Aligning the NWEA RIT Scale with the Nevada Criterion Referenced Assessment and the Iowa Test of Basic Skills

January, 2004

John Cronin, Ph.D. and Branin Bowe – Northwest Evaluation Association

Each year, Nevada students in grades 3, 4, 5, and 7 participate in testing as part of the Nevada assessment program. Students in grades 3 and 5 take the Nevada Criterion Referenced Assessment (Nevada CRT) while students in grades 4 and 7 take the Iowa Test of Basic Skills (ITBS). 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 Nevada Department of Education has developed scales that are used to assign students to one of four performance levels on the Nevada CRT. These are, from the lowest cut score to the highest: *developing, approaches, meets, and exceeds*. For purposes of NCLB, the *meets* level is considered the level that represents satisfactory performance. Students taking the Iowa Test of Basic Skills are also assigned to one of four levels. These levels simply reflect the four quartiles reported in the ITBS norms.

Many students who attend school in Nevada also take 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 Nevada have been specifically aligned to match the content of local and Nevada 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 Nevada.

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 one of an ongoing series of studies that are being conducted to identify the relationships between NWEA tests and state-mandated assessments. Studies 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 Nevada CRT?
- What RIT scores correspond to various performance levels on the Nevada CRT and Iowa Test of Basic Skills?
- How well can *passing* performance on the Nevada assessments (*meets* performance on the Nevada CRT and 50th percentile or better performance on the ITBS) be predicted from RIT scores when NWEA assessments are administered in the same time frame?

Method

Participating School Systems

An e-mail solicitation was sent in October, 2003 to all Nevada school systems who had two or more seasons of experience with NWEA testing prior to spring 2003 to secure participants for the study. Based on the response from this solicitation, spring 2003 Nevada CRT and NWEA student assessment records were collected from three school districts.

Data Preparation

For purposes of studying NWEA test alignment with the Nevada CRT, student level test records from spring 2003 Nevada testing and spring 2003 NWEA assessments were matched using district assigned student ID numbers. Matched records were then screened to remove invalid scores. Table I shows the number of student records included in the reading analysis for this study.

Table 1
Reading and Mathematics Tests Included by Grade

Subject	3	5	Total
Reading	1084	1184	2268
Mathematics	1087	1155	2242

We had enough student records at each grade to adequately cover the breadth of the scale and perform a robust analysis. Because the study involved a small number of districts, we recommend that schools validate our estimates by cross-checking their own students' performance against our cut scores.

Two prior studies have been completed aligning the RIT scale with performance level cutpoints on the Iowa Test of Basic Skills (Kingsbury, 1999; Kingsbury, Breithaupt, and Hauser, 2002). Results from these prior studies were used to estimate the cutpoints on the RIT scale that would be equivalent to the ITBS performance levels set for Nevada.

Analyses

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

Linking Nevada CRT and ITBS scores to the RIT scales. Three methods of estimating cut scores for Nevada CRT levels (abbreviated NEV) were used. The most straightforward was simple linear regression ($NEV_{pred} = a(RIT) + c$). 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 ($NEV_{pred} = a(RIT^2) + b(RIT) + c$). For each of these methods, the RIT score was determined by substituting the appropriate Nevada CRT score for NEV_{pred} and solving the equation for RIT.

A fixed-parameter Rasch model was also used to estimate RIT cut scores. In this method, the Nevada CRT performance level was treated as a NEV 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.

Our prior studies of the ITBS indicated that the highest predictive correlations were obtained when the regression equation included the linear and cubic terms. The results in the applied predictive model of the form are expressed as:

Predicted ITBS %ile = w0 + w1 * RIT + w3 * RIT * RIT * RIT

This equation was applied to estimate the points on the RIT scale that were equivalent to the quartile cutpoints used by Nevada on the ITBS.

Predicting Nevada CRT performance levels from RIT scores. RIT scores were first used to predict whether students were likely to achieve performance at or above the *meets* performance level on the Nevada CRT. 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 Nevada CRT 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 Nevada CRT 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 Nevada CRT levels based on observed RIT scores. The predictions of Nevada CRT 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.

Because the accuracy of prediction had been established by prior studies, we did not generate accuracy of prediction estimates for the ITBS as part of this study.

Content Validity

Formal comparisons of the content of NWEA and Nevada 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 Nevada assessments. Both NWEA assessments and the Nevada assessments

include multiple-choice items. The Nevada CRT also includes short answer and extended response questions. In reading, all of the Nevada test items are linked to extended reading passages of 300 to about 600 words in length. NWEA's reading tests do not link each item to an extended reading passage and the reading passages for grade 3 students are considerably shorter than those used on the Nevada CRT.

