	21.7%	.636	66.0%	55.5%
Grade 8	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	45.4%	31.7%	.300	21.9%	33.1%	56.0%	24.3%	.566	20.7%	42.1%
2nd Order	44.4%	34.0%	.235	46.2%	27.9%	58..2%	24.5%	.579	56.4%	38.0%
Rasch	58.1%	22.8%	.608	68.7%	51.2%	57.7%	22.8%	.605	63.9%	66.0%

Table 13
Accuracy of the RIT scale in predicting CST performance level – language usage

	Fall					Spring				
Grade 2	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	39.7%	18.3%	.538	28.3%	2.4%	55.2%	25.2%	.544	45.5%	35.1%
2nd Order	38.8%	16.8%	.565	25.2%	2.4%	55.1%	24.6%	.553	45.5%	23.5%
Rasch	50.5%	26.6%	.474	60.3%	32.5%	55.0%	25.8%	.531	68.0%	39.7%
Grade 3	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	42.8%	18.7%	.562	29.2%	63.2%	57.2%	24.3%	.575	45.5%	35.1%
2nd Order	43.0%	16.7%	.611	32.8%	56.5%	57.5%	23.1%	.597	56.8%	49.8%
Rasch	53.4%	25.8%	.516	58.6%	60.4%	57.3%	25.3%	.558	69.2%	52.2%
Grade 4	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	45.8%	29.6%	.355	44.3%	44.4%	58.0%	24.6%	.577	51.6%	52.8%
2nd Order	47.1%	26.8%	.431	48.5%	16.2%	60.6%	21.2%	.650	63.7%	30.2%
Rasch	58.2%	22.1%	.620	70.8%	47.6%	59.6%	22.2%	.628	73.8%	46.5%
Grade 5	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	42.2%	37.9%	.104	19.8%	33.6%	56.0%	22.9%	.590	19.1%	41.5%
2nd Order	44.4%	36.8%	.172	47.9%	29.2%	57.9%	22.0%	.620	50.1%	45.8%
Rasch	57.1%	22.6%	.603	59.6%	47.7%	56.8%	22.5%	.603	63.3%	51.0%
Grade 6	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	48.1%	33.9%	.295	42.6%	37.8%	55.7%	24.9%	.553	32.7%	48.4%
2nd Order	49.2%	30.3%	.384	55.2%	39.9%	57.8%	20.8%	.640	50.1%	45.8%
Rasch	57.1%	21.6%	.622	65.2%	55.9%	57.5%	23.2%	.596	69.3%	58.5%
Grade 7	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	46.9%	38.5%	.179	32.5%	25.5%	56.8%	23.8%	.582	29.9%	40.5%
2nd Order	49.1%	35.1%	.286	49.9%	20.7%	57.3%	20.5%	.642	44.9%	32.2%
Rasch	56.8%	22.7%	.601	66.8%	49.2%	56.8%	23.8%	.580	63.6%	50.7%
Grade 8	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	45.7%	34.0%	.255	21.3%	29.9%	55.0%	25.0%	.545	22.7%	37.4%
2nd Order	45.7%	30.1%	.340	38.1%	29.9%	55.9%	20.1%	.640	38.1%	37.4%
Rasch	54.7%	24.2%	.558	61.1%	51.2%	55.4%	24.9%	.551	62.7%	50.2%

Table 14
Accuracy of the RIT scale in predicting CST performance level – mathematics

	Fall					Spring				
Grade 2	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	51.0%	28.8%	.434	36.8%	0.0%	51.3%	26.8%	.478	58.7%	24.5%
2nd Order	50.3%	28.0%	.442	52.9%	0.0%	51.0%	24.4%	.521	53.6%	14.9%
Rasch	39.4%	49.6%	-.260	54.1%	8.2%	52.0%	25.7%	.505	70.3%	40.4%
Grade 3	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	48.9%	29.3%	.401	45.5%	40.1%	55.2%	24.3%	.560	78.9%	55.9%
2nd Order	49.6%	27.3%	.451	58.2%	26.6%	56.1%	21.6%	.615	78.9%	60.4%
Rasch	42.9%	42.7%	.003	52.8%	37.9%	56.4%	22.2%	.607	69.5%	55.6%
Grade 4	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	51.3%	28.3%	.447	49.2%	36.3%	56.4%	25.0%	.558	73.8%	56.6%
2nd Order	52.8%	29.0%	.450	62.5%	19.8%	56.6%	22.6%	.601	76.5%	60.5%
Rasch	48.9%	26.7%	.454	65.4%	39.4%	56.2%	22.7%	.596	69.2%	52.3%
Grade 5	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	51.7%	24.3%	.530	36.8%	46.2%	56.7%	20.4%	.640	85.3%	62.9%
2nd Order	53.9%	19.6%	.637	52.9%	42.5%	58.7%	16.6%	.717	78.2%	66.9%
Rasch	46.9%	22.1%	.529	54.1%	54.7%	58.5%	19.8%	.661	68.2%	60.7%
Grade 6	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	57.7%	24.0%	.584	45.5%	41.7%	61.2%	20.5%	.665	88.4%	41.3%
2nd Order	59.6%	19.3%	.676	58.2%	26.6%	64.0%	16.7%	.739	80.6%	47.2%
Rasch	50.4%	21.8%	.567	52.8%	37.9%	63.3%	19.9%	.685	69.7%	43.5%
Grade 7	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	60.8%	22.1%	.637	43.8%	42.4%	61.4%	22.7%	.630	42.6%	34.7%
2nd Order	62.1%	19.2%	.691	64.0%	31.1%	62.8%	22.1%	.648	66.1%	15.3%
Rasch	49.6%	17.8%	.640	43.5%	43.4%	62.0%	20.6%	.667	73.7%	43.3%
Grade 8	Accuracy	Type I Error	Prediction Index	% Adv. Found	% B.B. Found	Accuracy	Type I Error	Prediction Index	% Adv. Found	% BB Found
Linear	53.7%	24.9%	.536	34.8%	40.6%	52.9%	26.5%	.499	18.7%	38.9%
2nd Order	53.4%	27.0%	.495	54.5%	31.1%	52.9%	27.2%	.479	44.5%	25.9%
Rasch	49.9%	21.9%	.562	42.7%	49.8%	52.3%	25.9%	.504	60.1%	52.9%

Table 15
Prediction index scores by performance level assignment for previous NWEA state alignment Studies

State	Reading	State	Math
Washington	.874	Washington	.928
Texas	.868	Texas	.900
Indiana	.860	Illinois	.888
Colorado	.840	Colorado	.808
Illinois	.804	Washington	.805
Nevada	.776	Indiana	.804
Pennsylvania	.770	Pennsylvania	.769
South Carolina	.757	South Carolina	.764
Arizona	.756	Arizona	.756
Washington	.698	Nevada	.742
Minnesota	.627	Minnesota	.611
California	.600	California	.565

Best estimates of CST performance level cut scores

To determine the RIT scores that best predict the cut scores for the various California performance levels we did the following:

- **For the** *proficient* **and** *basic* **RIT cut score, we selected the methodology that produced the highest overall performance index score.**
- **For the** *far below basic* **RIT score and the** *advanced* **RIT score, we selected the cut scores that correctly predicted the largest proportion of students who actually achieved these levels of performance on the CST.**

The methodology that was ultimately applied to determine cut scores is bolded in Tables 12 through 14. Tables 16 and 17 (see following page) summarize the recommended cut scores for each performance level on the CST.

