A Study of the Ongoing Alignment of the NWEA RIT Scale with the Indiana Statewide Test for Educational Progress (ISTEP+)

John Cronin and Branin Bowe

September 20, 2005

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

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

A Study of the Ongoing Alignment of the NWEA RIT Scale with the Indiana Statewide Test for Educational Progress (ISTEP+)

John Cronin and Branin Bowe

September, 2005

Each fall, Indiana students participate in testing as part of the state's assessment program. Students in grades grades 3 through 9 take the Indiana Statewide Test for Educational Progress (ISTEP+) in English/Language Arts and Mathematics. These tests serve as an important measure of student achievement for the state's accountability system. Results from these assessments are used to make state-level decisions concerning education, to meet *Adequate Yearly Progress* (AYP) reporting requirements of the *No Child Left Behind Act* (NCLB), and to inform schools and school districts of their performance. The Indiana Department of Education has developed scales that are used to assign students to one of three performance levels on these tests.

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

In order to use the two testing systems to support each other, an alignment of the scores from the state and RIT-based tests is as important as curriculum alignment. A 2003 study first established estimated RIT scores that aligned with the equivalent cut points on the ISTEP+ scale (Cronin, 2003). Because Indiana expanded the number of grades tested in fall of 2004, we undertook a study to estimate the aligned cut scores for the grades added and attempted to determine whether previous estimates of cut scores had changed. We estimated the relative accuracy with which the NWEA assessments continued to predict ISTEP+ results. Finally, we developed estimates for both the fall and prior spring RIT scores so schools may use spring results to assess their students' likelihood of success on ISTEP. The primary questions addressed in this study are:

- What RIT scores correspond to various performance levels on the ISTEP+ tests?
- How do these RIT scores differ from the 2003 estimates of performance levels?
- How well can performance on the Indiana assessments be predicted from RIT scores when NWEA assessments are administered in the same fall and the prior spring?

Method

Our study included test records from over 25,500 students enrolled in 11 Indiana school systems. These students had taken both the state assessment and NWEA assessments in fall of 2004; many had also taken NWEA assessments in spring of 2004. Student records were included when a student had both a valid NWEA scale score and a valid ISTEP+ score in the equivalent subject for the fall season. We excluded records in which students had been given accommodations on the state assessment.

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

(www.nwea.org/research/national.asp) , for more detail about the methods we use to conduct scale alignment studies.

Results

Descriptive Statistics

Table 1 reviews descriptive statistics for the ISTEP+ and NWEA assessments. The median RIT scores for this sample in reading average about 3 points below the median for the 2005 NWEA norm population sample. In mathematics, the median RIT ranges from 3 to 5 points below the 2005 norm median. The distributions in all subjects show some evidence of a negative skew. Nevertheless, the sample provides reasonable numbers of students who perform at all levels on the test scales and this assures that the statistical methods applied have an adequately large sample to derive good estimates of performance levels that are at the higher and lower ends of a test scale.

Pearson correlations

Table 2 shows the Pearson correlations for each grade. Concurrent validity was tested by examining same subject Pearson correlations between the NWEA and ISTEP+ assessments. ISTEP+ ELA to NWEA reading coefficients were very high, ranging between .75 and .84 for testing during the same season, and between .74 and .82 when NWEA testing occurred in the prior spring. Correlations between the two NWEA reading assessments were slightly stronger, as expected, with coefficients ranging between .80 and .84. Language usage correlations with the ISTEP+ ELA were also strong with same season correlations ranging from .74 to .82 for testing during the same season and for NWEA that occurred during the prior spring. Correlations between the two NWEA language usage assessments ranged from .80 to .86. In mathematics, correlations between the ISTEP+ and the NWEA mathematics assessments were also high. Correlations between ISTEP and NWEA mathematics ranged from .75 to .88 for both same season administrations and for NWEA administrations that occurred during the prior spring. Correlations between the two NWEA mathematics assessments were very high, ranging from .80 to .90. The strength of the correlations among the assessments suggests that tests are measuring the same general constructs.

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

ISTEP + ELA							
		Fall 2004					
Grade	N*	Mean	Median	SD			
3	2490	445.17	443	64.12			
4	3013	456.57	458	62.07			
5	4030	473.14	476	61.78			
6	4242	489.954	491	57.45			
7	4226	511.70	514	70.86			
8	4336	528.90	531	58.81			
9	3152	532.02	536	67.06			
NWEA Reading							
	Fall 2004			Spring 2004			
Grade	Mean	Median	SD	N	Mean	Median	SD
3	189.96	192	15.28	2189	189.59	191	14.36
4	197.08	200	16.08	2766	197.67	200	14.94
5	203.22	205	15.11	3716	203.27	205	14.83
6	207.77	210	15.29	3886	208.59	210	14.08
7	211.06	213	15.72	3788	212.04	214	14.94
8	215.28	217	15.56	3939	215.62	217	15.25
9	218.42	220	15.03	2786	218.95	220	14.49
NWEA Language Usage							
	Fall 2004			Spring 2004			
Grade	Mean	Median	SD	N	Mean	Median	SD
3	194.21	197	14.44	2166	194.28	196	13.99
4	201.13	203	14.62	2765	201.14	203	14.33
5	206.58	208	13.57	3719	206.58	208	13.56
6	210.72	213	13.97	3887	211.21	213	13.00
7	213.30	215	14.02	3785	213.76	216	13.70
8	216.84	219	13.46	3945	216.84	219	13.75
9	219.15	221	12.84	2812	219.02	221	12.91
ISTEP Mathematics							
		Fall 2004					
Grade	N	Mean	Median	SD			
3	2492	425.75	425	61.81			
4	3000	442.16	443	59.52			
5	4001	450.81	454	62.63			
6	4208	490.84	493	62.67			
7	4105	503.14	506	72.06			
8	4262	543.35	545	80.18			
9	3096	556.22	558	72.28			
NWEA Mathematics							
	Fall 2004			Spring 2004			
Grade	Mean	Median	SD	N	Mean	Median	SD
3	193.41	195	11.52	2246	191.89	193	11.01
4	201.47	202	12.66	2745	201.12	201	12.01
5	208.38	209	13.54	3671	208.06	208	13.41
6	215.28	216	15.07	3832	216.31	217	14.57
7	220.56	221	16.43	3699	221.05	222	16.37
8	226.82	228	17.35	3870	226.94	227	17.97
9	229.95	231	17.12	2754	231.12	2321	17.01