Results

Descriptive Statistics

Table 2 reviews descriptive statistics for the Nevada CRT and NWEA assessments. The median RIT scores for this sample are slightly above those for the NWEA norm population with median scores for the sample falling within 2 points of the median for the norm group in all subjects and grades. These differences are small enough that they will not affect the ability of the sample to represent the typical range of student performance on NWEA's assessments. We do not anticipate, therefore, that the accuracy of cut score predictions will be affected by an unusual distribution of scores within the sample.

Table 2
Means, Standard Deviations, and Medians for the Nevada CRT and NWEA assessments

Grade	3	5
Nevada Reading CRT		
N	1084	1153
Mean	307.01	296.60
Median	206	301
Std. Deviation	67.55	67.35
NWEA Reading		
N	1090	1147
Mean	199.80	211.88
Median	202	214
Std. Deviation	13.58	13.35
Nevada Mathematics CRT		
N	1089	1155
Mean	303.16	302.74
Median	303	305
Std. Deviation	73.66	62.29
NWEA Mathematics		
N	1088	1156
Mean	202.93	217.29
Median	204	218
Std. Deviation	10.12	12.67

Pearson correlations

Table 3 shows the results of this analysis for each grade. Concurrent validity was tested by examining same subject Pearson correlations between the NWEA and Nevada CRTs. Correlations were quite high, ranging from .76 (grade 3 mathematics) to .86 (grade 5 mathematics) indicating that the tests were generally measuring the same constructs. Discriminant validity was tested by examining same subject Pearson correlations next to

correlations for the alternate subject (math against reading). In all cases the same subject correlations were higher than correlations against the alternate subject.

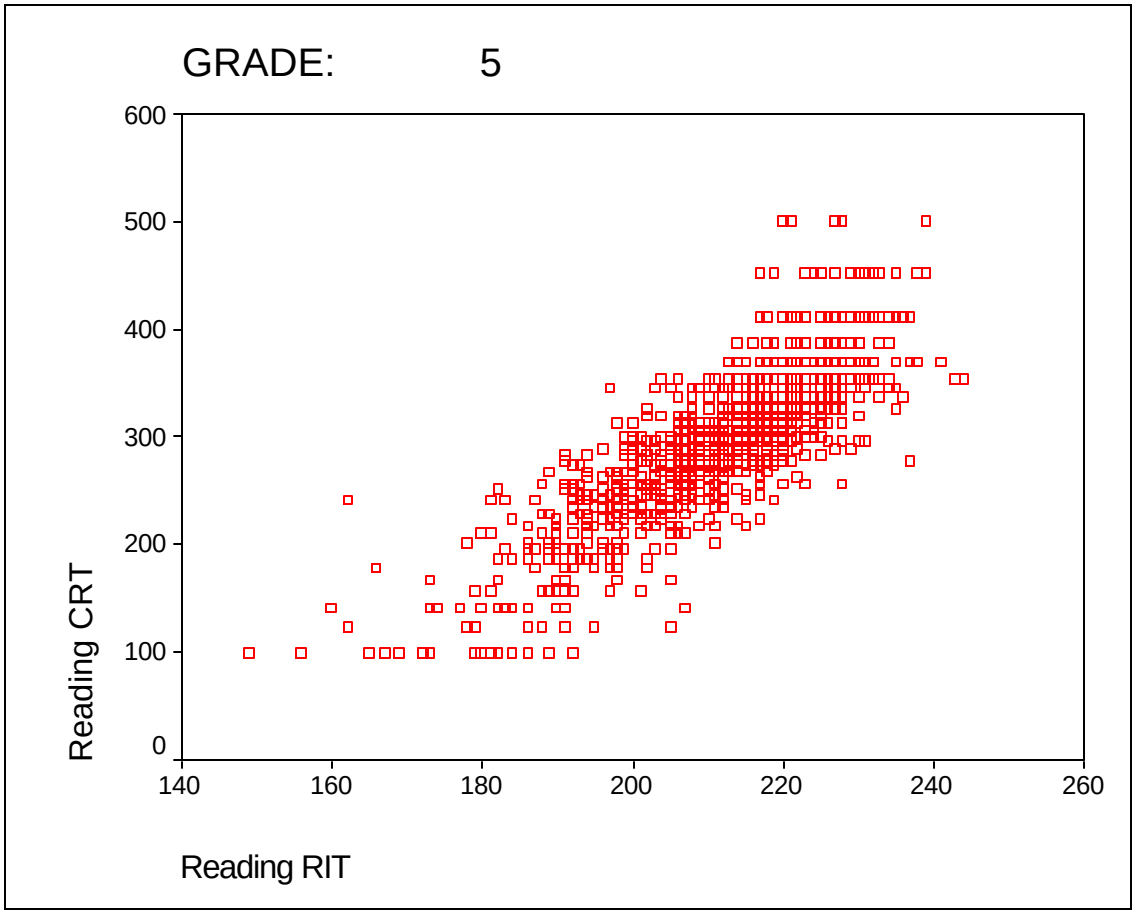
Table 3
Pearson Correlations for Nevada CRT and NWEA assessments by Subject - Reading

