

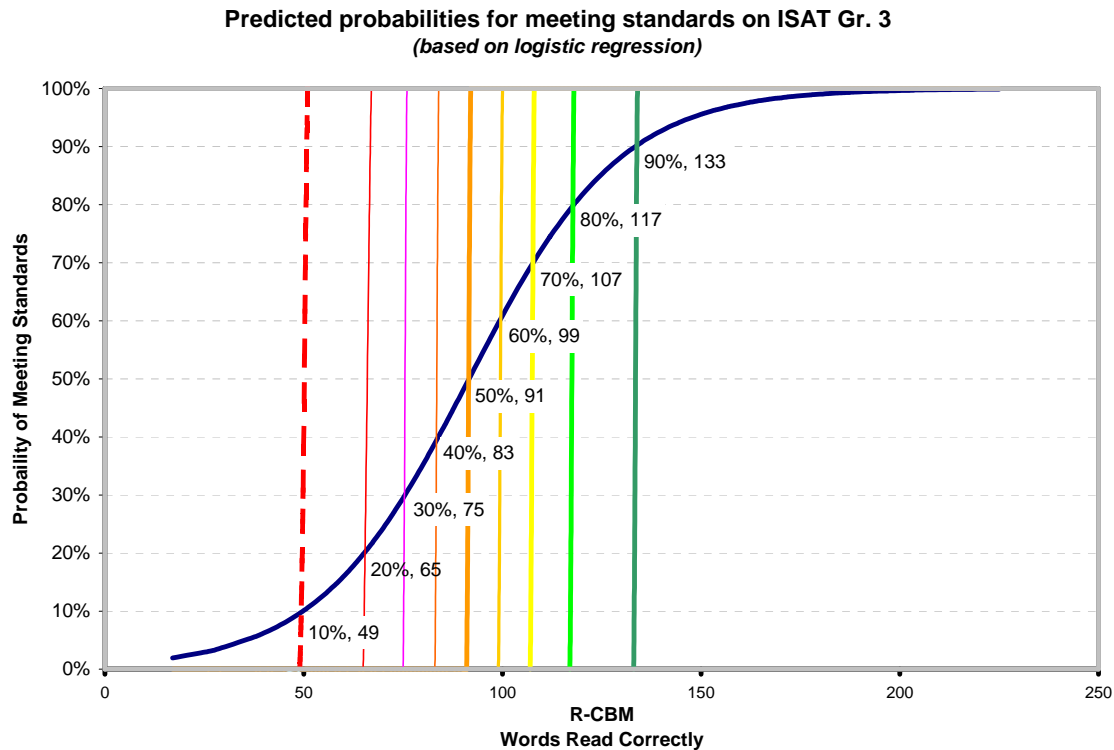
Running Head: The viability of Curriculum-Based Measurement as a predictor of high stakes tests in Illinois

Increasing demands associated with accountability for learning have led many schools and districts to move away from commercially published and scored standardized norm – reference tests to standardized measures of general outcomes that can be administered and scored in the confines of a teacher’s classroom. Although many practitioners of the traditional assessment ilk, have begun to struggle with the popularity of these quick, efficient measures of general outcomes, teachers and administrators have begun to see the value of these measures for predicting outcomes on High-stakes tests as well as monitoring the effectiveness of interventions. In many districts a Kuhnian battle of sorts (1962, p. 84) has been the result of this paradigmatic strife. This document is the result of many questions that have been asked about the technical adequacy of R-CBM and MAZE as predictors of High-Stakes State Tests in Illinois.

In the 2003-04 and 2004-05, school years, all students in a medium sized suburban school district were assessed three times per year by teachers with R-CBM and MAZE, both acquired from Aimsweb. For each grade 3, 5, and 8, students were also tested using the Illinois Standards Achievement Test (ISAT) in Reading and in grades 4 and 7 in Science. Predictive Validity coefficients (correlations between the previous year Science scores and subsequent year’s Reading scores) were moderate to strong (i.e., > .70) indicating that ISAT science was assessing in large part Reading competency. Taken together, data from between 440 and 710 students at each grade level were used. In accordance with NCLB (2004) scores from ISAT were dichotomized into Meeting and Not Meeting standards.

Using the dichotomies set up by the state two analyses were run to maximize predictive utility of cut scores. Logistic regression (LR) was used because using the default settings (i.e., classification cut off = .5), cut scores maximize specificity, and positive predictive values. As a non-parametric analysis, LR is robust to many of the violations of parametric analyses, and the sigmoidal curve lends itself to interpretation of a single predictor score (i.e., R-CBM or MAZE) as a probability of meeting standards on the criterion variable. See Figure 1, below.