Analysis of the performance level cut scores

We hope that the projected cut scores provide useful information to educators who use NWEA data to help students succeed in learning and on their state test. In addition to information that can be used to plan student programs, the study also provides a helpful external look at some important aspects of the California Standards Test. Some of these include the difficulty of the standards relative to other states, dither difficulty of the state's mathematics standards relative to the ELA standards, and the calibration of the state's standards between grades.

Table 16
Projected Minimum RIT Scores for FALL PRIOR that are Equivalent to Performance Levels on CST
(scores under the below basic cut score project to far below basic
NWEA percentile rank is in parenthses)

Grade	Reading to CST ELA				Language to CST ELA				Math			
	Below Basic	Basic	Proficient	Advanced	Below Basic	Basic	Proficient	Advanced	Below Basic	Basic	Proficient	Advanced
2	149 (2)	155 (9)	175 (43)	191 (78)	156 (2)	161 (10)	179 (48)	193 (79)	153 (2)	158 (3)	170 (24)	180 (62)
3	162 (8)	178 (23)	194 (59)	205 (86)	166 (7)	182 (24)	197 (61)	207 (87)	162 (8)	177 (15)	190 (49)	203 (87)
4	166 (4)	184 (17)	201 (53)	211 (81)	169 (4)	189 (18)	204 (55)	215 (86)	176 (4)	193 (25)	203 (57)	212 (84)
5	179 (6)	194 (20)	210 (59)	223 (91)	184 (6)	198 (21)	212 (60)	224 (92)	191 (9)	203 (31)	215 (68)	227 (92)
6	185 (6)	199 (20)	214 (56)	225 (85)	190 (6)	203 (21)	217 (61)	226 (87)	189 (5)	207 (28)	223 (70)	238 (94)
7	190 (6)	204 (16)	218 (56)	231 (89)	195 (7)	207 (23)	220 (61)	230 (90)	197 (8)	215 (35)	233 (77)	250 (97)
8	196 (8)	210 (20)	224 (62)	236 (90)	200 (8)	212 (27)	224 (64)	234 (91)	204 (10)	221 (35)	236 (61)	257 (96)

Table 17
Projected Minimum RIT Scores for SPRING that are Equivalent to Performance Levels on CST
(scores under the below basic cut score project to far below basic
NWEA percentile rank is in parenthses)

Grade	Reading to CST ELA				Language to CST ELA				Math			
	Below Basic	Basic	Proficient	Advanced	Below Basic	Basic	Proficient	Advanced	Below Basic	Basic	Proficient	Advanced
2	159 (7)	170 (16)	188 (50)	202 (83)	164 (7)	175 (20)	193 (59)	205 (86)	162 (2)	173 (11)	185 (39)	196 (74)
3	176 (10)	188 (25)	203 (61)	214 (88)	176 (9)	191 (26)	206 (66)	217 (91)	173 (3)	190 (22)	202 (56)	215 (90)
4	174 (4)	191 (18)	208 (56)	218 (82)	177 (4)	192 (16)	210 (59)	220 (86)	180 (3)	198 (21)	212 (59)	225 (89)
5	185 (6)	200 (22)	217 (65)	228 (90)	191 (7)	201 (19)	218 (65)	228 (92)	194 (8)	211 (36)	224 (69)	245 (97)
6	190 (6)	204 (20)	220 (60)	230 (86)	195 (8)	205 (20)	221 (54)	229 (80)	189 (3)	214 (32)	231 (71)	252 (96)
7	193 (5)	208 (21)	225 (64)	235 (89)	199 (7)	209 (22)	225 (68)	234 (86)	200 (7)	221 (35)	238 (71)	257 (95)
8	201 (8)	214 (24)	230 (67)	240 (91)	203 (9)	214 (26)	230 (75)	237 (91)	208 (10)	227 (35)	244 (67)	264 (95)

Comparing California proficiency standards with the estimated standards reported in other state test alignment studies

Northwest Evaluation Association tests have been aligned with the cut scores for the state proficiency test in 16 states. To get an estimate of the difficulty of the California standards in relation to other state tests, we evaluated the standard used as the cut score for *NCLB* reporting or the *proficient* performance level and compared it to the cut score representing the same standard in these other states. Although the number of states studied is rapidly increasing, the states studied may not reflect what is typical in regard to these kinds of standards.

The results are summarized in Table 18. California's cut scores in both reading and mathematics are well above the NWEA's national median scores in both reading and mathematics and rank among the most challenging of the state standards studied, generally requiring that students perform between the 55th and 70th percentile (with the notable exceptions of grade 2 and grade 10. We'd recommend caution about drawing any judgments about the quality of California's standards from this information alone. States establish standards for different purposes. Some states, Washington might be an example, set standards at a level they believe appropriate for students pursuing post-secondary education. Others may set standards at a lower level that reflects the literacy needed to be successful in the workplace. The No Child Left Behind Act requires schools to set targets that would result in all students achieving a proficient level of performance in about 11 years. While a few communities in California are no doubt close to achieving this already, many will have to improve the performance of large proportions of their students to reach this challenging goal. Our point is that standards should be judged on how well they align with the purposes the community originally wanted to reflect, not purely on how high or low the "bar" is set. The primary thing the tables make clear is that proficiency standards vary widely from state to state and that proficiency is not yet a concept that h as a shared definition.

Relative difficulty of the mathematics and ELA standards

Educators may assume that state standard setting processes are designed to produce standards across subjects that are equal in difficulty. Our previous studies show that this is not always the case. Arizona's math standards, for example, have been considerably more challenging than their standards for reading, although the state is taking steps to bring closer alignment between the two subjects. In general, California's standards for Math and English/Language Arts are similar to each other in difficulty.

