Technical Report # 1806

Supplementary Report on easyCBM PRF Measures: A Follow-Up to Previous Technical Reports

Julie Alonzo

Daniel Anderson

University of Oregon

Published by

Behavioral Research and Teaching
University of Oregon • 175 Education
5262 University of Oregon • Eugene, OR 97403-5262
Phone: 541-346-3535 • Fax: 541-346-5689
http://brt.uoregon.edu

Abstract

In response to a request for additional analyses, in particular reporting confidence intervals around the results, we re-analyzed the data from prior studies. This supplementary report presents the results of the additional analyses addressing classification accuracy, reliability, and criterion-related validity evidence. For ease of reference, we organize this technical report into sections based on the type of evidence being presented.

Supplementary Report on easyCBM PRF Measures:

A Follow-Up to Previous Technical Reports

This technical report is an addendum to previous technical reports. In response to a request for additional analyses, in particular reporting confidence intervals around the results, we re-analyzed the data from prior studies. This supplementary report presents the results of the additional analyses addressing classification accuracy, reliability, and criterion-related validity evidence. For ease of reference, we organize this technical report into sections based on the type of evidence being presented.

Classification Accuracy Methods

We used the Smarter Balanced English Language Arts Assessment as our criterion measure. This measure is completely independent from the screening measure. SBAS is a large-scale assessment in wide use across the United States as a state accountability measure. We used R statistical package to perform the classification analyses. The cut point of the score associated with the 40th percentile from the easyCBM National Norms was selected, as prior studies and wide-spread district policy suggests this is an appropriate cut-point for identifying students with intensive need. Although the 40th percentile might, initially, seem too high a cut-point for intensive need, the higher expectations for student performance aligns with the higher expectations for which schools are being held accountable in the past five years. (Prior to SBAS and the CCSS adoption, performance expectations in the states from which this sample was drawn were substantially lower – the 20th percentile was previously used for identifying students with intensive need. Expectations have increased, however, and thus our cut-point also had to raise.

Students who scored below the cut-point 40th percentile were assigned a variety of interventions, depending on specific pattern of need (performance on other parts of the literacy benchmark assessment such as vocabulary and reading comprehension, success of prior years' interventions, whether they also had identified mathematics needs) and resources available at the schools. Interventions ranged from one-on-one daily instruction on phonics to small group (2-6 students) twice-weekly supplemental fluency instruction, to after-school mentoring with a focus on oral reading fluency. A number of students concurrently received several of these interventions (typically only those students whose mathematics performance did not indicate a need for mathematics intervention as well because those students who also needed mathematics intervention simply did not have sufficient time in the school day to receive all the instructional interventions they needed). Interventions were delivered by a variety of personnel (depending on school/district resources): Special Education teachers, general education teachers during their "intervention block", instructional assistants, and student mentors (some adult, some older children). Sample demographics are reported in Table 1.

Table 1
Sample Demographics, Classification Accuracy Analyses

Grade	3	4	5	6	7	8
Criterion	SBAS ELA	SBAS ELA	SBAS ELA	SBAS ELA	SBAS ELA	SBAS ELA
National/Local Representation[1]	Pacific Northwest, OR and WA	Pacific Northwest, OR and WA	Pacific Northwest, OR and WA	Pacific Northwest, OR and WA	Pacific Northwest, OR and WA	Pacific Northwest, OR and WA
Date	SY2014-15	SY2014-15	SY2014-15	SY2014-15	SY2014-15	SY2014-15
Sample Size	26250	30567	30483	29800	29267	34250
Male	12667	12100	12517	12117	11817	13783
Female	11467	11800	11667	11417	11133	13317
Gender Unknown	2117	6667	6300	6267	6317	7150
Free or Reduced-price Lunch Eligible	8133	8233	7933	8300	7433	7717
White, Non-Hispanic	5617	4883	5617	4567	5283	7283
Other	20633	25683	24867	25233	23983	26967
Disability Classification	2683	2767	2550	2567	2283	2750
Language Proficiency Status (ELL)	2700	2467	2267	1783	1900	1667

Classification Accuracy Results

Results of our classification accuracy analyses are presented for fall (Table 2), Winter (Table 3), and Spring (Table 4).