## Running Head: The viability of Curriculum-Based Measurement as a predictor of high stakes tests in Illinois



Note. This chart shows the relative probabilities of Meeting Standards on Grade 3 ISAT for different obtained scores on a Spring R-CBM task. For example, a child who reads 91 WRC would have a predicted probability of meeting standards of about 50% while a child who reads at a rate of approximately 117 WRC has a predicted probability for meeting standards of 80%.

Although, somewhat more restrictive than Logistic Regression, the second analysis that was run was linear discriminant function analysis (LDFA). Again, using dichotomies based on results of ISAT testing as the dependent variable and a single continuous variable (i.e., R-CBM or MAZE) as the independent variable, using default settings cut scores maximize sensitivity and negative predictive power.

Using the combined rounded, results from LR and LDFA, cut scores that maximize specificity and positive predictive values are designated; i.e., Below Basic, as well as the designation of cut scores that maximize sensitivity and negative predictive values i.e., Proficient. The resulting figures can then be used for criterion referenced standards.

In general, this procedure approximates a rule of 30%, 60% and 90% probability for meeting standards for Below Basic, Indeterminate Risk, and Low Risk groups, respectively. The figures presented in Table 1 represent the cut scores, the number of student scores used in the calculation, the linear correlation (predictive, criterion-related validity), the percent of students in the sample who met standards and the percentage of students in each performance category who actually met standards.

Table 2. Cut scores for R-CBM and associated probabilities for meeting standards.

		Cut Score		N	Predictive coefficient	Percent of Group Met Standards	Probability of Meeting Standards		
		Below Basic	Proficient				Below Basic	Questionable Status	Proficient
Fall	Grade 3	50	55	462	.66	77.5	27.0	68.2	90.9%
Winter	Grade 3	80	90	470	.70	77.7	30.5	74.2	91.0%
Spring	Grade 3	95	115	451	.68	76.9	31.5	75.7	91.3%
Fall	Grade 5	95	130	527	.62	71.5	28.0	62.6	91.1%
Winter	Grade 5	105	145	544	.60	71.3	30.4	58.6	89.3%
Spring	Grade 5	130	160	532	.61	71.2	33.0	63.0	89.1%
Fall	Grade 8	120	155	702	.62	76.8	33.7	66.0	91.8%
Winter	Grade 8	125	165	695	.58	77.4	31.0	67.0	90.8%
Spring	Grade 8	138	175	708	.62	77.3	31.4	63.8	91.5%

Note. Below Basic includes scores at or below listed score; Questionable Status includes scores from Below Basic +1 to Proficient -1; Proficient includes scores at or above the listed scores

Table 3. Cut scores for Maze and associated probabilities for meeting standards.

		Cut Score		N	Predictive coefficient	Percent of Group Met Standards	Probability of Meeting Standards		
		Below Basic	Proficient				Below Basic	Questionable Status	Proficient
Fall	Grade 3	5	10	446	.57	78.2	30	69	91
Winter	Grade 3	10	17	441	.65	77.2	27	74	93
Spring	Grade 3	15	22	447	.63	76.6	28	71	93
Fall	Grade 5	11	17	527	.55	71.5	36	63	89
Winter	Grade 5	16	23	521	.58	71.4	26	65	88
Spring	Grade 5	21	28	533	.61	71.1	32	58	88
Fall	Grade 8	15	20	711	.52	77.2	47	71	93
Winter	Grade 8	18	25	698	.63	76.7	39	74	92
Spring	Grade 8	22	28	708	.68	76.0	44	69	93

Note. Below Basic includes scores at or below listed score; Questionable Status includes scores from Below Basic +1 to Proficient -1; Proficient includes scores at or above the listed scores

Running Head: The viability of Curriculum-Based Measurement as a predictor of high stakes tests in Illinois