Table 18 - Cut scores representing proficient or "meets standards" level of performance on 16 state assessments

Reading

Grade 3			Grade 4			Grade 5			Grade 6			Grade 7			Grade 8			Grade 9			Grade 10		
State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile
SC	205	67	WY	214	73	SC	220	73	SC	221	63	SC	227	70	WY	232	74	MT	224	43	OR	236	77
CA	203	61	SC	213	70	CA	217	65	CA	220	60	WA	226	67	SC	230	68	IA	224	43	WA	227	53
NV	202	58	CA	208	56	NV	215	59	MT	211	35	CA	225	64	CA	230	68	ID	221	37	ID	224	44
MN	193	35	WA	207	53	PA	212	50	ID	211	35	MT	218	43	OR	227	58	CO	204	9	MT	224	44
OR	193	35	ID	200	34	AZ	210	45	IN	210	32	IA	216	37	AZ	224	49				IA	223	42
ID	193	35	MT	196	26	OR	209	42	IA	209	30	ID	215	35	PA	223	46				CO	209	15
MT	193	35	IA	196	26	IL	207	37	TX	208	28	TX	210	24	IN	219	35				CA	208	14
IL	193	35	CO	191	18	MT	206	35	CO	197	11	CO	206	18	MT	219	35						
IN	192	32				ID	206	35							IA	219	35						
IA	191	31				IA	205	32							ID	218	32						
AZ	190	29				MN	204	30							IL	218	32						
TX	179	13				TX	204	30							MN	218	32						
CO	179	13				CO	197	18							CO	206	12						

Mathematics

Grade 3			Grade 4			Grade 5			Grade 6			Grade 7			Grade 8			Grade 9			Grade 10		
State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile	State	Cut Score	%ile
SC	208	75	WY	221	83	SC	227	76	SC	235	78	SC	242	78	WY	257	89	MT	242	47	WA	257	73
CA	202	56	WA	218	76	CA	224	69	CA	231	71	WA	242	78	SC	251	80	IA	241	44	MT	247	40
NV	203	59	SC	217	74	AZ	220	59	IN	221	47	CA	238	71	AZ	248	75	ID	240	42	IA	247	40
IN	201	50	CA	212	59	NV	216	48	ID	219	42	ID	225	44	CA	244	67	CO	235	32	OR	245	33
OR	199	46	ID	205	39	PA	216	48	IA	218	40	MT	224	42	PA	237	53				ID	242	25
AZ	199	46	IA	205	39	OR	215	46	MT	218	40	IA	222	38	OR	235	50				CO	233	14
MN	198	42	MT	205	39	ID	213	41	CO	207	19	TX	221	35	ID	233	46				CA	232	13
MT	197	39				MT	212	38				CO	216	26	MN	231	42						
IA	197	39				IA	212	38							IN	231	42						
ID	196	36				MN	210	33							IL	230	40						
IL	193	29				IL	210	33							MT	228	36						
						TX	209	31							IA	228	36						
						CO	201	15							CO	225	31						

- Indiana tests students in the fall. Their cut scores were adjusted to reflect equivalent spring performance
- Colorado uses the partially proficient level of performance for NCLB reporting. To maintain consistency we report the level each state uses for NCLB reporting here.
- The Texas estimate is based on the level for proficient performance that will be implemented in 2005.

Calibration of the California Standards Across Grades

When we say a standard should be calibrated across grades, we mean that a standard have the same difficulty at every grade level. Standards for grade 8 should not be considerably easier or more difficult than the standards for grade 3. Here are the reasons we take this position:

- **If standards are used to evaluate the effectiveness of teacher or school performance, equity requires that the standards be the same for all. It is simply unfair to hold some teachers and students to a higher standard than others simply because they work at different grade levels. From a practical point of view, teachers will be reluctant to accept teaching assignments at a grade level if it becomes known that the standards associated with that grade level are considerably more difficult to achieve than those imposed at other grades. If you doubt us, call any Arizona middle school principal and ask if it has been easier to fill 6th or 8th grade math positions in the last couple of years.**
- **If standards are used to tell teachers and students whether students are on-track to meet community expectations, it's important that proficiency at third grade truly projects to proficiency at eighth grade, assuming proficient children achieve normal growth. When this is not the case, teachers, students, and their parents receive an inaccurate message about the true performance of their children. In other words, if the third grade standard is considerably easier than the eighth grade standard, reports will tell some third grade families that their children are proficient, when, in fact, their performance is very likely to fall short of proficiency in the future.**

There are significant issues relative to the calibration of standards within the California State Tests. The most significant problem is that the standards for performance in the upper grades (grades 6, 7, and 8) are substantively higher than they are at the younger grades (grades 2, 3, and 4). Let's use mathematics to illustrate the problem.

Figure 2 (see following page) shows the percentile score associated with proficiency on the spring NWEA mathematics test. It shows that the percentile score required for passing the test at grades 2 through 4 is much lower than the near 70th percentile score required to pass the test at grades 6, 7, and 8. Were these patterns to hold up over time, about 13% of the total testing population identified as *proficient* in 3rd grade would fail to meet the standard in 8th for no reason other than lack of calibration in the standard.

Figure 3 is a line graph that compares the RIT score that actually meets the standard each grade with the score that would be required at every grade for a student to be on-track to meet the 8th grade standard. The figure shows that the score currently required by the standard ranges from 3 to 9 points less than the projected 8th grade cut score in grades 2, 3, and 4. While these differences do not immediately seem large, when applied over an entire state they result in thousands of students being identified as proficient in grades 2, 3, and 4 who will grow normally and not achieve proficiency at grade 8. This can result in the delay of needed interventions for these students and can wreak havoc on the stability of adequate yearly progress statistics.

Figure 2 – NWEA spring percentile score projecting to proficient level of performance on CST in mathematics