Grade 3		
	Nevada CRT Reading	**Nevada CRT Math**
NWEA Reading	.82	.65
NWEA Math	.70	.76
Grade 5		
	Nevada CRT Reading	**Nevada CRT Math**
NWEA Reading	.83	.72
NWEA Math	.72	.86

- Same subject correlations are shaded

Analysis of scatterplots suggested that some of the scale relationships might be curvilinear and revealed some tendency for correlations to break down near the lower end of the scales, possibly indicating a floor effect on the Nevada CRT. Figure 1 provides an example that illustrates both the scale relationships and some breakdown in correlation near the bottom of the Nevada CRT scale. Readers might also note the "rows" near the top of the Nevada CRT scale, which are a product of successively larger differences in scale scores as the student's raw score increases near the top of the scale.

Figure 1 – Scatterplot depicting Grade 5 NWEA reading against the Grade 5 Nevada CRT reading score



Linking Nevada CRT cut scores to the RIT scale

The primary purpose of this study was to estimate the RIT scale scores that most closely correspond to the cut scores for different performance levels on the Nevada CRT and ITBS. 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 4 shows several estimations of the Spring 2003 RIT score that correspond to the cut scores for the various performance levels on the Nevada CRT scale. Table 5 shows our estimate of the Spring 2003 RIT score that corresponds to the performance levels established for the ITBS.

Table 4
Estimated points on the RIT scale equating to the minimum scores (rounded) for performance levels on the Nevada CRT

	Linear Regression				Second-order Regression				Rasch Status-on-Standard			
Reading	**Dev**	**App**	**Meet**	**Exc**	**Dev**	**App**	**Meet**	**Exc**	**Dev**	**App**	**Meet**	**Exc**
Grade 3	<175	175	**200**	213	<173	173	**201**	212	<178	178	**202**	212
Grade 5	<189	189	**213**	234	<189	189	**213**	232	<193	193	**215**	230
	Linear Regression				**Second-order Regression**				**Rasch Status-on-Standard**			
Mathematics	**Dev**	**App**	**Meet**	**Exc**	**Dev**	**App**	**Meet**	**Exc**	**Dev**	**App**	**Meet**	**Exc**
Grade 3	<185	185	**203**	216	<184	184	**203**	215	<189	189	**204**	215
Grade 5	<193	193	**216**	235	<194	194	**218**	236	<199	199	**218**	236

Table 5
Estimated points on the RIT scale equating to the minimum scores (rounded) for Nevada performance levels on the Iowa Test of Basic Skills

Reading to ITBS Reading				
Perf. Level	**1**	**2**	**3**	**4**
Grade 4	<184	185	**194**	206
Grade 7	<206	206	**215**	223
Perf. Level	**1**	**2**	**3**	**4**
Grade 4	<193	193	**200**	206
Grade 7	<212	212	**220**	236

Predicting Nevada CRT pass-fail status from RIT scores

Once the 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 Nevada CRT. The most accurate method of prediction was generally used to derive the best estimate of RIT cut scores that equate to the different Nevada CRT performance levels. 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 *meets* standard level on the Nevada CRT. Next we assessed the accuracy with which the RIT predicted 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 *meets* level and beyond. RIT scores can provide teachers with advance notice about students who may not reach these goals on the Nevada assessment that corresponds to their grade level.

Table 6 shows the results for reading. When using the methodology chosen to produce the most reasonable estimate, the rate of accurate prediction was consistently between 83% and 84%. Type I error ranged between about about 8% and 11% of the sample population and prediction index scores were above .860 at every grade. The results suggest that the NWEA reading assessment predicts *pass-fail* status on the Nevada CRT reasonably well.

Table 7 shows the results for mathematics. The accuracy of proficiency status predictions ranged from about 77% to 83% when using the best methodology. Type 1 errors ranged from about 10% to 14%. Prediction index scores were all above .920 when using the best method. The results again suggest an acceptable rate of prediction for all grades.