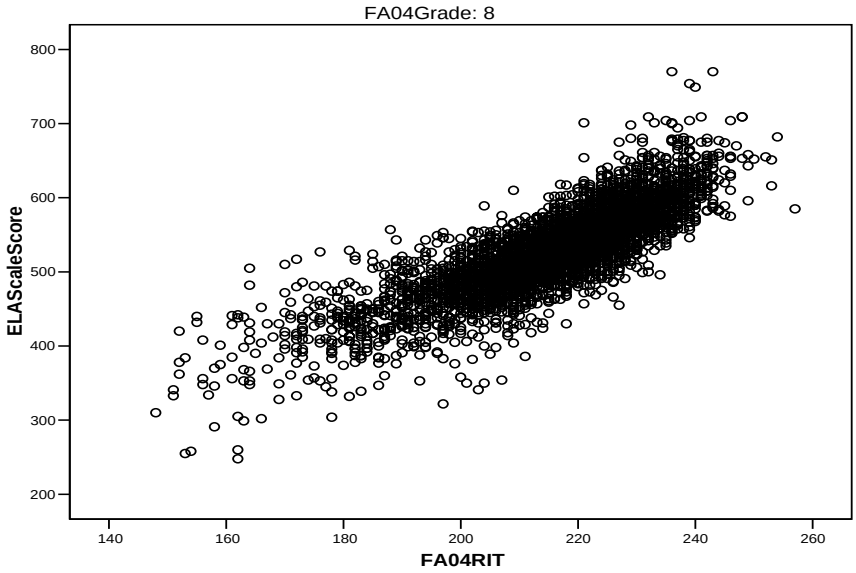
Table 2 – Inter-test Correlations for ISTEP+ and NWEA assessments by Subject

	English/Language Arts			Mathematics		
Grade	ISTEP+ - Fall Reading RIT	ISTEP+ - Spring Reading RIT	Fall Reading RIT – Spring Reading RIT	ISTEP+ - Fall RIT	ISTEP+ - Spring RIT	Fall RIT – Spring RIT
3	.75	.74	.82	.75	.75	.80
4	.82	.80	.83	.81	.77	.80
5	.80	.79	.83	.82	.79	.84
6	.82	.80	.82	.85	.83	.85
7	.82	.81	.84	.87	.86	.88
8	.84	.82	.83	.88	.87	.90
9	.80	.81	.80	.87	.88	.87

	English/Language Arts		
Grade	ISTEP+ - Fall Lang. Usage RIT	ISTEP+ - Spring Lang. Usage RIT	Fall Lang. Usage RIT – Spring Lang. Usage RIT
3	.74	.74	.83
4	.81	.78	.82
5	.78	.76	.80
6	.81	.78	.82
7	.82	.80	.84
8	.82	.82	.86
9	.82	.80	.83

In general, scatterplots showed that relationships between NWEA and ISTEP+ scores were linear with some evidence of mild dispersion at the low end of the scales. Because state assessments are typically designed to generate estimates of performance using the grade level standards and content, some of the dispersion effect may be a product of the limitations inherent in a grade level test's ability to deliver items that accurately measure students in the extremes of the performance range. In addition, the level of motivation to take tests differs from day-to-day, especially for low performers, and this probably also explains some of the dispersion of scores at the low end of the scale. Figure 1 shows an example that is illustrates both the strength of the linear correlation and the issue of dispersion.

Figure 1 – Grade 8 Reading ISTEP+ ELA score plotted against Reading RIT score



Linking ISTEP+ performance level cut scores to the RIT scale

The primary purpose of this study was to generate new estimates of the RIT scale scores that most closely correspond to the cut scores for different performance levels on the ISTEP+. 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.

Our alignment studies employ three methods to estimate cut scores, linear regression, second order regression, and a Rasch status on standards (Rasch SOS) method that estimates cut scores using a design based in item-response theory.

Tables 3 and 4 show several estimations of the Fall and Spring 2004 RIT scores that correspond to the cut scores for the various performance levels on the ISTEP+ scales. As a rule the three methodologies came to similar estimates of cut scores for each of the performance levels, although the Rasch SOS methodology did produce somewhat higher estimates of the RIT score required to meet the ***pass*** standard at each grade. The Rasch SOS methodology also produced lower estimates of the RIT score required to meet the ***pass*** **+** standard.

Table 3 – Estimated points on the RIT scale equating to the minimum scores (rounded) for performance levels on the ISTEP+ based on FALL testing

	Grade 3								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<176	176	212	<176	176	210	<179	179	206
Language - ELA	<180	180	215	<181	181	213	<184	184	210
Mathematics	<184	184	210	<184	184	210	<186	186	206
	Grade 4								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<188	188	221	<189	189	218	<191	191	215
Language ELA	<192	192	223	<194	194	221	<195	195	218
Mathematics	<194	194	219	<195	195	219	<196	196	216
	Grade 5								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<194	194	228	<194	194	224	<197	197	222
Language - ELA	<198	198	229	<199	199	227	<200	200	224
Mathematics	<202	202	228	<202	202	228	<203	203	225
	Grade 6								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<201	201	234	<203	203	231	<204	204	228
Language - ELA	<205	205	235	<206	206	230	<207	207	228
Mathematics	<208	208	234	<208	208	234	<209	209	232
	Grade 7								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<207	207	237	<208	208	233	<208	208	232
Language - ELA	<209	209	236	<211	211	233	<211	211	231
Mathematics	<215	215	241	<215	215	241	<215	215	240
	Grade 8								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<211	211	242	<211	211	237	<213	213	236
Language - ELA	<213	213	239	<214	214	237	<215	215	234
Mathematics	<221	221	247	<222	222	247	<221	221	246
	Grade 9								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<214	214	248	<216	216	242	<216	216	239
Language - ELA	<216	216	243	<216	216	239	<217	217	237
Mathematics	<226	226	253	<227	227	252	<226	226	250

Table 3 – Estimated points on the RIT scale equating to the minimum scores (rounded) for performance levels on the ISTEP+ based on prior SPRING testing

	Grade 3								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<176	176	211	<176	176	208	<181	181	205
Language - ELA	<181	181	215	<181	181	212	<184	184	210
Mathematics	<183	183	208	<183	183	207	<185	185	204
	Grade 4								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<189	189	221	<190	190	219	<192	192	215
Language ELA	<192	192	223	<194	194	221	<195	195	221
Mathematics	<194	194	218	<193	193	217	<196	196	215
	Grade 5								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<194	194	228	<194	194	224	<198	198	223
Language - ELA	<197	197	226	<198	198	226	<200	200	224
Mathematics	<201	201	228	<200	200	227	<202	202	225
	Grade 6								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<203	203	233	<204	204	230	<205	205	227
Language - ELA	<205	205	235	<207	207	232	<208	208	228
Mathematics	<208	208	235	<208	208	234	<210	210	232
	Grade 7								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<208	208	236	<209	209	234	<210	210	231
Language - ELA	<210	210	236	<211	211	234	<212	212	231
Mathematics	<215	215	242	<216	216	242	<215	215	241
	Grade 8								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<210	210	241	<211	211	238	<213	213	235
Language - ELA	<212	212	240	<214	214	237	<215	215	234
Mathematics	<220	220	248	<220	220	247	<220	220	246
	Grade 9								
	Linear			Second-order regression			Rasch SOS		
	NP	P	P+	NP	P	P+	NP	P	P+
Reading - ELA	<215	215	246	<216	216	242	<217	217	245
Language - ELA	<215	215	244	<216	216	239	<218	218	237

Mathematics	<227	227	254	<228	228	253	<227	227	252

Establishing RIT score estimates for ISTEP+ performance levels.

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

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

- **No Pass and Pass.** We calculated a *prediction index* statistic for the proposed *pass* cut score. This is calculated as 1 – (correct predictions/type I errors). A test with a high prediction index statistic typically reflects both a high rate of accuracy and a low rate of Type I errors. We generally selected the method that produced the highest prediction index number.