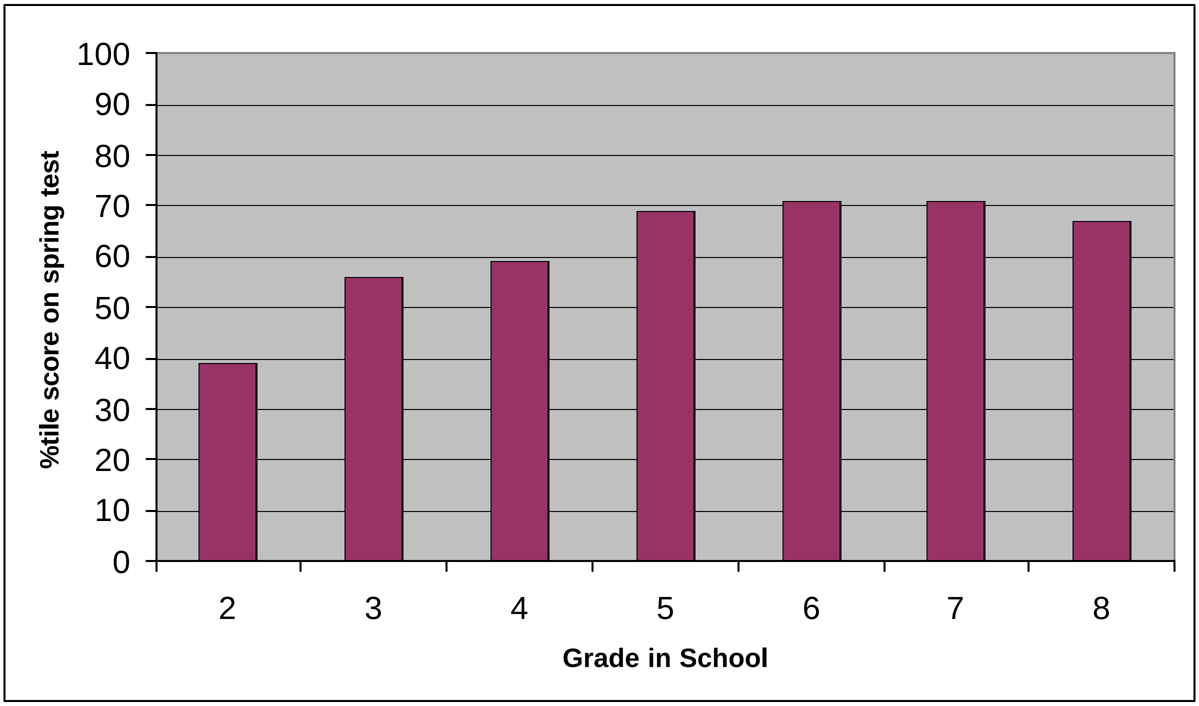
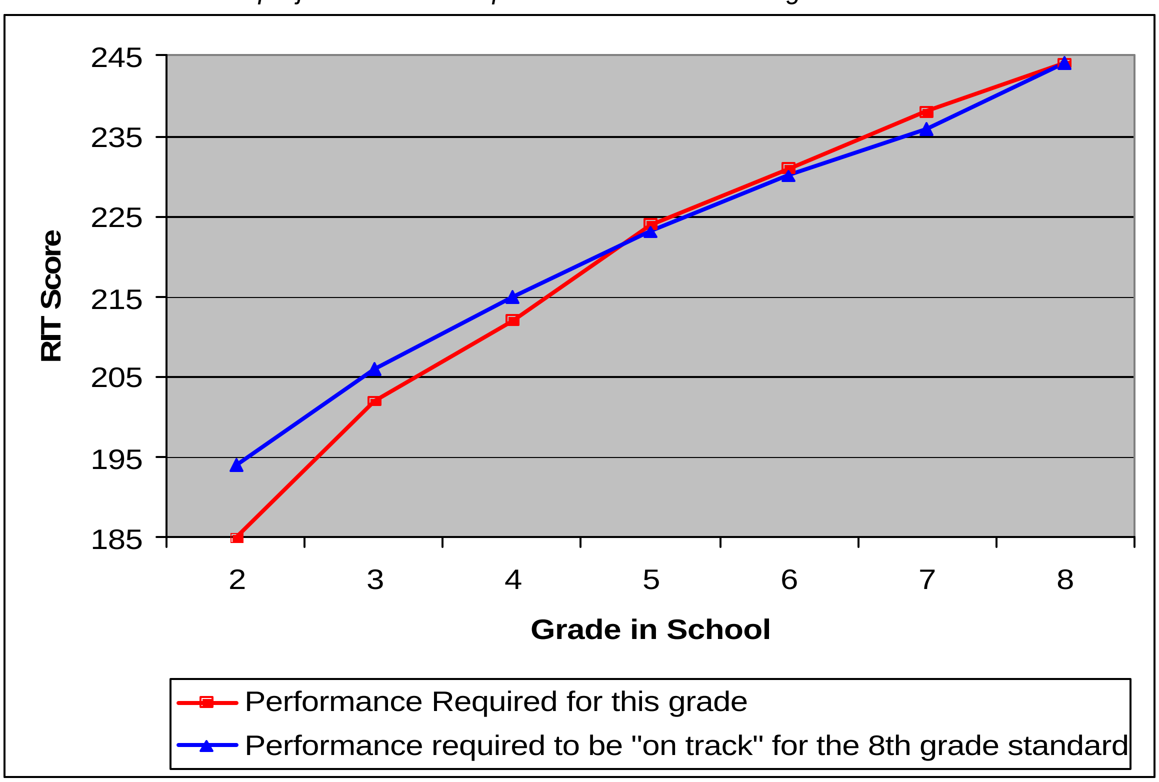


Figure 3 – RIT score projected to achieve proficient score on one grade's CST vs. RIT score required to project to achieve a proficient score on the 8th grade CST



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

Helping students pass the state test is not the primary reason our members use NWEA assessments. We hope they are used to provide teachers information that will allow them to improve the learning of all students. Nevertheless, state test results are important and failing to do well on them can have deleterious effects on students and their schools. Because of this, we believe educators would benefit from knowing more about the probability that a student's RIT score would lead to a passing score on the CST. This would allow educators to more reliably identify students who will need additional resources to reach this level of performance. Equally important, however, it will allow educators to know which students are "safe" against California standards so they can focus their time with these students on providing new challenges that better suit their current needs.

Tables 19 through 24 on the following pages, and the accompanying graphs show the proportion of students at each RIT level who earned scores at or above the *proficient* level on the CST assessments. Using Table 19 as an example, we find that about 12% of the 5th grade students who achieved a reading RIT score between 205 and 209 went on to achieve a passing score on the California ELA assessment. A 5th grade teacher with ten students performing in this range would know that only about one in ten of these students will be proficient on the CST unless they work harder, receive more focused instruction, or have access to additional resources.

On the other hand, about 92% of 5th grade students performing at 225 to 229 level achieved proficiency on the ELA assessment. Teachers should feel free to focus their efforts with these students on new and more difficult challenges than the basic fifth grade standards might provide.

Table 19
Proportion of students achieving proficient performance level on the CST English/Language Arts assessment based on PRIOR FALL RIT score - Reading

RIT Score	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
140	4.35%						
145	4.67%						
150	7.32%						
155	13.51%	0.33%					
160	21.59%	2.03%	0.00%				
165	29.02%	2.25%	2.16%				
170	39.34%	4.23%	1.08%	0.00%	0.00%		
175	52.78%	7.94%	0.82%	0.59%	1.16%		
180	66.32%	12.84%	3.17%	0.94%	0.81%	0.00%	
185	79.00%	23.18%	7.77%	2.11%	0.88%	2.04%	0.00%
190	90.49%	42.46%	14.01%	5.93%	0.83%	1.47%	1.78%
195	95.92%	65.23%	31.37%	15.88%	1.40%	1.06%	0.61%
200	98.50%	83.98%	55.79%	34.83%	5.88%	2.53%	1.48%
205	100.00%	93.80%	79.18%	58.44%	17.73%	6.28%	1.60%
210		98.51%	90.77%	81.40%	40.91%	18.71%	6.33%
215		99.38%	97.77%	94.04%	68.37%	39.19%	12.02%
220		98.28%	99.76%	98.17%	87.01%	68.57%	35.48%
225		100.00%	100.00%	99.45%	96.23%	88.77%	62.88%
230				100.00%	99.33%	96.95%	84.45%
235					100.00%	99.76%	96.08%
240						100.00%	98.23%
245							99.04%
250							100.00%

