

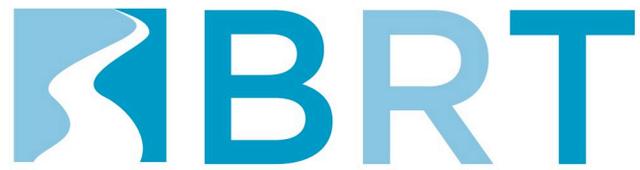
Technical Report # 1801

**Classification Accuracy of the easyCBM
Kindergarten – Grade 2 Reading Measures**

Julie Alonzo

Daniel Anderson

University of Oregon



behavioral research & teaching

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>

Copyright © 2018. Behavioral Research and Teaching. All rights reserved. This publication, or parts thereof, may not be used or reproduced in any manner without written permission.

The University of Oregon is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation. This document is available in alternative formats upon request.

Abstract

In this technical report, we present results of classification accuracy analyses to identify cut scores to optimize sensitivity and specificity for the easyCBM literacy assessments in Kindergarten through Grade 2. In conducting these analyses, we used the following approach: We kept sensitivity above 0.80 and maximized specificity from there. If specificity crossed 0.90, we kept sensitivity above 0.90 and maximized specificity from there. If there were multiple cut scores with the same specificity maximum while keeping sensitivity above 0.80 or 0.90, respectively, we selected the cut score that maximized sensitivity. For the Kindergarten measures, we used the raw score associated with the 50th percentile on the Spring Letter Sounds measure as the outcome. For the Grade 1 measures, we used the raw score associated with the 50th percentile on the Spring Word Reading Fluency measure as the outcome. For the Grade 2 measures, we used the raw score associated with the 50th percentile on the Spring Passage Reading Fluency measure as the outcome. With the exception of the Phoneme Segmenting measure when administered in the Fall of Kindergarten, all measures demonstrated strong classification accuracy.

Classification Accuracy of the easyCBM Kindergarten – Grade 2 Reading Measures

The easyCBM assessment system (Alonzo, Tindal, Ulmer, & Glasgow, 2006) has been widely adopted for use in school districts operating under Response to Intervention (RTI) or Multi-Tiered Systems of Support (MTSS) approaches to meeting student learning needs. The system includes both universal screening and progress monitoring assessments for identifying Kindergarten through Grade 8 students at risk in literacy and mathematics. In this technical report, we focus on the English language literacy universal screening measures available for use with students in Kindergarten through Grade 2. Findings from the study reported here are also applicable to the progress monitoring measures available at those grades, as both the universal screening (benchmark) and progress monitoring measures were developed concurrently, using an identical process, and designed to be of equivalent difficulty. In other words, each of the progress monitoring and universal screening measures of the same measure type (e.g., Word Reading Fluency, Letter Names) should be viewed as equivalent alternate forms.

The easyCBM English-language literacy measures we analyzed in this study include Letter Names, Phoneme Segmenting, Letter Sounds, Word Reading and Passage Reading Fluency. A brief description of each measure is provided below.

The *Letter Names Fluency* measure (LN) assesses alphabetic principles and rapid naming fluency (Alonzo & Tindal, 2007a). Students are presented with a sheet of paper on which letters in both their capital and lower-case forms are printed in a table. Students have 60 seconds to name as many letters as they can, reading across the paper from left to right, then down to the next row. Teachers follow along on their own test protocol (either a touch-screen tablet for instant scoring or a paper copy of the assessor version of the test), marking as errors any letters skipped or named incorrectly. If a student pauses more than three seconds on a letter, the

assessor supplies the letter and marks it as incorrect. The final score is reported as correct letters named per minute. The Letter Names Fluency measure is included as one of the Kindergarten Benchmark assessments in the fall only.