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

Tables 4 and 5 show the recommended RIT cut scores for each of the ISTEP+ performance levels. In general, Rasch SOS methods were most reliable for establishing predictive cut scores.

Table 4 – Recommended FALL RIT cut scores for ISTEP+ ELA performance levels – Reading

	Not Passing		Pass			Pass +			
Grade	Score	% of non-passing students found	Score	Method	% correct for P/F prediction	Prediction Index*	Score	Method	% of students ID
3	<179	62.0%	179	R	86.9%	.917	212	R	50.3%
4	<191	68.5%	191	R	86.2%	.910	221	R	58.6%
5	<197	61.5%	197	R	83.8%	.892	228	R	51.1%
6	<204	70.4%	204	R	83.5%	.899	234	R	50.3%
7	<208	67.0%	208	S, R	82.1%	.868	237	R	51.4%
8	<213	71.4%	213	R	83.9%	.888	242	R	50.0%
9	<216	71.5%	216	S,R	82.3%	.877	248	R	52.4%

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

Table 5 – Recommended PRIOR SPRING RIT cut scores for ISTEP+ ELA performance levels – Reading

	Not Passing		Pass			Pass +			
Grade	Score	% of non-passing students found	Score	Method	% correct for P/F prediction	Prediction Index*	Score	Method	% of advanced students found
3	<181	63.1%	181	R	86.5%	.917	205	R	47.8%
4	<192	63.1%	192	R	84.1%	.910	215	R	53.0%
5	<198	63.4%	198	R	83.8%	.892	223	R	41.6%
6	<205	66.7%	205	R	82.9%	.899	227	R	45.2%
7	<210	69.5%	210	R	82.2%	.868	231	R	52.1%
8	<213	70.6%	213	R	82.1%	.877	235	R	52.6%
9	<216	70.3%	216	R	82.3%	.854	243	R	31.7%

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

Table 6 – Recommended FALL RIT cut scores for ISTEP+ ELA performance levels – Language Usage

	Not Passing		Pass			Pass +			
Grade	Score	% of non-passing students found	Score	Method	% correct for P/F prediction	Prediction Index*	Score	Method	% of advanced students found
3	<184	64.6%	184	R	87.2%	.923	210	R	48.4%
4	<195	67.9%	195	R	84.8%	.907	218	R	57.1%
5	<200	59.3%	200	R	82.5%	.884	224	R	50.5%
6	<208	66.6%	208	R	80.8%	.883	228	R	47.6%
7	<211	69.9%	211	S, R	82.1%	.880	231	R	49.0%
8	<215	72.1%	215	R	82.5%	.884	234	R	49.0%
9	<217	73.4%	217	R	82.0%	.897	237	R	54.3%

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

Table 7 – Recommended PRIOR SPRING RIT cut scores for ISTEP+ ELA performance levels – Language Usage

	Not Passing		Pass			Pass +			
Grade	Score	% of non-passing students found	Score	Method	% correct for P/F prediction	Prediction Index*	Score	Method	% of advanced students found
3	<181	50.1%	181	L,S,R	75.5%	.892	210	R	48.9%
4	<194	57.0%	194	S	76.5%	.854	221	R	56.2%
5	<198	49.9%	198	S	76.1%	.846	224	R	46.3%
6	<207	60.0%	207	S	75.1%	.833	228	R	48.5%
7	<211	66.5%	211	S	74.0%	.856	231	R	49.0%
8	<214	67.8%	214	S	75.9%	.843	234	R	49.7%
9	<216	64.8%	216	S	72.6%	.841	237	R	52.9%

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

Table 8 – Recommended FALL RIT cut scores for ISTEP+ performance levels – Mathematics

	Not Passing		Pass			Pass +			
Grade	Score	% of non-passing students found	Score	Method	% correct for P/F prediction	Prediction Index*	Score	Method	% of advanced students found
3	<186	58.8%	186	R	83.4%	.887	206	R	52.5%
4	<196	66.3%	196	R	84.8%	.897	216	R	59.1%
5	<203	67.6%	203	R	83.6%	.896	225	R	65.3%
6	<209	67.4%	209	R	85.4%	.900	232	R	68.9%
7	<215	75.6%	215	L,S,R	87.1%	.917	240	R	73.7%
8	<222	79.7%	222	S	87.4%	..929	246	R	75.2%
9	<227	82.1%	227	S	76.2%	.923	250	R	79.0%

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

Table 9 – Recommended PRIOR SPRING RIT cut scores for ISTEP+ performance levels – Mathematics

	Not Passing		Pass			Pass +			
Grade	Score	% of non-passing students found	Score	Method	% corr for P/F prediction	Prediction Index*	Score	Method	% of advanced students found
3	<185	56.9%	185	R	81.9%	.879	204	R	51.5%
4	<196	64.1%	196	R	82.5%	.888	215	R	60.4%
5	<202	63.6%	202	R	82.4%	.883	225	R	62.5%
6	<210	70.2%	210	R	84.5%	.908	232	R	69.8%
7	<216	78.0%	216	S	86.1%	.925	241	R	69.7%
8	<220	75.9%	220	L,S,R	86.5%	.918	246	R	74.5%
9	<228	82.4%	228	S	86.5%	.925	252	R	75.4%

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

We evaluate the relative accuracy of state alignment study results by comparing the prediction index statistics generated by these studies for their accuracy in assessing proficiency status and performance level. Table 10 summarizes the accuracy of proficiency status prediction for this study relative to other state alignment studies and Table 11 summarizes the accuracy of performance level prediction. The results show that the prediction index statistics for proficiency status are low when compared to other state studies, but in line with the prior Indiana studies.

It's very likely that the prediction index statistics for the ISTEP+ ELA test are influenced by the fact that the NWEA reading and language usage tests were used independently to predict ISTEP results. Since neither test standing alone is fully aligned with the content, this had a slight effect on the strength of correlations and also may have had a slight effect on prediction accuracy. It should also be noted that the accuracy of prediction is also affected by the placement of the standard. States with the highest prediction index scores for proficiency, Texas for example, also tended to have low standards relative to the other states. The reason is that if a standard is set at a very low level, most students will be performing so far above the standard that there is no possibility of a missed prediction. If the standard is set in the middle of a normal distribution, however, a large group of students performs very close to the cut score which increases the probability of missed predictions significantly.

Nevertheless, the Indiana index statistics showed rates of correct prediction for proficiency that were consistently above 80% for scores collected in the same season and ratios of correct prediction to Type I error that ranged from about 6 to 1 to nearly 14 to 1.