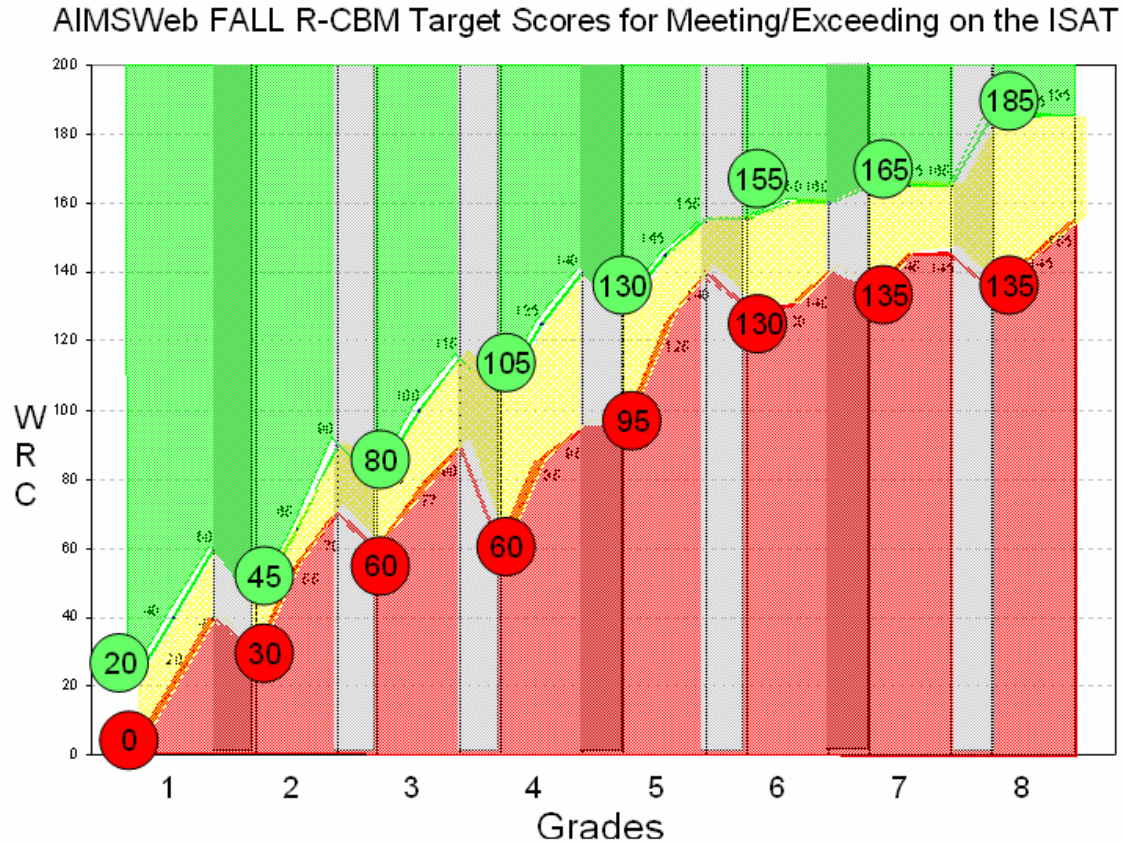
In Table 4 are displayed the criterion referenced cut-scores generated and validated for the Illinois Standards Achievement Test.

Grade	Measure	Fall		Winter		Spring	
		Below Basic	Proficient	Below Basic	Proficient	Below Basic	Proficient
1	R-CBM	0	20	20	40	40	60
	MAZE	x	x	2	6	5	10
2	R-CBM	30	45	55	65	70	90
	MAZE	2	4	5	10	8	15
3	R-CBM	60	80	80	100	90	115
	MAZE	5	10	10	17	15	22
4	R-CBM	60	105	85	125	95	140
	MAZE	7	11	13	18	15	20
5	R-CBM	95	130	125	145	140	155
	MAZE	11	16	16	22	21	28
6	R-CBM	130	155	130	160	130	160
	MAZE	12	17	18	24	23	30
7	R-CBM	135	165	145	165	145	165
	MAZE	14	19	17	23	23	30
8	R-CBM	135	185	145	185	155	185
	MAZE	15	20	18	25	22	30

Note. A backward design was used to predict cut scores for grades 6, 2 and 1. Specifically, scores for grade 2 are based on predicting scores indicating proficiency on R-CBM (or MAZE) in the fall of grade 3.

These data can be depicted visually as well to illustrate the expected growth in R-CBM based on meeting standards. The visual depiction highlights the region of questionability that falls between the prediction of not meeting standards, and the prediction of likely to meet standards.

Running Head: The viability of Curriculum-Based Measurement as a predictor of high stakes tests in Illinois



Taken together, the data appear to indicate that approximately 20 to 30% of students identified as Below Basic, are in fact False Positive predictions (i.e., these students meet standards); approximately 60% of students in the questionable region meet standards; and approximately 85 to 90% of students in the proficient region meet standards (i.e., 10 to 15% false negatives). *Note. These outcomes have been confirmed with samples of students from additional districts in Illinois, and for third grade, using DIBELS Oral Reading Fluency, have been confirmed for 2005-06 ISAT and the Illinois Measure of Annual Progress in English (IMAGE).*

In general, these analyses show that R – CBM does a great job of predicting outcomes for students, but certainly a one minute sample of behavior (or even a three minute sample) is not a perfect indicator of literacy or even of outcomes on high stakes tests. To maximize accuracy, we have to admit that to predict status for some children we require more than a one minute, or the median of 3 minutes of a sample of behavior.