The Rasch method of calibration generally produced more accurate predictions of proficiency status than either regression method in all subjects.

Table 6
Accuracy of the RIT scale in predicting Nevada CRT pass-fail status – reading

Grade 3	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	200	83.56%	11.35%	.864
Second Order	201	84.59%	9.41%	.889
Rasch SOS	202	84.23%	8.67%	.897*
Grade 5	**Cut Score**	**Accuracy**	**Type I Error**	**Prediction Index**
Linear	213	83.49%	10.77%	.871
Second Order	213	83.49%	10.77%	.871
Rasch SOS	215	84.10%	7.91%	.906*

* Indicates methodology chosen for recommended estimate

Table 7
Accuracy of the RIT scale in predicting Nevada CRT pass-fail status – mathematics

Grade 3	Cut Score	Accuracy	Type I Error	Prediction Index
Linear	203	77.11%	14.34%	.814
Second Order	203	77.11%	14.34%	.814
Rasch SOS	204	77.76%	11.76%	.849*
Grade 5	**Cut Score**	**Accuracy**	**Type I Error**	**Prediction Index**
Linear	216	81.44%	14.40%	.823
Second Order	218	83.09%	9.80%	.882
Rasch SOS	218	83.09%	9.80%	.882*

* Indicates methodology chosen for recommended estimate

Table 8 summarizes the accuracy of prediction for this study relative to other state alignment studies. The prediction indices for Nevada are low relative to those we have generated for other states in reading and mathematics.

One possible reason for this is that the Nevada cut scores are estimated to be near the middle of the student performance distribution (the 50th percentile on NWEA norms). Prediction index scores in states with extremely high or low standards are often higher because it is easier to predict a passing score for students when we know in advance that 90% typically pass than it is to predict a passing score when only 50% typically pass.

There are some other factors that may also play a role in explaining the lower than usual prediction index scores. In reading, one possibility is the length of the passages in the Nevada CRT which range from 300 to 500 words at grade 3 and 300 to 600 words at grade 5. These are

considerably longer than NWEA passages for those grades and the grade 3 passages are longer than those used for many state tests at that grade level. In addition, all Nevada questions are linked to reading passages, a design that is different from those of both the NWEA and many state assessments. Finally, the use of constructed response questions on the Nevada CRT may influence the quality of prediction, although most other state tests introduce some constructed response items in reading.

We could find nothing in descriptions of the Nevada CRTs for mathematics that would be helpful in explaining the lower rate of prediction relative to other states in this subject.

In any case, the rates of correct prediction remain high enough to provide useful information to educators who are planning instruction to ensure all students perform at a level that meets the standards. In all grades and subjects, NWEA assessments generate a minimum of four correct predictions of pass/fail status for each Type I error.

Table 8
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	Indiana '01	.907	Wyoming	.961
Minnesota	.944	Colorado '03	.903	Colorado '01	.957
Wyoming	.931	Indiana '03	.894	Washington	.949
Colorado '03	.931	Arizona	.874	Illinois	.946
Illinois	.928			California	.944
California*	.921			Colorado '03	.943
Arizona	.912			South Carolina	.943
Colorado '01	.910			Minnesota	.936
Nevada	**.902**			Washington	.936
South Carolina	.902			Arizona	.919
Indiana '01	.902			Indiana '01	.899
Indiana '03	.900			**Nevada**	**.866**
Washington	.886			Indiana '03	.860

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

Predicting Nevada CRT Performance Levels from RIT Scores

The Nevada CRT reports four levels of performance. Three cut scores are set to define these four levels. Analyzing the capacity of RIT scores to predict students' Nevada CRT 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. The *exceeds* level on the Grade 5 Nevada CRT in reading is especially hard to project because so few students attained this level of performance.

When predicting performance levels, a case is identified as accurate when the performance level assigned by the Nevada CRT 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 *exceeds* performance for the student and the Nevada test result is *meets*, we declare the case a Type I error because the RIT score overestimated performance.