Table 10 – Prediction Indices (Based on Proficiency Status) for Previous NWEA State Alignment Studies

State	Reading	State	Language	State	Math
Texas	.967*	Texas	.968*	Texas	.969*
Minnesota	.944*	South Carolina Exit	.938*	Wyoming	.961
South Carolina Exit	.940*	California	.913*	Colorado '01	.957
Pennsylvania	.935*	Indiana '01	.907*	Illinois	.946*
Wyoming	.931	Colorado '03	.903*	Colorado '03	.943*
Colorado '03	.931*	Indiana '03	.894*	South Carolina '03	.943*
Illinois	.928*	**Indiana '05**	**.891**	Minnesota	.936*
California	.925*	South Carolina '04	.889*	South Carolina Exit	.933*
Arizona	.912*	Arizona	.874*	Pennsylvania	.926*
Colorado '01	.910*			Washington '99	.920
Nevada	.902*			Arizona	.919*
South Carolina '03	.902*			South Carolina '04	.914*
Indiana '01	.902*			Washington '04	.912*
Indiana '03	.900*			California	.910*
Washington '99	.893			**Indiana '05**	**.906**
Indiana '05	**.892**			Indiana '01	.899*
Washington '04	.886*			Nevada	.866*
South Carolina '04	.884*			Indiana '03	.860*

Table 11 – Prediction index scores by performance level assignment for previous NWEA state alignment Studies

State	Reading	State	Math
Texas	.868	Texas	.900
Indiana '05	**.867**	Illinois	.888*
Indiana '03	.860	**Indiana '05**	**.863**
Colorado	.840	Colorado	.808
Illinois	.804*	Indiana '03	.804*
Nevada	.776*	Pennsylvania	.769*
Pennsylvania	.770*	South Carolina '03	.764*
South Carolina '03	.757*	Arizona	.726*
Arizona	.756*	Nevada	.742*
South Carolina '04	.717*	South Carolina '04	.741*
Washington '04	.667	Washington '04	.721
South Carolina Exit	.649*	South Carolina Exit	.705*
Minnesota	.627*	Minnesota	.611*
California	.600*	California	.565*

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

Although the predicted RIT cut scores can help teachers and students establish targets for NWEA assessments that can help assure success on the state test, teachers should be aware that students performing near the proficient cut score on the RIT scale have only about a 50% probability of passing the ISTEP+. The information in Tables 12 through 17 provide educators with more precise data related to students' probabilities of achieving proficiency.

These tables show the proportion of students at each 5 point RIT level who earned scores at or above the *proficient* level on their respective ISTEP+ assessment both when the NWEA test was administered in the same season as the state test (fall) and also when the NWEA test was administered during the prior spring. Using fall reading (Table 12) as an example, we find that about 27% of the Grade 4 students who achieved a reading RIT score between 180 and 184 went on to achieve a passing score on the ISTEP+ assessment. A reading teacher would know that only about one in four of students performing in this range in fall is likely to achieve a proficient score on the ISTEP+ unless they work harder, receive more focused instruction, or have access to additional resources.

On the other hand, about 99% of students who scored between RITs of 210 and 214 achieved proficiency on the Indiana assessment. Teachers should feel free to focus their efforts with these students on content and skills that go beyond the minimum expectations for performance.

Figures 3 through 8 are graphic depictions of the data in the tables.

Table 12 – Proportion of students passing the ISTEP+ ELA based on FALL RIT reading score

RIT	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9
150	0.00%	0.00%	0.00%				
155	5.41%	7.69%	14.29%				
160	11.11%	9.52%	18.18%	0.00%			
165	22.22%	6.38%	27.27%	11.11%			
170	50.49%	14.55%	12.50%	15.00%			
175	52.23%	14.14%	25.00%	14.71%	0.00%		
180	70.81%	27.01%	23.46%	17.02%	15.56%		
185	80.66%	39.89%	28.21%	9.80%	8.45%		
190	93.69%	60.08%	37.00%	20.43%	7.69%	6.94%	
195	96.36%	78.09%	58.17%	29.15%	16.06%	7.63%	3.80%
200	99.08%	92.07%	72.69%	45.61%	26.77%	16.95%	13.43%
205	100.00%	97.22%	88.66%	70.81%	48.63%	24.86%	20.74%
210		98.68%	95.37%	85.69%	68.69%	46.50%	29.10%
215		100.00%	98.41%	94.01%	85.50%	72.87%	59.05%
220			99.68%	98.83%	93.35%	90.08%	77.05%
225			100.00%	100.00%	97.51%	94.52%	88.16%
230					99.14%	99.17%	97.81%
235					100.00%	100.00%	99.60%
240							100.00%

Table 13 – Proportion of students passing the ISTEP+ ELA based on PRIOR SPRING RIT reading score

RIT	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9
145	0.00%						
150	6.67%						
155	20.83%						
160	23.08%	0.00%					
165	24.44%	7.41%					
170	33.73%	10.81%					
175	43.70%	18.18%	8.47%				
180	70.00%	23.48%	21.95%	9.76%			
185	83.70%	47.09%	29.66%	9.38%	4.17%		
190	91.67%	58.33%	38.57%	14.53%	4.17%	5.36%	0.00%
195	97.40%	75.78%	58.48%	20.25%	14.38%	12.30%	7.25%
200	97.86%	88.56%	77.04%	44.65%	25.77%	12.87%	5.50%
205	98.77%	96.20%	87.52%	67.66%	42.86%	33.70%	18.57%
210	100.00%	98.25%	93.85%	81.95%	65.65%	47.58%	31.73%
215		99.47%	97.44%	93.36%	82.23%	75.69%	57.81%
220		100.00%	100.00%	98.39%	92.93%	88.45%	76.20%
225				98.76%	97.75%	96.11%	90.81%
230				100.00%	100.00%	99.43%	96.17%
235						100.00%	99.05%
240							100.00%

Table 14 – Proportion of students passing the ISTEP+ ELA based on FALL Language Usage score

RIT	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9
150	9.09%						
155	11.63%						
160	25.00%						
165	19.44%	0.00%					
170	42.22%	15.69%	8.00%	4.35%			
175	47.71%	19.51%	22.58%	8.57%	0.00%	0.00%	
180	78.80%	30.25%	23.77%	12.82%	3.03%	9.52%	
185	85.02%	47.09%	27.67%	16.13%	11.27%	8.33%	4.35%
190	91.62%	61.59%	49.29%	19.47%	12.03%	9.64%	8.16%
195	97.63%	83.29%	60.65%	38.20%	14.06%	11.83%	6.32%
200	99.02%	93.85%	80.49%	56.39%	36.03%	20.74%	15.02%
205	98.88%	99.10%	93.92%	74.36%	57.75%	39.53%	27.12%
210	100.00%	100.00%	98.24%	89.41%	82.06%	67.95%	52.78%
215			99.40%	97.97%	90.66%	84.90%	75.42%
220			99.53%	99.21%	96.86%	94.31%	92.83%
225			98.18%	100.00%	99.28%	98.79%	98.59%
230			100.00%		100.00%	99.46%	100.00%
235						100.00%	

Table 15– Proportion of students passing the ISTEP+ ELA based on PRIOR SPRING RIT Language Usage score