The *Phoneme Segmenting Fluency* measure (Seg) assesses phonemic awareness and is administered entirely orally (Alonzo & Tindal, 2007a). The teacher reads from a list of words and asks the student to segment each word into its constituent phonemes. The measure is administered individually for 60 seconds. Teachers follow along on their own test protocol (either a touch-screen tablet for instant scoring or a paper copy of the assessor version of the test), marking as errors any phonemes skipped or segmented incorrectly. If a student pauses more than three seconds, the assessor supplies the segmented word and marks it as incorrect. The final score is reported as correct phonemes segmented per minute. The Phoneme Segmenting Fluency measure is included as one of the Benchmark assessments for the fall, winter, and spring of Kindergarten, as well as for the fall of first grade.

The *Letter Sounds Fluency* measure (LS) assesses phonics (Alonzo & Tindal, 2007a). Students are presented with a sheet of paper on which letters in both their capital and lower-case forms are printed in a table. They are prompted to produce the sound the letter makes, reading across the paper from left to right, then down to the next row. Some common digraphs (sh, ph, th) are also included on this measure. Students are given 60 seconds to complete this measure. Teachers follow along on their own test protocol (either a touch-screen tablet for instant scoring or a paper copy of the assessor version of the test), marking as errors any letters skipped or sounded incorrectly. If a student pauses more than three seconds on a letter, the assessor supplies the letter sound and marks it as incorrect. The final score is reported as correct letter sounds

produced per minute. The Letter Sounds Fluency measure is included as one of the Benchmark assessments for the fall, winter, and spring of both Kindergarten and first grade.

The *Word Reading Fluency* measure (WRF) assesses oral reading fluency and rapid naming fluency (Alonzo & Tindal, 2007b). Students are shown a piece of paper with a variety of decodable and sight-words arranged in a table. They are instructed to read the words aloud, moving left to right and then down the rows. This test is administered individually for 60 seconds. Teachers follow along on their own test protocol (either a touch-screen tablet for instant scoring or a paper copy of the assessor version of the test), marking as errors any words skipped or read incorrectly. If a student pauses more than three seconds on a word, the assessor supplies the word and marks it as incorrect. The final score is reported as words read correct per minute. The Word Reading Fluency measure is included as one of the Benchmark assessments for the winter and spring of Kindergarten and fall, winter, and spring of Grade 1.

The *Passage Reading Fluency* measure (PRF) also assesses oral reading fluency (Alonzo & Tindal, 2007b). Students are given 60 seconds to read aloud a short (approximately 250 word) narrative passage presented to them on a single side of a sheet of paper. Teachers follow along on their own test protocol (either a touch-screen tablet for instant scoring or a paper copy of the assessor version of the test), marking as errors any words skipped or read incorrectly. If a student pauses more than three seconds on a word, the assessor supplies the word and marks it as incorrect. The passages used are written to be at middle of the year reading level for each grade. The Passage Reading Fluency measure is included as one of the Benchmark assessments for the winter and spring of Grade 1 and fall, winter, and spring of Grades 2-8. Because classification accuracy of the PRF measures in the upper elementary and middle school grades have been reported elsewhere, we constrain the current study to Grades 1 and 2. The *Passage Reading*

Fluency measure (PRF) also assesses oral reading fluency (Alonzo & Tindal, 2007b). Students are given 60 seconds to read aloud a short (approximately 250 word) narrative passage presented to them on a single side of a sheet of paper. Teachers follow along on their own test protocol (either a touch-screen tablet for instant scoring or a paper copy of the assessor version of the test), marking as errors any words skipped or read incorrectly. If a student pauses more than three seconds on a word, the assessor supplies the word and marks it as incorrect. The passages used are written to be at middle of the year reading level for each grade.