Table 20
Proportion of students achieving proficient performance level on the CST English/Language Arts assessment based on same SPRING RIT score - Reading

RIT Score	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
140							
145	2.11%						
150	0.90%						
155	2.82%						
160	2.99%		0.00%				
165	5.06%		0.65%				
170	7.57%		1.00%	0.00%			
175	12.45%	0.57%	0.63%	0.57%	0.00%		
180	22.92%	2.24%	1.22%	0.95%	1.39%		
185	40.57%	4.73%	1.42%	0.75%	1.06%		
190	61.97%	9.61%	4.16%	0.34%	1.22%	1.94%	
195	77.30%	23.13%	8.98%	2.59%	0.93%	0.66%	0.65%
200	89.68%	47.97%	21.06%	6.02%	1.74%	1.84%	1.38%
205	96.41%	69.27%	45.98%	12.73%	5.78%	3.40%	1.39%
210	99.31%	89.07%	72.52%	32.18%	15.05%	6.87%	2.27%
215	98.18%	96.09%	88.99%	55.45%	37.58%	20.82%	6.38%
220	96.67%	99.35%	97.73%	80.43%	68.46%	44.77%	14.10%
225	100.00%	99.24%	98.95%	92.41%	87.32%	70.15%	36.78%
230		100.00%	100.00%	99.39%	96.04%	89.28%	60.06%
235				99.45%	99.24%	96.97%	86.88%
240				100.00%	99.31%	98.10%	94.36%
245					100.00%	100.00%	98.68%
250							98.98%
255							100.00%

Figure 4 – Proportion of students achieving proficient performance level on the CST English/Language Arts assessment based on PRIOR FALL RIT score - Reading

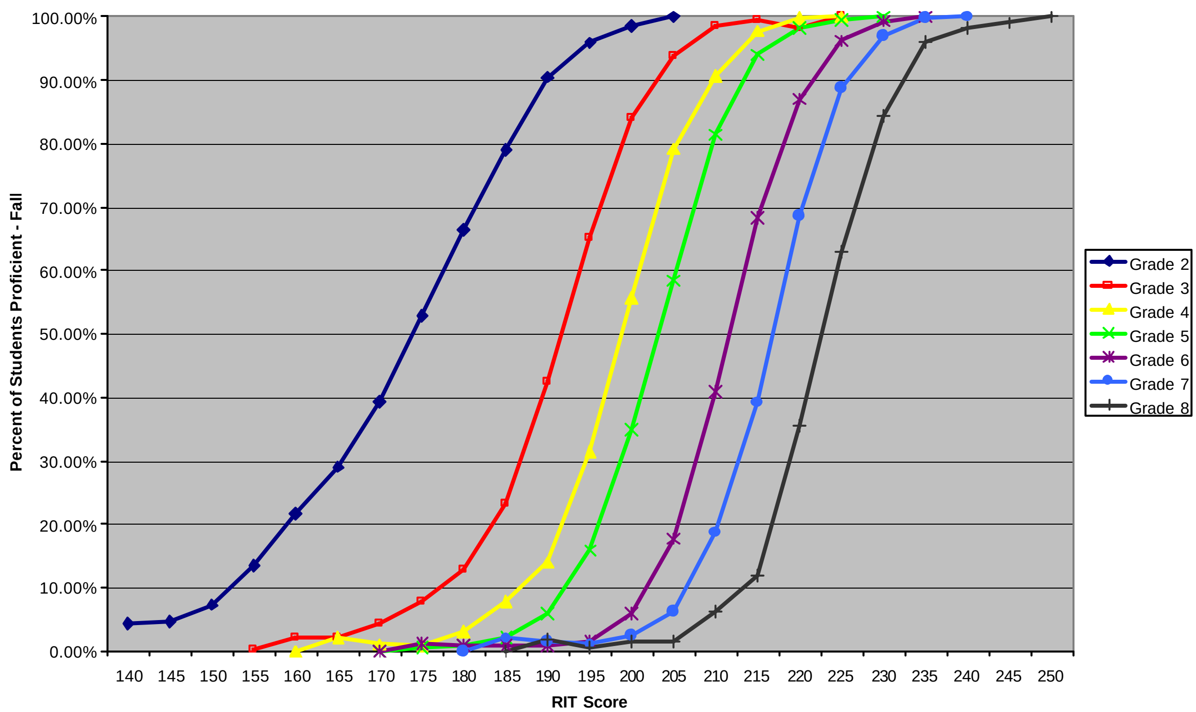


Figure 5 – Proportion of students achieving proficient performance level on the CST English/Language Arts assessment based on same SPRING RIT score - Reading