Table 2
Classification Accuracy: Fall easyCBM PRF Predicting SBAS ELA Performance

Grade	3rd	4th	5th	6th	7th	8th
Criterion	SBAS English Language Arts	SBAS English Language Arts	SBAS English Language Arts	SBAS English Language Arts	SBAS English Language Arts	SBAS English Language Arts
Cut points	40th percentile	40th percentile	40th percentile	40th percentile	40th percentile	40th percentile
False Positive Rate	0.21	0.24	0.11	0.20	0.23	0.29
False Negative Rate	0.32	0.31	0.36	0.36	0.30	0.30
Sensitivity	0.66	0.67	0.43	0.60	0.62	0.55
Specificity	0.80	0.78	0.95	0.83	0.83	0.82
Positive Predictive Power	0.79	0.76	0.89	0.80	0.77	0.71
Negative Predictive Power	0.68	0.69	0.64	0.64	0.70	0.70
Overall Classification Rate	0.73	0.73	0.70	0.70	0.73	0.70
Area Under the Curve (AUC)	0.82	0.82	0.83	0.79	0.79	0.76
AUC Estimate's 95% Confidence Interval: Lower Bound	0.79	0.79	0.81	0.77	0.77	0.74
AUC Estimate's 95% Confidence Interval: Upper Bound	0.84	0.84	0.85	0.82	0.82	0.79
Specificity Value at 90% Sensitivity	0.49	0.51	0.49	0.44	0.39	0.42
Specificity Value at 80% Sensitivity	0.65	0.69	0.70	0.61	0.62	0.59
Specificity Value at 70% Sensitivity	0.76	0.78	0.81	0.76	0.76	0.68

Table 3
Classification Accuracy: Winter easyCBM PRF Predicting SBAS ELA Performance

Grade	3rd	4th	5th	6th	7th	8th
Criterion	SBAS English Language Arts	SBAS English Language Arts	SBAS English Language Arts	SBAS English Language Arts	SBAS English Language Arts	SBAS English Language Arts
Cut points	40th percentile	40th percentile	40th percentile	40th percentile	40th percentile	40th percentile
False Positive Rate	0.17	0.21	0.17	0.16	0.21	0.23

False Negative Rate	0.35	0.32	0.28	0.37	0.32	0.33
Sensitivity	0.60	0.64	0.65	0.55	0.55	0.47
Specificity	0.86	0.82	0.87	0.88	0.87	0.89
Positive Predictive Power	0.83	0.79	0.83	0.84	0.79	0.77
Negative Predictive Power	0.65	0.68	0.72	0.63	0.68	0.67
Overall Classification Rate	0.72	0.72	0.76	0.70	0.72	0.70
Area Under the Curve (AUC)	0.82	0.81	0.84	0.81	0.80	0.78
AUC Estimate's 95% Confidence Interval: Lower Bound	0.80	0.79	0.82	0.78	0.77	0.76
AUC Estimate's 95% Confidence Interval: Upper Bound	0.84	0.83	0.86	0.83	0.82	0.80
Specificity Value at 90% Sensitivity	0.50	0.50	0.52	0.47	0.42	0.42
Specificity Value at 80% Sensitivity	0.67	0.65	0.73	0.67	0.60	0.60
Specificity Value at 70% Sensitivity	0.78	0.77	0.84	0.77	0.76	0.72

Table 4

Classification Accuracy: Spring easyCBM PRF Predicting SBAS ELA Performance

Grade	3rd	4th	5th	6th	7th	8th
Criterion	SBAS English Language Arts	SBAS English Language Arts	SBAS English Language Arts	SBAS English Language Arts	SBAS English Language Arts	SBAS English Language Arts
Cut points	40th percentile	40th percentile	40th percentile	40th percentile	40th percentile	40th percentile
False Positive Rate	0.15	0.22	0.19	0.16	0.22	0.21
False Negative Rate	0.34	0.32	0.28	0.39	0.32	0.32
Sensitivity	0.61	0.66	0.64	0.52	0.58	0.46
Specificity	0.88	0.80	0.85	0.88	0.84	0.90
Positive Predictive Power	0.85	0.78	0.81	0.84	0.78	0.79
Negative Predictive Power	0.66	0.68	0.72	0.61	0.68	0.68
Overall Classification Rate	0.73	0.73	0.75	0.69	0.71	0.71
Area Under the Curve (AUC)	0.83	0.82	0.83	0.81	0.79	0.78
AUC Estimate's 95% Confidence Interval: Lower Bound	0.81	0.79	0.81	0.79	0.77	0.76