Table 9
Accuracy of the RIT scale in predicting Nevada CRT performance level – reading

Grade 3	**Accuracy**	**Type I Error**	**Prediction Index**	**% Exceeds found**	**% Developing found**
Linear	67.44%	18.91%	.720	58.6%	49.3%
Second Order	68.08%	18.82%	.724	63.4%*	43.7%
Rasch SOS	67.25%	16.88%	.749*	63.4%*	62.0%*
Grade 5	**Accuracy**	**Type I Error**	**Prediction Index**	**% Exceeds found**	**% Developing found**
Linear	70.98%	17.03%	.760	15.4%	44.9%
Second Order	70.72%	17.64%	.751	21.5%	44.9%
Rasch SOS	70.98%	14.25%	.799*	30.8%*	69.2%*

* indicates most accurate method

Table 10
Accuracy of the RIT scale in predicting Nevada CRT performance level – mathematics

Grade 3	**Accuracy**	**Type I Error**	**Prediction Index**	**% Exceeds found**	**% Developing found**
Linear	61.40%	23.53%	.617	42.3%	44.9%
Second Order	61.03%	24.72%	.595	46.8%*	37.8%
Rasch SOS	62.32%	20.04%	.678*	46.8%*	61.2%*
Grade 5	**Accuracy**	**Type I Error**	**Prediction Index**	**% Exceeds found**	**% Developing found**
Linear	71.03%	21.94%	.691	69.4%	37.3%
Second Order	73.20%	16.48%	.775	67.1%*	41.0%
Rasch SOS	73.03%	15.09%	.793*	67.1%*	60.2%*

* indicates most accurate method for this purpose

The results reported in tables 9 and 10 suggest that the Rasch method generally produced the most accurate predictions for students at all levels of performance. The accuracy of performance level prediction for the grade 3 mathematics test was low relative to the other assessments.

NWEA has reported estimated performance level assignments for prior studies conducted in ten states. Table 11 compares the accuracy with which these tests predict performance level. The results show that the Nevada CRT performance index scores near the median in reading but second from the lowest in mathematics among those studies conducted to date. Nevada's use of 4 performance levels depresses the prediction index number (a few states use three), but this would not be the primary factor explaining the lower score.

Table 11
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
South Carolina	.757	South Carolina	.764
Arizona	.756	Arizona	.756
Washington	.698	**Nevada**	**.742**
Minnesota	.627	Minnesota	.611

Best estimates of Nevada CRT performance level cut scores

To estimate the RIT scores that best predict the cut scores for the various Nevada CRT performance levels we did the following:

- For the *meets* RIT score, we selected the methodology that produced the highest performance index score in predicting "pass/fail" alone.
- For the *developing/approaches* RIT score and the *exceeds* RIT score, we selected the cut scores that correctly predicted the largest proportion of students who actually achieved these levels of performance on the Nevada CRT.

Table 12 summarizes the recommended cut scores for each performance level on the Nevada CRT. Based on NWEA student growth norms, the table also includes an estimated cut score for grade 4 that would indicate "on-track" performance for students who will be taking the grade 5 Nevada CRT in that subject the following year.

Table 13 reports the estimated cut scores for each Nevada defined performance level on the ITBS.

Table 12
Projected RIT Scores Equivalent to Performance Levels on Nevada CRT
(estimated scores for years not tested are in blue)

Reading	Developing			Approaches	Meets			Exceeds		
	Score Range	% of pop. identified	Method	Cut Score	Cut Score	Perf. Index	Method	Cut Score	% of pop. Identified	Method
Grade 3	<178	62.0%	Rasch	178	202	.749	Rasch	212	63.4%	Rasch
Grade 4	<186			186	209			222		
Grade 5	<193	69.2%	Rasch	193	215	.799	Rasch	230	30.8%	Rasch
Mathematics	**Developing**			**Approaches**	**Meets**			**Exceeds**		
	Score Range	% of pop. identified	Method	Cut Score	Cut Score	Perf. Index	Method	Cut Score	% of pop. Identified	Method
Grade 3	<189	61.2%	Rasch	189	204	.678	Rasch	215	46.8%	Rasch/2nd Order
Grade 4	195			194	212			226		
Grade 5	<199	60.2%	Rasch	199	218	.793	Rasch	236	67.1%	Rasch

Table 13
Estimated points on the RIT scale equating to the minimum scores for Nevada performance levels on the Iowa Test of Basic Skills

Reading to ITBS Reading				
Perf. Level	1	2	3	4
Grade 4	<184	185	**194**	206
Grade 7	<206	206	**215**	223
Perf. Level	1	2	3	4
Grade 4	<193	193	**200**	206
Grade 7	<212	212	**220**	236

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

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 believed educators would benefit from knowing more about the probability that a student's RIT score would lead to a passing score on the Nevada CRT. 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 Nevada standards so they can focus their time with these students on providing new challenges that better suit their current needs.

Tables 14 and 15 show the proportion of students at each RIT level who earned scores at or above the *meets* level on the Nevada CRT reading and mathematics assessments. Using Table 14 as an example, we find that about 27% of the 3rd grade students who achieved a reading RIT score between 195 and 199 went on to achieve a passing score on the Nevada reading assessment. A 3rd grade teacher with ten students performing in this range would know that only about three in ten of these kinds of students will meet standard on the Nevada CRT unless they work harder, receive more focused instruction, or have access to additional resources.