The Multiple Choice Reading Comprehension (MCRC) measure is included as one of the Benchmark assessments for the fall, winter, and spring of Grade 2-8 (Alonzo, Liu, & Tindal, 2008). The second grade MCRC measure includes an original work of fictional narrative, approximately 900 words long, followed by 12 selected response questions. Seven of the questions target literal comprehension while the remaining five target inferential comprehension. Each question includes three possible answer choices: a correct answer, a near-distractor, and a far-distractor. The MCRC measures are designed to be group-administered on computers. Scores are reported as total items correct. Questions for which no answer is selected are counted as incorrect. Because classification accuracy of the MCRC measures in the upper elementary and middle school grades have been reported elsewhere, we constrain the current study to Grade 2.

The Vocabulary (VOC) measure is included as one of the Benchmark assessments for the fall, winter, and spring of Grade 2-8 (Alonzo, Anderson, Park, & Tindal, 2012). The VOC measure presents students with vocabulary words presented in context (typically embedded within a sentence) and asks students to select the best answer from three possible selected response options. Each question includes three possible answer choices: a correct answer, a near-distractor, and a far-distractor. The VOC measures are designed to be group-administered on

computers. Scores are reported as total items correct. Questions for which no answer is selected are counted as incorrect. Because classification accuracy of the VOC measures in the upper elementary and middle school grades have been reported elsewhere, we constrain the current study to Grade 2.

Methods

In this section we provide a description of the sample and analytic method used in this study.

Sample and Data Collection

Data for this study were provided by a small school district in the Pacific Northwest. The data set included all students enrolled in the district from the fall of 2016 through the spring of 2017 who were present when the fall, winter, and spring benchmark assessments were administered. In all, data were included from 288 kindergarten students, 315 first-grade students, and 326 second-grade students. Sample demographics are included in Table 1.

Table 1
Demographics of the Sample

Demographic Variables	Kindergarten	Grade 1	Grade 2
Female	47%	50%	51%
Receiving Special Education Services	19%	17%	14%
English Learners	32%	30%	31%
Hispanic	37%	40%	36%
White	84%	83%	71%
Two or More Races	7%	6%	15%
American Indian	8%	10%	13%
Native Hawaiian or other Pacific Islander	< 1%	< 1%	< 1%
Black or African American	< 1%	< 1%	< 1%
Asian	< 1%	< 1%	< 1%

Assessments were administered by school district personnel (either the students' assigned teachers or instructional assistants assigned to work with them in their classrooms), using their district protocol for assessing students in grades K-2. Prior to administering the assessments, personnel were provided training on standardized test administration and scoring via the online training resources offered on the easyCBM system. Per District policy, all staff had to reach proficiency on the easyCBM training site on administering and scoring the specific measures for which they would be responsible prior to assessing students. The district provided the research team with the data set at the end of the school year to enable analysis of student performance in the district.

The total number of students with scores varied by measure and benchmark season, with more students having assessment data in the dataset as the year progressed. In kindergarten, students were administered the Phoneme Segmenting (SEG), Letter Names (LN), and Letter Sounds (LS) fluency measures in the fall, and they were administered the Seg, LS, and Word Reading Fluency (WRF) in the winter and spring. Descriptive statistics for the kindergarten sample are reported in Table 2.

Table 2
Descriptive Statistics: Kindergarten

	Fall			Winter			Spring		
	SEG	LN	LS	SEG	LS	WRF	SEG	LS	WRF
n	264	244	262	254	266	264	263	265	265
M	6.41	12.55	5.36	32.70	23.71	7.83	50.18	38.77	15.18
SD	11.66	13.21	9.20	18.07	14.46	10.92	15.48	16.60	13.66

The sample's mean performance roughly matches the 50th percentile of the national norms on all measures and seasons with the exception of the fall LN measure, on which the mean of the sample matched the 31st percentile of the national norms.

In first grade, students were administered the Seg, LS, and WRF measures in the fall, and they were administered the LS, WRF, and PRF in the winter and spring. Descriptive statistics for the first-grade sample are reported in Table 3. Again, the first-grade sample's performance closely matched the national norms.