AUC Estimate's 95% Confidence Interval: Upper Bound	0.85	0.84	0.85	0.83	0.81	0.81
Specificity Value at 90% Sensitivity	0.50	0.51	0.50	0.47	0.42	0.41
Specificity Value at 80% Sensitivity	0.67	0.67	0.69	0.64	0.62	0.61
Specificity Value at 70% Sensitivity	0.81	0.76	0.80	0.77	0.71	0.70

Reliability Methods

The PRF measures provide an efficient and easy-to-administer assessment of students' oral reading fluency. For the results to be most interpretable, however, it is important that alternate forms of the measure be of equivalent difficulty/return equivalent results in the absence of changes in students' underlying oral reading fluency proficiency. Test-retest reliability provides an estimate of the consistency of scores obtained when a single form is administered to students more than once in a short period of time (in this case, with one week in between administrations). Alternate form reliability provides an estimate of the consistency of scores were different test forms to be administered. This type of reliability gives us information about how consistent results might be if the winter measure were used in place of the fall measure. This consistency in performance across testing occasions (test-retest) or forms (alternate form) is important when evaluating the trustworthiness of screening results. The G-theory studies extend on the test-retest and alternate form reliability analyses, further examining the degree to which variation in score can be attributed to alternate forms and/or alternate testing occasions.

Sample and Setting: Reliability Analyses

Students from three public elementary schools in the Pacific Northwest participated in test-retest and alternate form reliability studies, with sample size varying by grade. In grade 1, 41

students participated. In grade 2, 48 students participated. In grade 3, 50 students participated. In grade 4, 55 students participated. In grade 5, 50 students participated. A sub-sample of 38 grade 1, 34 grade 2, 38 grade 3, 39 grade 4, and 18 grade 5 students also participated in G-theory studies. No demographic information was collected in this study (see Tables 1a and b for descriptive statistics); however, on average, the participating schools comprised of 53% male students, 2% American Indian/Alaskan, 2% Asian/Pacific Islander, less than 1% of Black, 23% Hispanic, 67% White, and 8% two or more races students. 70% of the students are eligible for Free and Reduced Lunch programs. The district consists of 6% English Language Learners and 17% of students with Individualized Education Program (IEP).

Reliability Analyses

For our generalizability theory study (G-Study) we calculated the variances associated persons and two facets: forms and occasions. We then conducted decision studies (D-Studies) to help determine the necessary conditions for reliable measurement. Data for this study were analyzed in a two-facet fully crossed design (i.e., all students in the analysis were included in both testing occasions and administered the same test forms). The test forms were often administered in a different order on the separate occasions to mitigate order effects. The forms themselves remained constant across occasions in all analyses. For each grade level, we conducted 4 different G-theory analyses for passage reading fluency (PRF) to investigate 8 different test forms. The first facet in the analysis, form, was generally counterbalanced across occasions. The second facet was occasion.

Reliability Results

Table 5
Reliability Results

Type of Reliability	Grade	n	Coefficient	95% Confidence Interval*: Lower Bound	95% Confidence Interval*: Upper Bound
Alternate Form	1	41	.97	.94	.98
Alternate Form	2	48	.93	.91	.95
Alternate Form	3	50	.95	.94	.96
Alternate Form	4	55	.95	.93	.98
Alternate Form	5	50	.95	.92	.97
Test-Retest	1	41	.96	.95	.98
Test-Retest	2	48	.95	.93	.96
Test-Retest	3	50	.90	.87	.94
Test-Retest	4	55	.95	.86	.96
Test-Retest	5	50	.91	.90	.94
G-Theory	1	38	See text,		
G-Theory	2	34	See text,		
G-Theory	3	28	See text,		
G-Theory	4	39	See text,		
G-Theory	5	18	See text,		

Discussion: Reliability

The results of the test-retest and alternate-form reliability analyses suggested acceptable form equivalence for subsequent G-Theory analyses. For the Grade 1 Passage Reading Fluency analyses, 95% of the variance was associated with the 38 persons included in the analysis, 0% was associated with forms, and 0% was associated with occasion. The relative error variance was 30.78, while the absolute variance was 45.16. The G-Coefficient was .99, while the phi coefficient was .87.