RIT	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9
150							
155	7.69%						
160	23.53%						
165	15.38%	4.55%	0.00%				
170	31.15%	6.06%	36.36%				
175	35.23%	19.57%	24.00%	5.26%	0.00%		
180	49.64%	21.25%	30.23%	12.50%	13.33%	0.00%	0.00%
185	69.31%	32.86%	27.88%	7.69%	16.67%	13.33%	13.33%
190	87.76%	49.21%	28.13%	12.90%	8.00%	5.71%	5.71%
195	90.55%	63.41%	42.91%	15.89%	11.11%	6.58%	6.58%
200	97.87%	79.68%	62.15%	34.60%	21.40%	11.24%	11.24%
205	99.59%	91.69%	84.49%	52.35%	30.03%	21.55%	21.55%
210	100.00%	97.90%	90.72%	74.25%	58.86%	40.14%	40.14%
215		98.94%	96.53%	89.66%	79.46%	69.81%	69.81%
220		100.00%	98.58%	96.69%	89.77%	83.16%	83.16%
225			99.49%	97.94%	96.44%	96.81%	96.81%
230			100.00%	100.00%	98.68%	98.39%	98.39%
235					100.00%	100.00%	100.00%

Table 16– Proportion of students passing the ISTEP+ mathematics test based on FALL RIT Mathematics Score

RIT	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9
150							
155	0.00%						
160	23.08%						
165	23.53%	0.00%					
170	20.00%	17.14%					
175	27.83%	14.63%	8.00%				
180	42.31%	19.10%	29.27%				
185	59.47%	19.15%	20.83%	0.00%			
190	79.82%	35.16%	22.10%	5.19%	4.26%	3.23%	0.00%
195	90.87%	57.75%	30.03%	17.42%	9.18%	6.67%	3.45%
200	97.60%	82.67%	48.91%	22.39%	11.17%	3.74%	1.89%
205	100.00%	96.13%	72.97%	46.83%	16.54%	8.84%	6.09%
210		99.73%	90.00%	65.32%	35.17%	13.52%	6.17%
215		100.00%	98.58%	85.69%	64.07%	29.66%	15.83%
220			99.42%	96.31%	84.98%	54.98%	31.56%
225			99.56%	99.56%	93.95%	79.62%	50.00%
230			100.00%	100.00%	99.25%	92.29%	74.74%
235					100.00%	99.29%	91.64%
240						100.00%	98.76%
245							99.58%
250							100.00%

Table 17 Proportion of students passing the ISTEP+ mathematics test based on PRIOR SPRING Mathematics RIT score

RIT	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9
150	0.00%						
155	14.29%						
160	16.67%						
165	19.05%						
170	29.63%						
175	33.05%	14.63%					
180	45.83%	13.75%	7.50%				
185	62.33%	27.69%	15.79%	8.00%			
190	83.20%	40.73%	31.72%	12.50%			0.00%
195	94.43%	58.13%	33.22%	18.18%	3.45%		3.57%
200	98.04%	81.70%	51.79%	26.24%	15.61%	6.48%	2.22%
205	100.00%	97.54%	75.99%	39.33%	20.65%	13.44%	2.67%
210		98.92%	90.58%	65.98%	36.28%	19.40%	8.84%
215		99.40%	97.10%	82.41%	60.00%	39.57%	13.66%
220		100.00%	99.13%	93.52%	84.01%	59.50%	24.80%
225			100.00%	99.14%	93.25%	79.72%	54.30%
230				99.40%	99.74%	94.12%	66.43%
235				100.00%	99.66%	98.71%	94.10%
240					100.00%	99.68%	97.23%
245						99.63%	99.53%
250						100.00%	99.41%
255							100.00%

Figure 3 – Percent of Students Passing ISTEP+ ELA by FALL Reading RIT Performance Range

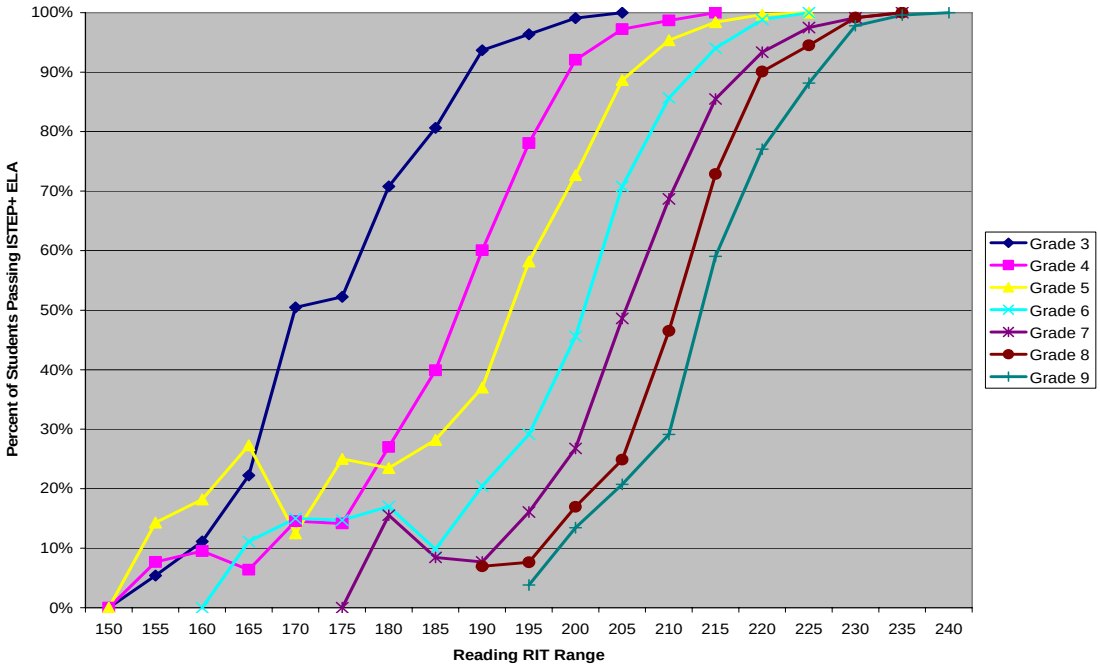


Figure 4 – Percent of Students Passing ISTEP+ ELA by PRIOR SPRING Reading RIT Performance Range