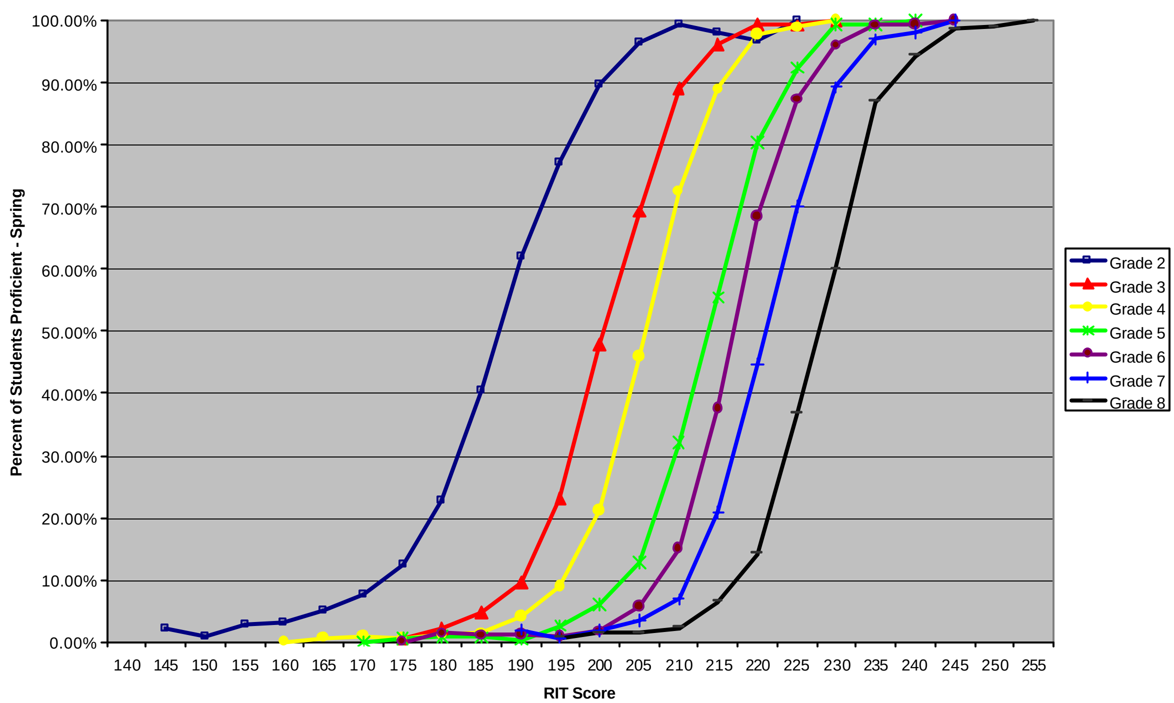


Table 21
Proportion of students achieving proficient performance level on the CST English/Language Arts assessment based on PRIOR FALL RIT score – Language Usage

RIT Score	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
145	0.00%						
150	2.73%						
155	6.23%	0.00%	0.00%				
160	9.01%	0.76%	1.21%				
165	19.62%	2.01%	1.12%				
170	28.57%	1.90%	1.38%				
175	42.90%	5.52%	1.06%				
180	61.93%	7.85%	0.86%	0.00%			
185	80.98%	15.58%	3.39%	0.83%		0.00%	
190	89.38%	29.07%	8.25%	0.58%	1.04%	0.44%	0.00%
195	94.09%	50.86%	18.64%	3.50%	1.74%	0.94%	1.06%
200	96.53%	74.91%	40.48%	10.48%	3.58%	1.61%	0.88%
205	100.00%	90.80%	64.02%	24.05%	11.31%	5.06%	1.32%
210		96.71%	83.35%	47.65%	26.76%	14.44%	4.43%
215		99.68%	94.52%	72.06%	52.46%	34.76%	15.16%
220		98.84%	98.65%	90.64%	79.13%	62.03%	33.77%
225		100.00%	99.50%	96.21%	95.17%	85.59%	64.32%
230			100.00%	100.00%	98.68%	94.72%	84.48%
235					99.44%	98.65%	95.55%
240					100.00%	100.00%	99.33%
245							98.68%
250							100.00%

Table 22
Proportion of students achieving proficient performance level on the CST English/Language Arts assessment based on same SPRING RIT score – Language Usage

RIT Score	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
160	0.70%						
165	2.17%						
170	7.02%						
175	12.39%						
180	22.42%	0.86%					
185	46.21%	3.39%	1.00%	0.00%			
190	63.61%	5.59%	1.24%	0.72%			3.33%
195	81.88%	13.96%	4.27%	0.72%		1.86%	1.83%
200	94.20%	28.80%	12.67%	1.59%	2.48%	0.96%	1.31%
205	97.92%	57.36%	33.75%	6.87%	3.96%	2.51%	1.59%
210	97.35%	80.25%	59.75%	24.02%	12.78%	5.96%	3.09%
215	94.90%	92.53%	80.31%	46.75%	32.32%	19.16%	7.58%
220	93.10%	97.05%	93.29%	71.78%	57.84%	43.76%	18.65%
225	100.00%	98.82%	97.58%	91.13%	83.26%	67.31%	43.75%
230		91.30%	98.68%	96.45%	94.98%	90.69%	72.36%
235		92.31%	100.00%	98.95%	98.92%	96.63%	88.36%
240		100%		100.00%	100.00%	98.36%	96.46%
245					96.97%	99.19%	98.31%
250					100.00%	100.00%	98.41%
255							100.00%

Figure 6
Proportion of students achieving proficient performance level on the CST English/Language Arts assessment based on PRIOR FALL RIT score – Language Usage

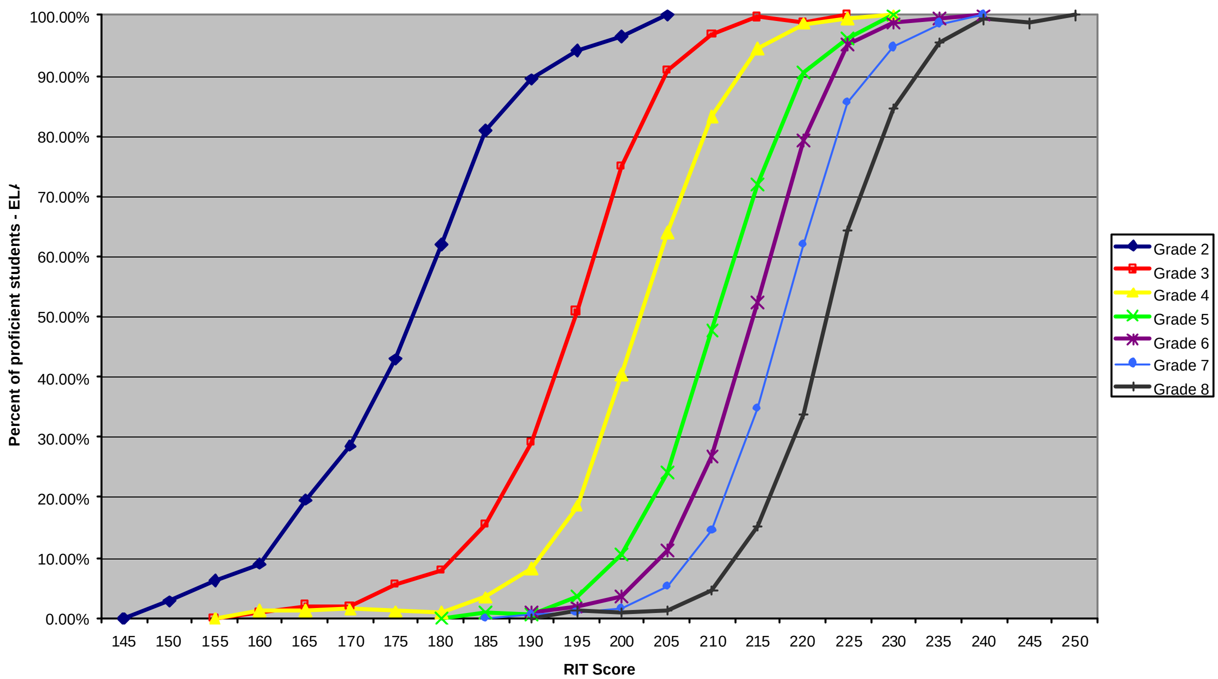


Figure 7
Proportion of students achieving proficient performance level on the CST English/Language Arts assessment based on same SPRING RIT score – Language Usage