For the Grade 2 Passage Reading Fluency analyses, 90% of the variance was associated with the 34 persons included in the analysis, 0% was associated with forms, and 0% was

associated with occasion. The relative error variance was 25.54, while the absolute variance was 37.18. The G-Coefficient was .98, while the phi coefficient was .97.

For the Grade 3 Passage Reading Fluency analyses, 82% of the variance was associated with the 28 persons included in the analysis, 0% was associated with forms, and 0% was associated with occasion. The relative error variance was 70.97, while the absolute variance was 97.12. The G-Coefficient was .95, while the phi coefficient was .93.

For the Grade 4 Passage Reading Fluency analyses, 88% of the variance was associated with the 39 persons included in the analysis, 0% was associated with forms, and 0% was associated with occasion. The relative error variance was 30.00, while the absolute variance was 64.07. The G-Coefficient was .98, while the phi coefficient was .96.

For the Grade 5 Passage Reading Fluency analyses, 89% of the variance was associated with the 18 persons included in the analysis, 0% was associated with forms, and 0% was associated with occasion. The relative error variance was 38.41, while the absolute variance was 58.53. The G-Coefficient was .98, while the phi coefficient was .96.

Validity Methods

We analyzed criterion validity using data from two studies. For Study 1, we used the Smarter Balanced English Language Arts Assessment as our criterion measure. This measure is completely independent from the screening measure. SBAS is a large-scale assessment in wide use across the United States as a state accountability measure. Because it is used by so many states for their accountability measure, school districts are quite interested in the relation between SBAS and easyCBM PRF. For Study 2, we used the DIBELs ORF measure to gather construct-related validity evidence. DIBELs ORF is a well-established measure for estimating students' oral reading fluency with a long history of published validity evidence. Like SBAS, DIBELs is

external to the easyCBM system. Unlike SBAS, however, the DIBELs ORF and the easyCBM PRF are designed to measure the exact same construct: Oral Reading Fluency. Thus, higher correlations between easyCBM and DIBELs ORF than between easyCBM and SBAS ELA provide strong evidence in support of the PRF measuring the intended construct (oral reading fluency).

Setting and Sample

Study 1: Data for the study examining the relation between the easyCBM PRF and the Smarter Balanced English Language Arts assessment came from a convenience sample of students provided by two school districts in the Pacific Northwest. All students enrolled in school and present during the three-week easyCBM Benchmark Assessment windows in the fall (September 2014), winter (January 2015) and spring (May 2015) were administered the easyCBM assessments. All enrolled students were likewise administered the Smarter Balanced assessments during the testing window provided by the state in the spring of 2015. The data set provided by the districts included easyCBM CCSS Math, Passage Reading Fluency, Vocabulary, and Multiple Choice Reading Comprehension (MCRC) as well as Smarter Balanced Math and English Language Arts total scores for students enrolled in grades 3-8. District 1 provided data for Grades 3-8, while District 2 provided data for Grades 4-8. In addition, District 1 provided demographic information, while District 2 (approximately ¼ the size of the first district) did not. Demographics of the sample are provided in Table 1. Because of the missing demographics from a large proportion of the sample, the percentages for each of the demographic variables are calculated based on the students in the sample whose data included full-resolution demographic information.

Table 6

Sample Demographics

Grade	Missing Demographic Data		Female		Hispanic		SpEd		ELL	
	#	%	#	%	#	%	#	%	#	%
3	33	3	492	48	187	18	87	8	67	7
4	328	24	523	50	217	21	100	10	62	6
5	295	23	483	48	159	16	89	9	39	4
6	291	22	505	49	180	17	95	9	27	3
7	280	23	456	48	185	19	78	8	29	3
8	266	20	526	50	192	18	83	8	22	2

During data cleaning, data from students who were administered the Alternate Assessment rather than the General Education assessment were removed from the dataset prior to further analyses. In all, six students each from Grades 4, 6, and 7 and three students from Grade 5 were removed from the dataset in this step. Data from all additional students were retained.