On the other hand, about 92% of third grade students performing at a 210 to 214 level achieved proficiency on the Nevada reading assessment. Teachers should feel free to focus their efforts with these students on new and more difficult challenges than the basic third grade standards might provide.

Figures 2 and 3 are graphic depictions of the data in the tables.

Table 14
Proportion of students passing the Nevada CRT based on same spring RIT score - Reading

RIT	Grade 3	Grade 5
175	0.00%	
180	6.38%	
185	3.33%	
190	7.83%	0.00%
195	27.03%	2.82%
200	52.17%	6.98%
205	80.34%	17.01%
210	92.00%	33.33%
215	100.00%	69.74%
220	97.44%	86.08%
225	100.00%	94.57%
230		95.56%
235		93.33%
240		100.00%

Figure 2 - Percentage of students receiving score of "proficient" on the Nevada CRT in reading by grade and RIT range

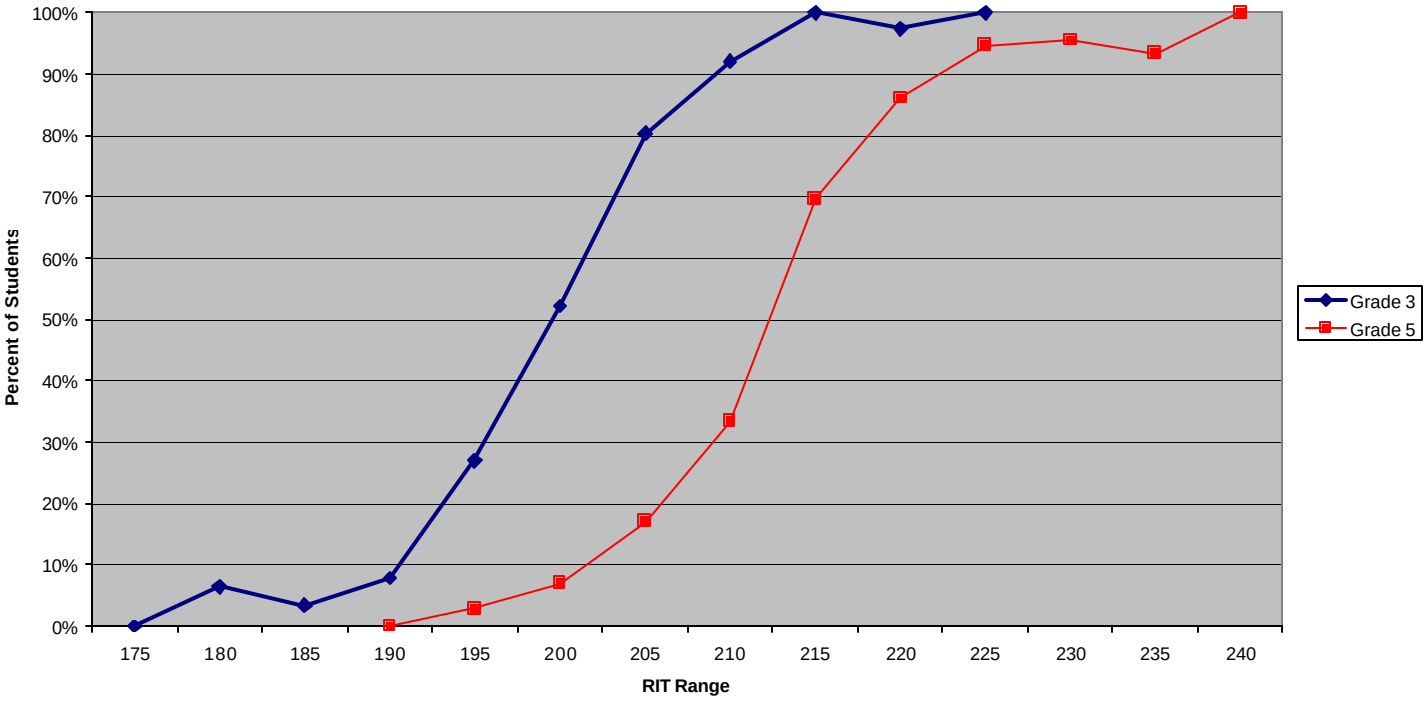
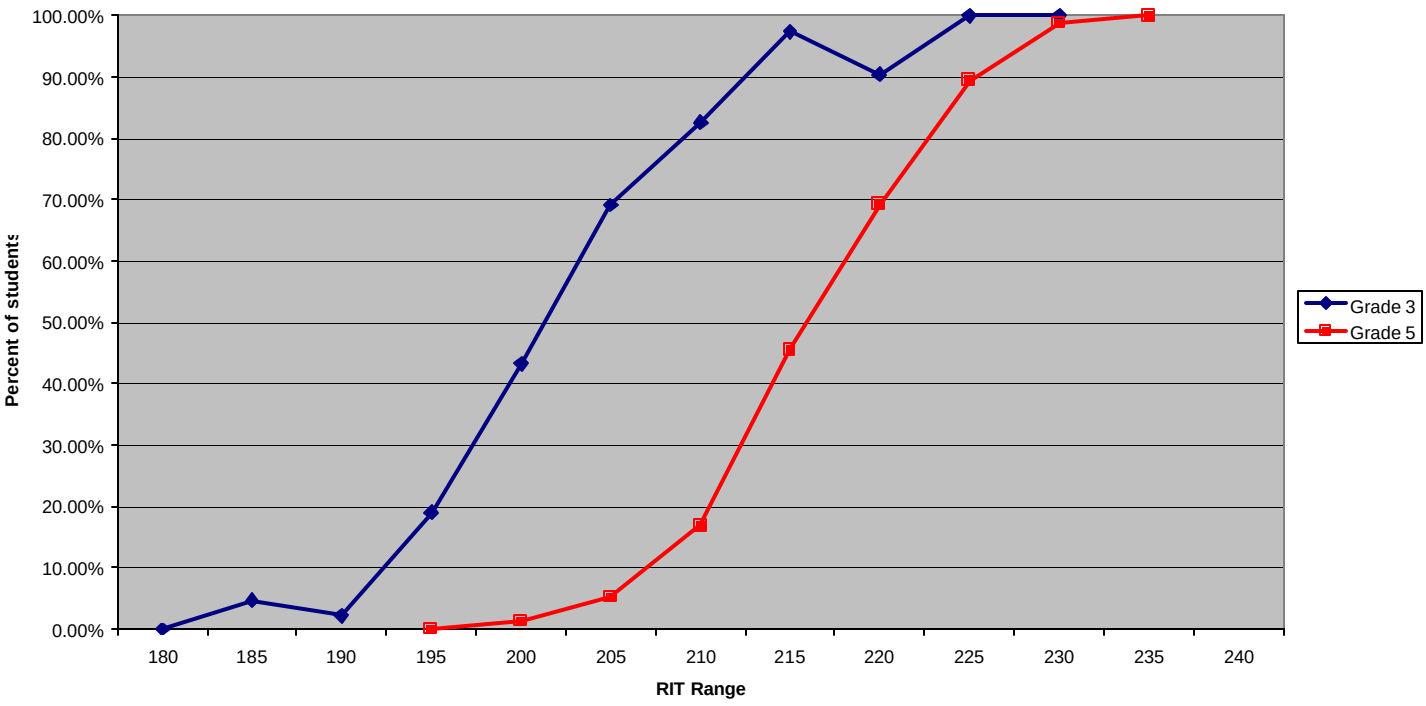