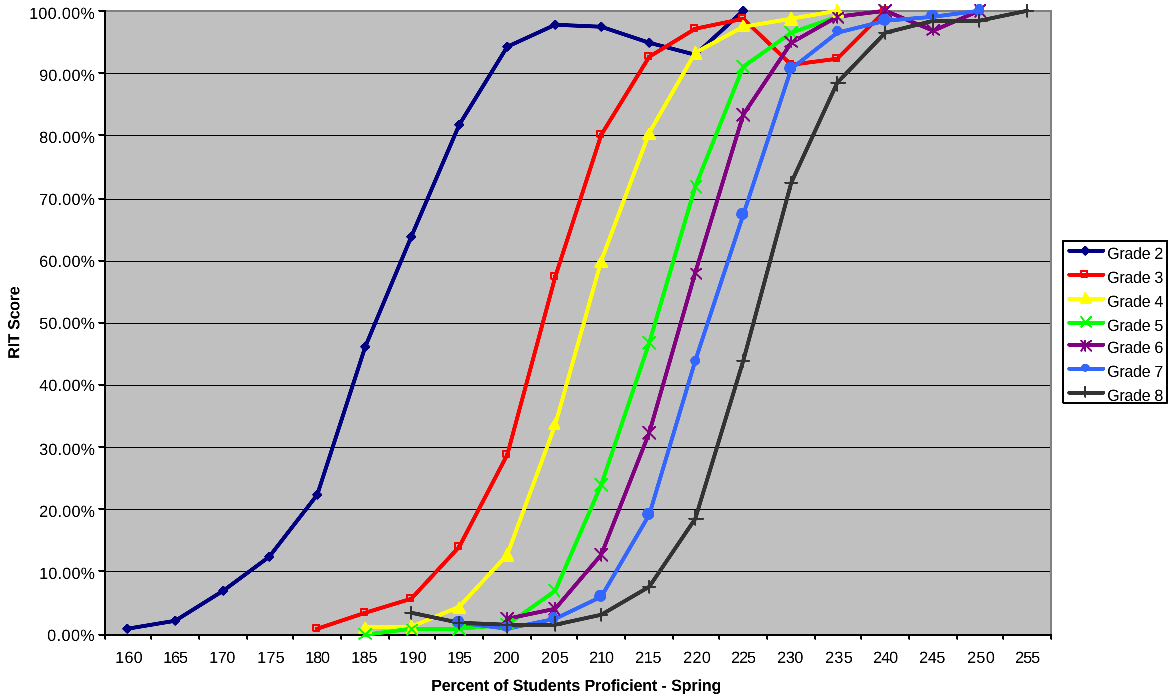


Table 23

Proportion of students achieving proficient performance level on the CST mathematics assessment based on PRIOR FALL RIT score – Mathematics

RIT Score	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
140	0.00%						
145	9.09%	0.00%					
150	25.30%	7.35%					
155	21.86%	2.46%					
160	32.80%	1.50%					
165	38.84%	7.33%					
170	56.48%	8.33%	0.00%	0.00%			
175	74.60%	15.62%	1.47%	0.76%	0.00%		
180	91.60%	25.88%	1.35%	1.05%	0.42%		
185	96.96%	40.11%	6.31%	1.88%	1.04%		
190	98.52%	59.13%	11.14%	1.57%	0.82%		
195	99.36%	80.96%	28.39%	1.98%	0.77%		
200	97.56%	90.81%	46.91%	4.46%	5.01%	0.41%	0.00%
205	100.00%	97.53%	68.60%	13.95%	11.42%	0.71%	0.73%
210		100.00%	89.44%	31.62%	28.56%	1.25%	1.70%
215			96.03%	57.67%	53.88%	5.43%	5.93%
220			99.40%	80.34%	76.44%	11.25%	11.49%
225			100.00%	93.74%	89.88%	28.39%	24.52%
230				97.84%	96.98%	50.20%	37.60%
235				100.00%	99.71%	73.90%	52.27%
240					100.00%	87.91%	63.81%
245						97.84%	70.92%
250						99.66%	81.92%
255						100.00%	91.93%
260							98.53%
265							99.08%
270							100.00%

Table 24

Proportion of students achieving proficient performance level on the CST mathematics assessment based on same SPRING RIT score – Mathematics

RIT Score	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
145	0.00%						
150	4.76%						
155	4.71%						
160	3.01%						
165	7.62%						
170	9.80%	0.00%				0.00%	
175	17.78%	1.08%	0.00%	0.00%		2.56%	0.00%
180	22.03%	1.91%	0.87%	0.77%	0.00%	1.39%	4.17%
185	38.69%	4.42%	0.57%	0.00%	0.68%	1.79%	1.39%
190	56.28%	11.46%	2.22%	0.32%	1.07%	0.94%	2.82%
195	79.27%	27.60%	4.34%	0.20%	0.28%	0.68%	0.86%
200	93.62%	54.31%	13.69%	1.87%	1.34%	0.87%	0.37%
205	97.87%	76.17%	30.12%	2.32%	1.09%	1.98%	3.37%
210	99.05%	91.88%	54.99%	6.84%	1.40%	0.89%	1.48%
215	98.71%	97.46%	74.68%	14.50%	5.74%	0.83%	2.08%
220	95.95%	98.41%	89.47%	32.50%	15.65%	2.87%	2.76%
225	97.87%	97.93%	97.22%	57.68%	35.25%	7.84%	7.78%
230	100.00%	96.52%	98.34%	80.26%	63.58%	17.49%	17.08%
235		92.50%	100.00%	94.11%	80.33%	38.68%	25.18%
240		84.62%	99.22%	98.86%	93.72%	59.58%	43.37%
245		100.00%	100.00%	99.37%	98.83%	82.71%	52.78%
250				100.00%	99.70%	92.49%	59.75%
255					100.00%	98.51%	74.06%
260						99.12%	88.65%
265						99.33%	95.78%
270						99.12%	98.70%
275						96.67%	100.00%
280						100.00%	

Figure 8
Proportion of students achieving proficient performance level on the CST mathematics assessment based on PRIOR FALL RIT score – Mathematics