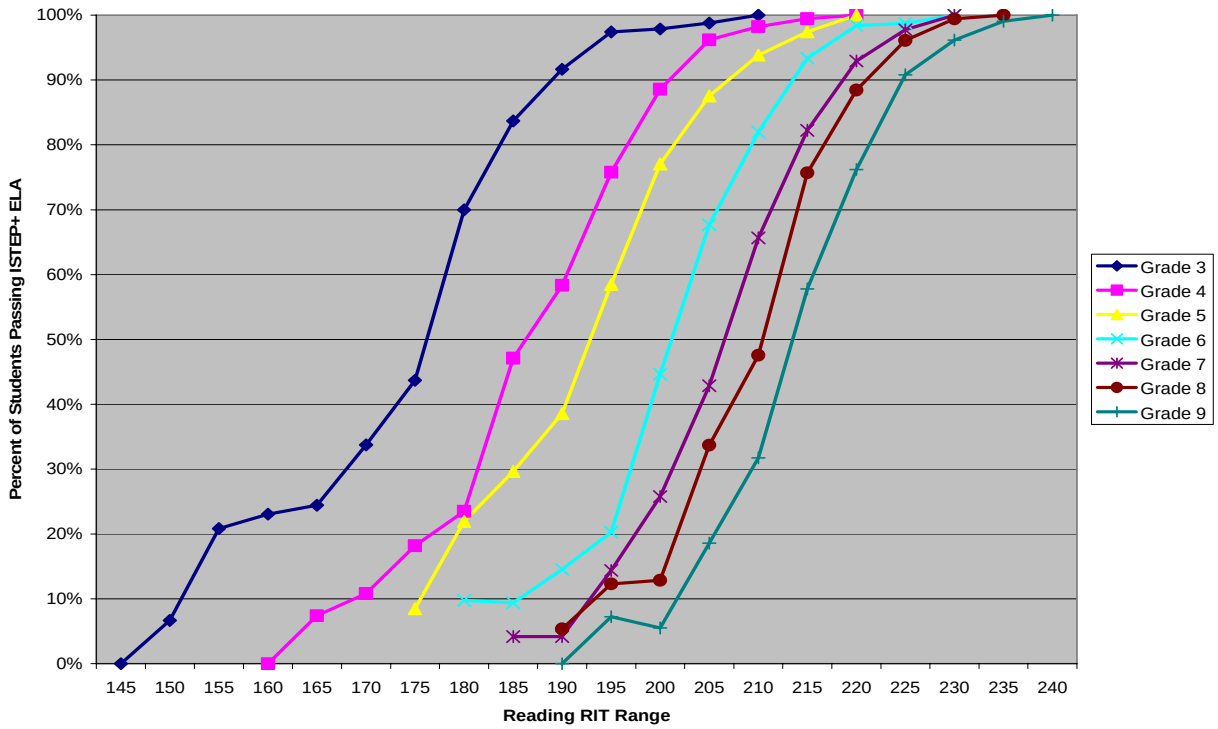


Figure 5 – Percent of Students Passing ISTEP+ ELA by FALL Language Usage RIT Performance Range

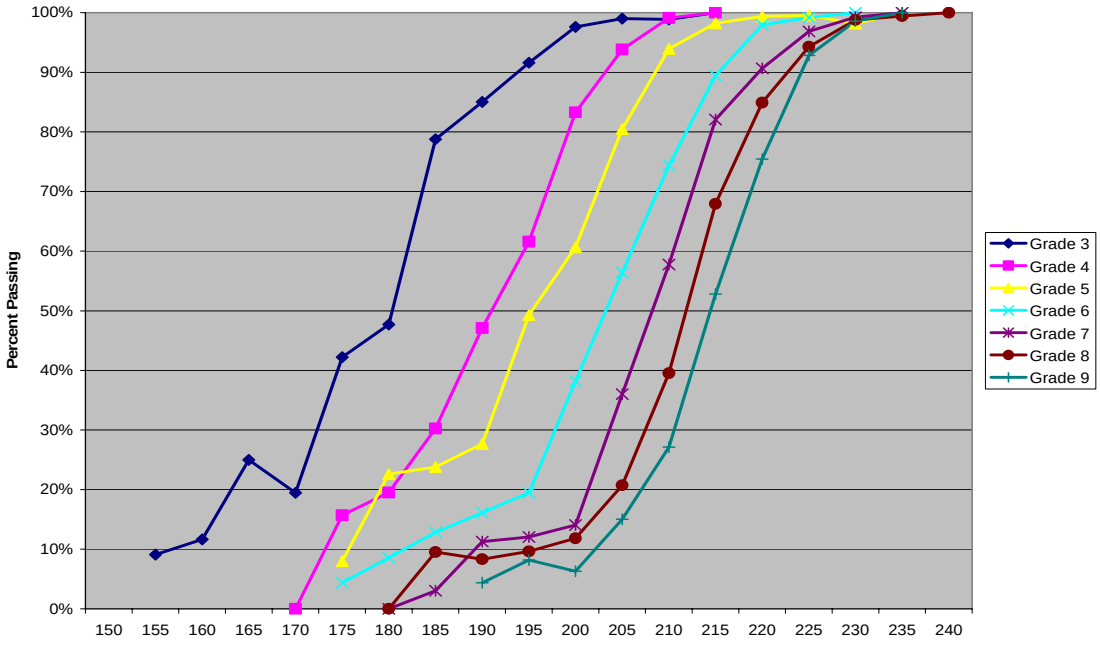


Figure 6 – Percent of Students Passing ISTEP+ ELA by PRIOR SPRING Language Usage RIT Performance Range

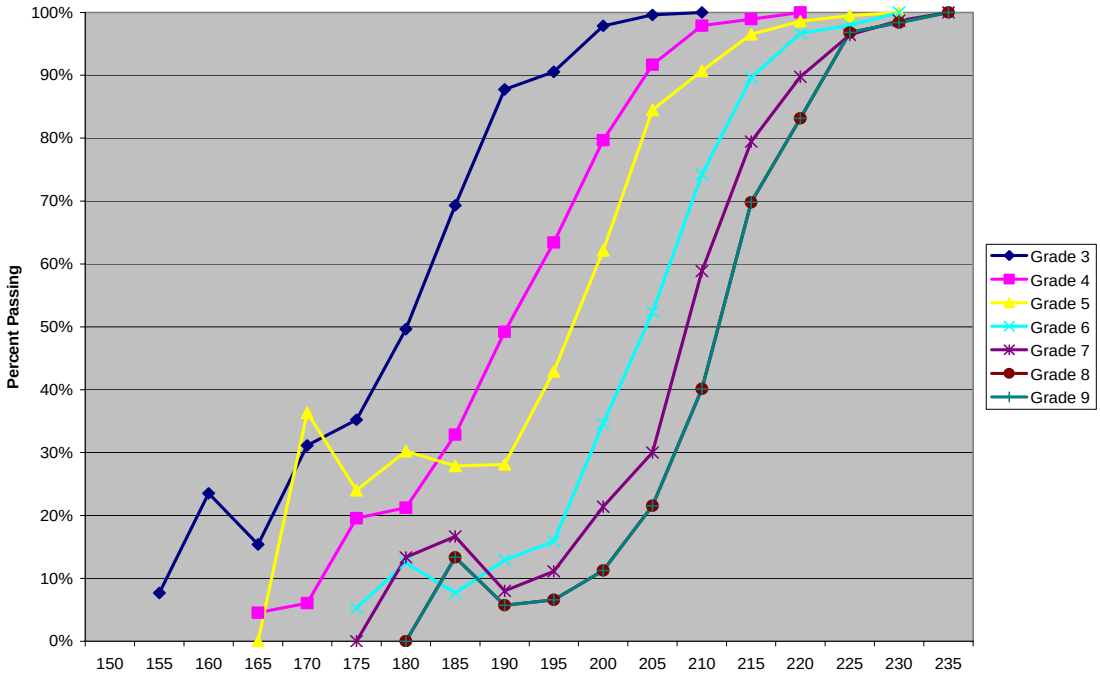


Figure 7 – Percent of Students Passing ISTEP+ Mathematics by FALL Mathematics RIT Performance Range