Study 2: For the study examining the relation between the easyCBM PRF and the DIBELs ORF measures, Data came from a convenience sample of students from ten schools in an Oregon school district that uses easyCBM® reading measures as part of its Response to Intervention (RTI) model. This study was conducted in January 2013, with the initial duration of the study extended from one month to 1.5 months, due to an unexpected severe flu season, which caused a high absenteeism rate. At the beginning of the study, a total of 1017 students from grade 2 (n=240), grade 3 (n=311), grade 4 (n=247), and grade 5 (n=219) were recruited. As a result of the high absenteeism rate, the final sample consisted of 204 2nd-grade students, 288 3rd-grade students, 184 4th-grade students, and 206 5th-grade students. No demographic information

was collected in this study; however, data came from participating schools with 53% male students, 2% American Indian/Alaskan, 2% Asian/Pacific Islander, less than 1% of Black, 23% Hispanic, 67% White, and 8% two or more races students. 70% of the students are eligible for Free and Reduced Lunch programs. The district consists of 6% English Language Learners and 17% of students with Individualized Education Program (IEP).

Validity Analyses

For Study 1, we used linear regression to analyze the predictive validity of the easyCBM PRF measures to the Smarter Balanced English Language Arts assessment. For Study 2, We used bivariate correlations to analyze concurrent validity for easyCBM PRF to DIBELs ORF measures.

Table 7
Criterion-Related Validity Evidence

Type of Validity	Grade	Criterion	n	Coefficient	95% Confidence Interval*: Lower Bound	95% Confidence Interval*: Upper Bound
Predictive	3	SBAS English Language Arts	1303	0.67	0.63	0.71
Predictive	4	SBAS English Language Arts	1520	0.64	0.60	0.68
Predictive	5	SBAS English Language Arts	1539	0.68	0.64	0.71
Predictive	6	SBAS English Language Arts	1467	0.61	0.57	0.65
Predictive	7	SBAS English Language Arts	1415	0.62	0.58	0.66
Predictive	8	SBAS English Language Arts	1475	0.57	0.53	0.61
Predictive	3	SBAS English Language Arts	1280	0.67	0.63	0.71
Predictive	4	SBAS English Language Arts	1489	0.63	0.59	0.67
Predictive	5	SBAS English Language Arts	1575	0.68	0.64	0.71
Predictive	6	SBAS English Language Arts	1494	0.63	0.59	0.67

Table 7
Criterion-Related Validity Evidence

Type of Validity	Grade	Criterion	n	Coefficient	95% Confidence Interval*: Lower Bound	95% Confidence Interval*: Upper Bound
Predictive	7	SBAS English Language Arts	1463	0.63	0.59	0.67
Predictive	8	SBAS English Language Arts	1535	0.60	0.56	0.64
Concurrent	3	SBAS English Language Arts	1303	0.67	0.63	0.71
Concurrent	4	SBAS English Language Arts	1520	0.64	0.60	0.68
Concurrent	5	SBAS English Language Arts	1593	0.66	0.62	0.70
Concurrent	6	SBAS English Language Arts	1500	0.62	0.58	0.66
Concurrent	7	SBAS English Language Arts	1478	0.62	0.58	0.66
Concurrent	8	SBAS English Language Arts	1526	0.62	0.58	0.66
Concurrent	2	DIBELs ORF	229	.95	.94	.95
Concurrent	3	DIBELs ORF	290	.94	.94	.96
Concurrent	4	DIBELs ORF	236	.93	.91	.94
Concurrent	5	DIBELs ORF	208	.88	.88	.91

Validity Discussion

For Study 1, the provided data indicate a moderate positive relation between the easyCBM PRF measures and the large-scale Smarter Balanced English Language Arts assessment at all tested grades and seasons. For Study 2, the provided data indicate a very strong positive relation between the easyCBM PRF measures and the DIBELs ORF measures at all tested grades. These findings, taken in concert with one another, provide strong evidence of the easyCBM PRF measure as an appropriate assessment of students' oral reading fluency. The correlations between the easyCBM PRF measures and the DIBELs ORF measures suggest they are measuring the same construct (as intended). Because oral reading fluency has consistently been shown to predict other reading outcomes, such as direct measures of comprehension (e.g., the SBAS ELA assessment), coefficients ranging from .57 to .68 support the validity of

including the easyCBM PRF measures as part of an assessment battery for screening students at risk for not meeting end-of-year performance expectations. The PRF measures are one of three different measures that together comprise the easyCBM Benchmark Assessments in reading.