Table 15
Proportion of students passing the Nevada CRT based on same spring RIT score – Mathematics

RIT	Grade 3	Grade 5
185	4.65%	
190	2.22%	
195	19.02%	0.00%
200	43.21%	1.33%
205	69.12%	5.22%
210	82.53%	16.85%
215	97.37%	45.50%
220	90.48%	69.27%
225	100.00%	89.39%
230	100.00%	98.80%
235		100.00%

Figure 3 - Percentage of students achieving score of "proficient" on Nevada CRT in math by grade and RIT range



Comparing Nevada 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 sixteen states. To get an estimate of the difficulty of the Nevada standards in relation to other state tests, we evaluated the standard used as the cut score for *NCLB reporting or passing* 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 16. One finding is that Nevada's cut scores for the Nevada CRT and the ITBS are quite different. Because the ITBS norms do not reflect the distribution of performance in Nevada, the actual cut scores representing 50th percentile performance at grades 4 and 7 are considerably lower than the *meets* standard on the Nevada CRT, when these two are equated by RIT. These differences could be confusing to teachers and the public. For example, a 3rd grade student who we'll call Marita with a reading RIT of 197 would be labeled *approaches* on the Nevada CRT for that grade. In 4th grade, without showing any growth whatsoever, Marita would project to achieve Level 3 or *meets* performance on the ITBS in reading. However, Marita, even if he or she made normal progress between grades 4 and 5, would again be identified as *approaches* when taking the 5th grade Nevada CRT, indeed her score would fall far below the standard for *meets* at that grade. It is not clear to us whether the Iowa Testing cut scores were intended to align with the standards in place for Nevada, but the data from this study makes it clear that the cut scores do not align. Educators should be aware of this fact when making instructional decisions from their Nevada data.