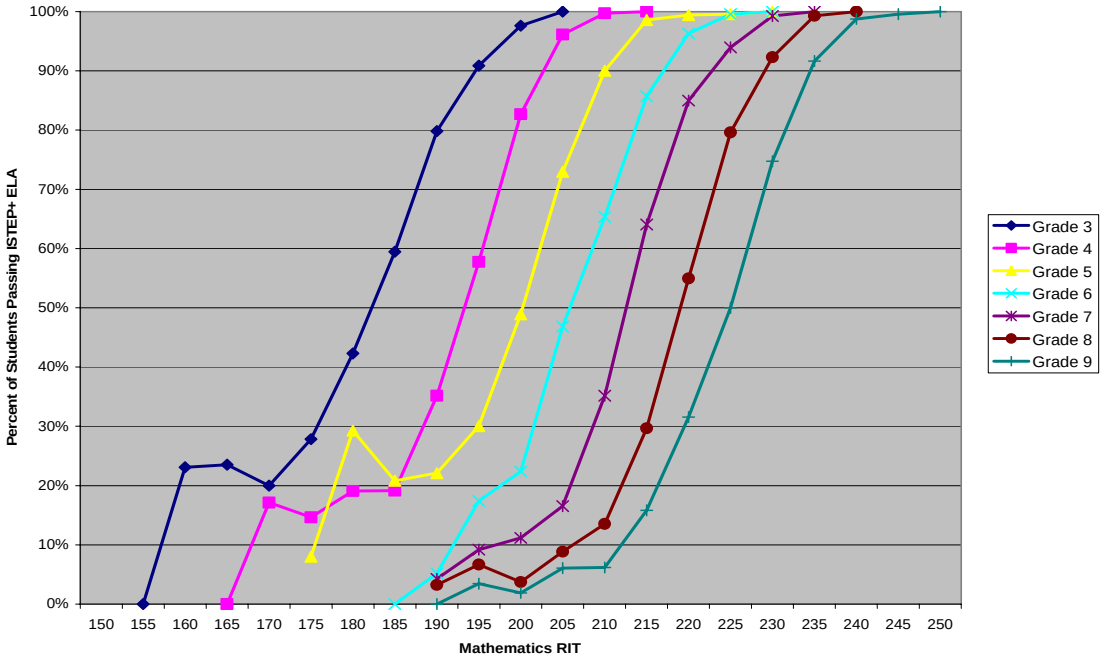
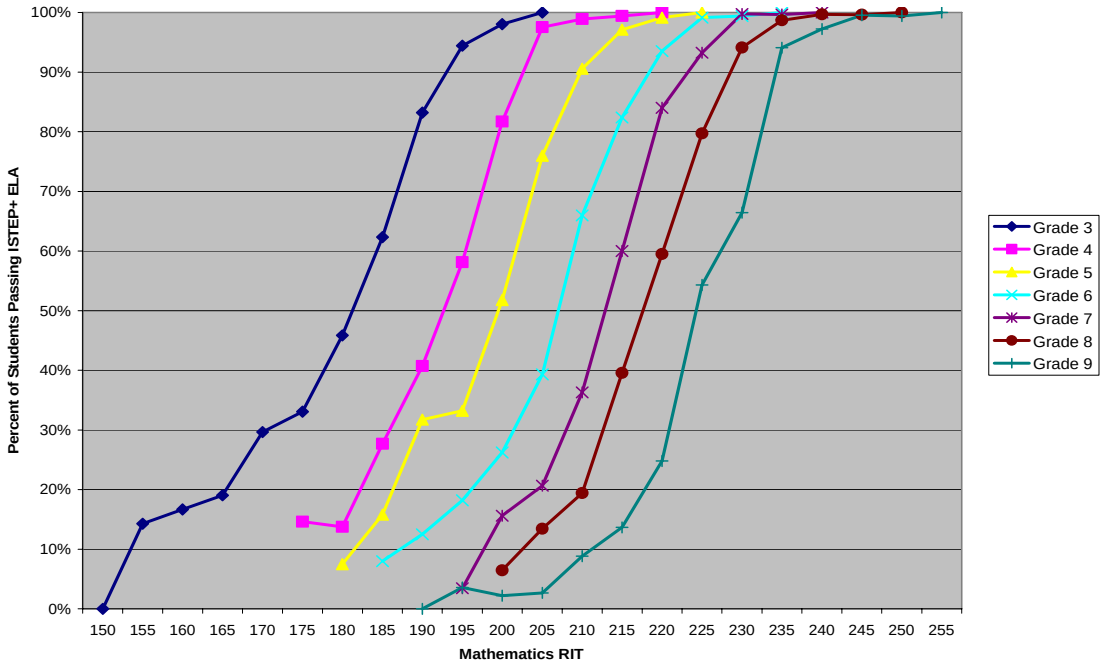


Figure 8 – Percent of Students Passing ISTEP+ Matheamtics by PRIOR SPRING Mathematics RIT Performance Range



Comparing changes in the estimated ISTEP+ standards relative to the prior alignment study

Scale Stability

It is impossible to accurately measure improvement without maintaining a stable scale. Imagine that Sid is working on his golf game and that he uses the 250 yard marker at the local driving range to estimate his driving distance. He notes that 40% of his drives roll beyond this marker in his first practice session. Sid does some weight work and takes a couple of lessons from his pro and returns to the driving range two weeks later. Now 60% of Sid's drives roll beyond the 250 marker. He naturally assumes that practice has led to improvement.

Suppose, however, that the range manager had moved the tee boxes forward ten yards so that golfers would hit off of fresh grass. If that happened, then we don't know whether Sid's improvement was a product of his hard work or a product of a change in the scale. In other words, the 250 yard marker represents a different distance today than it did two weeks ago.

Similarly, it's impossible to measure improvement on an academic test without maintaining a very stable scale. Even small changes in a test's difficulty relative to its predecessors can have a noticeable effect on proficiency rates that is independent of instruction. If a test is slightly easier than its predecessor's for example, proficiency rates may improve (just like Sid's driving distance seemed to improve) without an actual improvement in learning having occurred.

Table 18 compares the cut scores found for the current study with those generated by our most recent prior study (2003). In general the estimates for the 2005 ISTEP+ were about 1 to 3 RIT points lower than estimates from the 2003 study. While these differences are quite small, they have a substantive effect on the number of students we would project to achieve proficiency on the 2005 version of the ISTEP.

Let's use 8th grade as an example. The reading RIT score that aligns with the minimum score for a score of ***pass*** on the ISTEP+ is 3 points lower in the 2005 study than it was when we studied the ISTEP in 2003. If we used the 2005 norming group as our sample, we would have estimated that 69% of that group would have achieved a score of ***pass*** or better on the 2005 ISTEP while only 61% of that population would have achieved a score of ***pass*** or better on the 2003 version. Put another way, the change in estimated cut scores means that 8% more 8th graders might pass the ISTEP English/Language Arts test without achieving any improvement in their RIT score.

In 8th grade mathematics the 2005 study estimates a cut score that is 2 points lower than the 2003 study. Applied to the 2005 norming group, we'd estimate that about 4% more students would be likely to pass the 2005 version of the ISTEP than the 2003 version, without any improvement in RIT score.