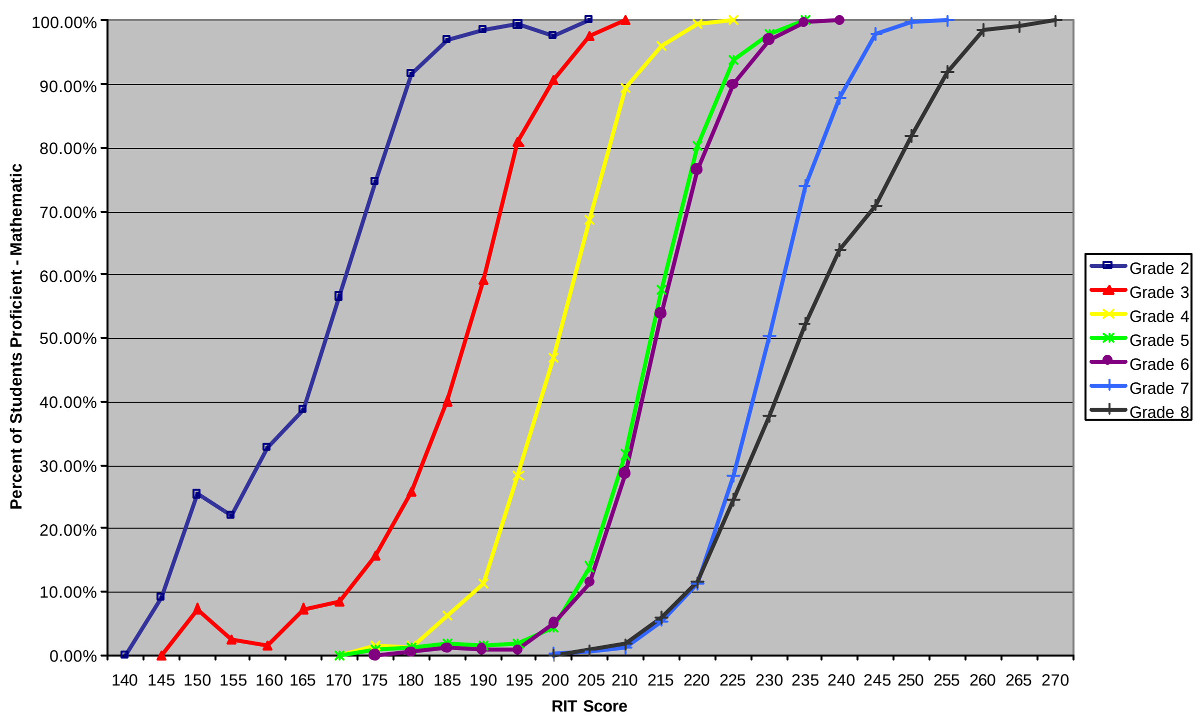
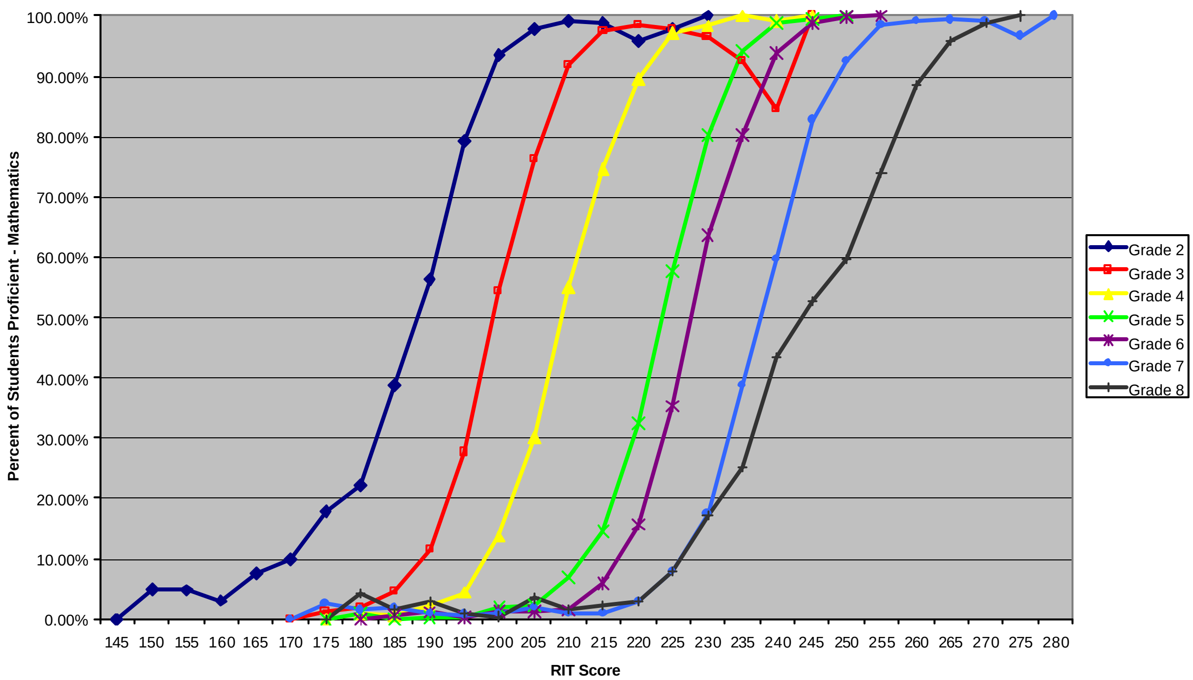


Figure 9
Proportion of students achieving proficient performance level on the CST mathematics assessment based on same SPRING RIT score – Mathematics



Using RIT scores and data from this alignment study to set individual growth targets

NWEA encourages educators and parents to collaborate on setting individual growth targets for students based on what we call a "hybrid-growth model". The *proficient standard* **cut score for each grade reflect benchmarks that students who are "on-target" would meet if they were to achieve the state's benchmark for the** *No Child Left Behind Act***. For students who are behind this benchmark, we recommend a growth target that would reflect the norm for their grade and RIT range (see the 2002 NWEA norms study for this information) plus some proportion of the gap between their current performance and the benchmark that the student would try to close during this school year. For those students whose performance is ahead of the benchmark, we suggest a target that reflects their current RIT range norm.**

This approach assures that each student has a growth target that is challenging. It also assures that low performing students have targets that will assure they eventually reach proficiency standards. Schools that achieve high rates of success on these kinds of targets will assure that no child is left behind (to borrow a phrase) while also making sure that all children have the opportunity to get ahead, regardless of where they stand against a standard. More information on this approach can be obtained by contacting the Research team at NWEA.

Summary and Conclusions

This study investigated the relationship between the scales used for the CST assessments and the RIT scales used to report performance on Northwest Evaluation Association tests. The study determined RIT score equivalents for the CST performance levels in English/Language Arts and mathematics. Test records for more than 73,000 students were included in this study.

Three methods generated an estimate of RIT cut scores that could be used to project CST performance levels. Rasch SOS and second-order regression methods generally produced the most accurate projections of cut scores. Accuracy of predicting proficient performance on the CST from spring NWEA assessments was above 83% for all grades and above 82% for all grades when fall NWEA scores were used.

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 California educators use NWEA assessments to better inform, plan and deliver student instruction. Good information, when matched with the professionalism and commitment of our colleagues, will assure that every student has the opportunity to reach their aspirations.