One other brief note related to standards calibration. The cut score for *exceeds* performance on the 5th grade Nevada CRT in reading is considerably more difficult than the 3rd grade *exceeds* standard. Students who achieve several points over the cut point for *exceeds* in grade 3 are very likely to slip back to *meets* standards performance on the grade 5 test, even if they show growth in grades 4 and 5 that is well beyond our norms. This change is NOT a product of poor instruction of high performing readers in these grades, rather it is a result of setting a 5th grade exceeds standard that is normatively far more difficult than the corresponding 3rd grade standard.

In general Nevada's cut scores are near the average of that of most of the other states studied. We'd recommend caution about drawing any judgments about the quality of Nevada's standards from that information. 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 ***meets standard o***r proficient level of performance in about 12 years. Some communities in Nevada are no doubt close to achieving this already, but many will have to improve the performance of large proportions of their students to reach this goal. Standards should be judged on how well they align with the purposes the community has set for establishing standards, not purely on how high or low the "bar" is set. One thing the tables make clear is that proficiency standards vary widely from state to state and that proficiency is not yet a concept that has a shared definition, although greater consensus in standard setting seems to be emerging. It would be fair to say that most states that we have studied who have set standards since implementation of No Child Left Behind has begun have tended to establish standards near or below the 50th percentile on our norms.

Table 16 - 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
NV	202	58	SC	213	70	NV	215	59	CA	216	46	WA	226	67	SC	230	68	IA	224	43	WA	227	53
CA	200	51	WA	207	53	CA	214	56	MT	211	35	CA	221	50	OR	227	58	ID	221	37	ID	224	44
MN	193	35	CA	205	46	PA	212	50	ID	211	35	MT	218	43	CA	226	54	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	NV	215	35	PA	223	46				CO	209	15
MT	193	35	IA	196	26	IL	207	37	TX	208	28	ID	215	35	IN	219	35				CA	208	14
IL	193	35	NV	194	22	MT	206	35	CO	197	11	TX	210	24	MT	219	35						
IN	192	32	CO	191	18	ID	206	35				CO	206	18	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	204	63	WA	218	76	CA	225	71	CA	230	68	WA	242	78	SC	251	80	IA	241	44	MT	247	40
NV	203	59	NV	200	26	AZ	220	59	IN	221	47	CA	238	71	AZ	248	75	ID	240	42	IA	247	40
IN	201	50	SC	217	74	NV	216	48	ID	219	42	ID	225	44	CA	240	60	CO	235	32	OR	245	33
OR	199	46	CA	212	59	PA	216	48	IA	218	40	MT	224	42	PA	237	53				ID	242	25
AZ	199	46	ID	205	39	OR	215	46	MT	218	40	IA	222	38	OR	235	50				CO	233	14
MN	198	42	IA	205	39	ID	213	41	CO	207	19	TX	221	35	ID	233	46				CA	232	13
MT	197	39	MT	205	39	MT	212	38				NV	220	33	MN	231	42						
IA	197	39				IA	212	38				CO	216	26	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.
- Nevada used a criterion-referenced test at grades 3 and 5 and the ITBS at grades 4 and 7 in 2003. They will discontinue use of the ITBS in 2004.

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 *meets 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 Nevada state assessments and the RIT scales used to report performance on Northwest Evaluation Association tests. The study determined RIT score equivalents for Nevada CRT and ITBS performance levels in reading and mathematics. Nevada CRT test records for more than 2,000 students were included in this study.

Three methods generated an estimate of RIT cut scores that could be used to project Nevada CRT performance levels. Rasch SOS methods generally produced the most accurate cut score estimates. Accuracy of predicting Nevada CRT passing performance was above 84% for all grades when using the best methodology. Type I errors ranged from about 8% to 14% when the best methodology was employed.

ITBS performance levels were estimated using a regression equation adapted from previous studies to reflect the Nevada performance levels for that test.

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

References

Kingsbury, G. (1999, April) A comparison of test scores from the Iowa Test of Basic Skills and Meridian Checkpoint Assessment Level Tests. Portland, OR. Northwest Evaluation Association.

Kingsbury, G., Breithaupt, D. and Hauser, C. (2002). An empirical comparison of scores from NWEA assessments and the Iowa Test of Basic Skills. Portland, OR. Northwest Evaluation Association.