Table 18 – Estimated RIT cut scores for the Proficient level of performance on the ISTEP+ 2001-2005*

	Reading		Language Usage		Mathematics	
	2003	2005	2003	2005	2003	2005
Grade 3	182 (27)	179 (22)	187(31)	184 (24)	189 (38)	186 (29)
Grade 4		191 (25)		195 (27)		196 (27)
Grade 5		197 (23)		200 (23)		203 (26)
Grade 6	205 (29)	204 (27)	208 (32)	208 (32)	212 (34)	209 (27)
Grade 7		208 (27)		212 (33)		215 (28)
Grade 8	216 (39)	213 (31)	217 (40)	215 (34)	224 (36)	222 (32)
Grade 9		216 (32)		218 (37)		227 (35)

*NWEA percentile score (based on 2005 norms study) is in parentheses

The Indiana Department of Education and their test vendors employ best practices relative to scale design and maintenance in order to maintain a stable scale for the ISTEP. We have no way of knowing first hand what might cause an across the board decline in estimated cut scores relative to the NWEA assessment. We do know that there are many possible factors that could contribute to this change, and many of them are independent of scaling methodologies. Nevertheless, districts should make note of this change when monitoring their longitudinal ISTEP results vis-à-vis their results from NWEA assessments. Ideally an improving school system would want to see simultaneous improvement in their state and NWEA results. In this case, some school systems may see greater improvement in their state results than in NWEA assessments because of factors independent of improvement in instruction that related to the design, scaling or delivery of the state assessment.

Calibration

It is also desirable for proficiency cut scores to calibrate across grades. By this we mean that the proficiency standard for performance in one subject and great, say grade 3 mathematics for example, should be no easier or more difficult than the standard at other grades.

Our 2003 study found that the ISTEP+ English/Language Arts test did not calibrate very well relative to NWEA reading scores. In this case the third grade estimated standard, which was estimated at the 22nd percentile on the 2005 NWEA norms, was considerably easier than the 8th grade standard, which was estimated at the 39th percentile. The results of the current study (see Figure 5) show that the test now seems to calibrate somewhat more closely across grades, although the 8th grade test remains somewhat more challenging, relative to NWEA scores, than the 3rd grade version. The mathematics tests calibrate relatively closely from grade 3 through grade 6, but become more difficult in grades 7 and 8.

Table 19 shows estimated reading and mathematics cut scores for each grade that would be calibrated with the ISTEP+ 9th grade standard in each grade. These cut scores represent a consistent level of difficulty across grades that can be used to gauge whether students who are projected to pass their current grade's ISTEP+ are, in fact, performing at a level that would project them to pass the ISTEP+ in future grades.

Figure 5 – NWEA percentile score required to achieve proficient performance on ISTEP+ 2003 - 2004

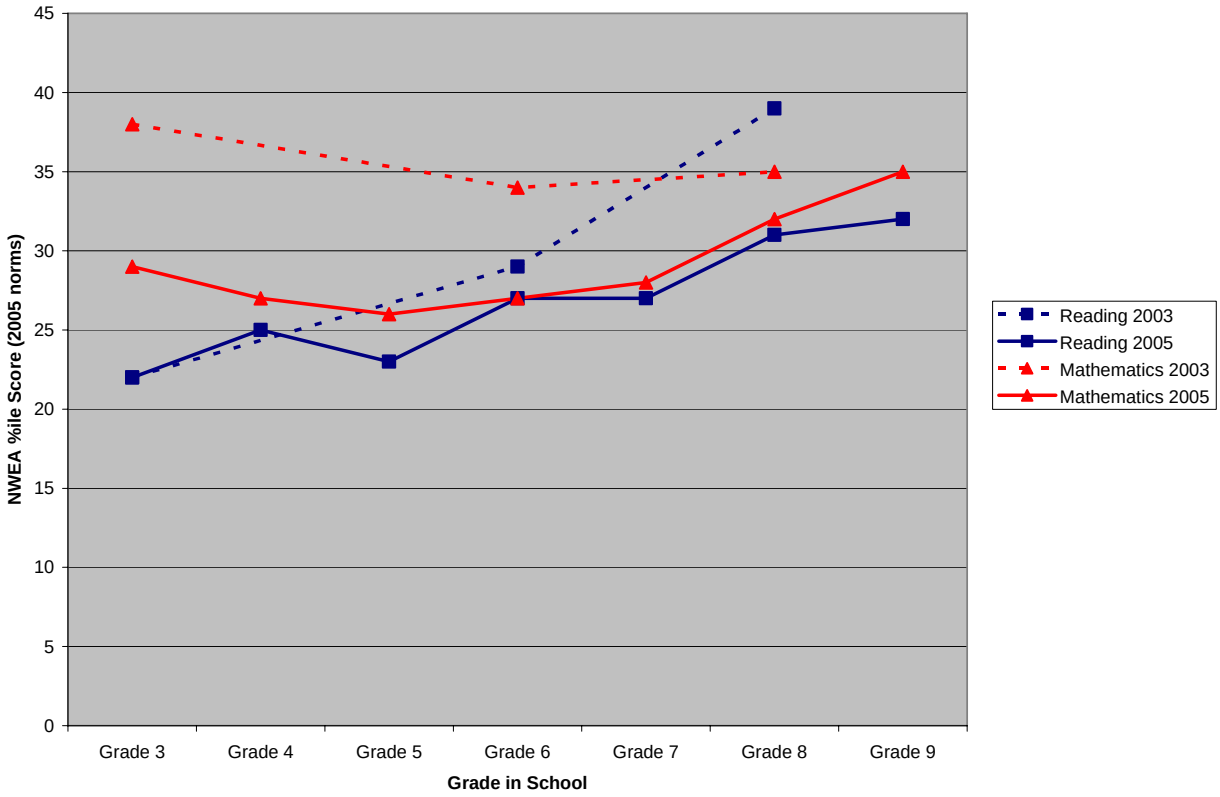


Table 19 – RIT score that calibrates (based on percentile) to proficient performance for grade 9 in reading and mathematics

	Reading		Mathematics	
	Estimated cut score for this grade	Cut score calibrated to the grade 8 standard (32th %ile)	Estimated cut score for this grade	Cut score calibrated to the grade 8 standard (35th %ile)
Grade 3	179 (22)	185	186 (29)	188
Grade 4	191 (25)	194	196 (27)	199
Grade 5	197 (23)	201	203 (26)	207
Grade 6	204 (27)	206	209 (27)	213
Grade 7	208 (27)	210	215 (28)	218
Grade 8	213 (31)	214	222 (32)	224
Grade 9	216 (32)	216	227 (35)	227

Comparing the ISTEP+ standards relative to those in place in other states

Northwest Evaluation Association tests have been aligned with the cut scores state assessments in 17 states. To get an estimate of the difficulty of the ISTEP+ in relation to other state tests, we evaluated the

standard defined as the NCLB passing score and compared it to the cut score representing the same standard in these other states. You can view the results of this analysis at the following web location: www.nwea.org/research/state.

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

Summary and Conclusions

This study investigated the relationship between the scales used for the ISTEP+ assessments and the RIT scales used to report performance on Northwest Evaluation Association tests. The study estimated the changes in reading and mathematics RIT score equivalents for the ISTEP+ performance levels in those subjects. Test records for more than 20,000 students were included in this study.

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

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

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

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

References

Cronin J, (2003). Aligning the NWEA RIT Scale with *the Indiana Statewide Testing for Educational Progress Plus(ISTEP+).* Lake Oswego, OR. Northwest Evaluation Association.

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