Table 3
Descriptive Statistics: Grade 1

	Fall			Winter			Spring		
	Seg	LS	WRF	LS	WRF	PRF	LS	WRF	PRF
n	278	278	278	293	293	300	303	303	304
M	36.63	30.22	18.48	42.23	27.63	36.36	50.40	44.00	57.53
SD	14.46	13.01	19.81	18.24	23.75	36.77	18.30	27.37	44.70

In second grade, students were administered the Multiple Choice Reading Comprehension (MCRC), Vocabulary (VOC) and PRF measures in the fall, winter, and spring. Descriptive statistics for the second-grade sample are reported in Table 4.

Although second-grade sample students' performance closely matched the national norms on MCRC and VOC across all three seasons, PRF performance was closer to the 43rd percentile in the fall and dropped to the 34th percentile in the winter, and the 37th percentile in the spring.

Table 4
Descriptive Statistics: Grade 2

	Fall			Winter			Spring		
	MCRC	Voc	PRF	MCRC	Voc	PRF	MCRC	Voc	PRF
n	302	302	302	307	309	308	310	311	310
M	6.34	8.18	55.76	7.80	9.31	70.65	8.13	9.45	87.32
SD	2.70	3.56	33.84	2.93	3.33	38.71	3.32	3.16	44.33

Data Analysis

We analyzed classification accuracy by conducting a Sensitivity/Specificity ROC analysis. This approach yields an Area Under the Curve (AUC) statistic, which provides an overall indication of the diagnostic accuracy of the Receiver Operating Characteristic (ROC) curve that generalizes the set of potential combinations of sensitivity and specificity for predictors. In conducting these analyses, we used the following approach: We kept sensitivity above 0.80 and maximized specificity from there. If specificity crossed 0.90, we kept sensitivity above 0.90 and maximized specificity from there. If there were multiple cut scores with the same specificity maximum while keeping sensitivity above 0.80 or 0.90, respectively, we selected the cut score that maximized sensitivity (Silberglitt & Hintze, 2005).

For the Kindergarten measures, we used the raw score associated with the 50th percentile on the Spring Letter Sounds measure (32 letter sounds correct per minute) as the outcome. For the Grade 1 measures, we used the raw score associated with the 50th percentile on the Spring Word Reading Fluency measure (49 words correct per minute) as the outcome. For the Grade 2 measures, we used the raw score associated with the 50th percentile on the Spring Passage Reading Fluency measure (101 words correct per minute) as the outcome.

Results

Results of the ROC analysis are presented in Table 5.

Table 5
AUC, Threshold, Sensitivity, and Specificity Statistics, by Grade and Measure

Measure	Grade	Season	AUC	Threshold	Sensitivity	Specificity
LN	K	fall	0.76	6.50	0.82	0.66
LS	K	fall	0.68	1.50	0.80	0.49
SEG	K	fall	0.37	---	1.00	0.00
LS	K	win	0.87	18.50	0.82	0.81
SEG	K	win	0.75	40.50	0.81	0.45
SEG	K	spr	0.85	49.50	0.80	0.77
SEG	1	fall	0.73	46.50	0.81	0.36
LS	1	fall	0.84	33.50	0.81	0.75
WRF	1	fall	0.95	11.50	0.84	0.89
LS	1	win	0.81	45.50	0.81	0.63
WRF	1	win	0.98	22.50	0.90	0.93
MCRC	2	fall	0.76	7.50	0.83	0.57
VOC	2	fall	0.82	10.50	0.85	0.64
PRF	2	fall	0.92	54.50	0.81	0.85
MCRC	2	spr	0.81	10.50	0.84	0.61
VOC	2	win	0.82	11.50	0.82	0.59
MCRC	2	win	0.81	9.50	0.81	0.63
PRF	2	win	0.94	84.50	0.90	0.74
VOC	2	spr	0.85	11.50	0.85	0.64

Because the kindergarten spring Letter Sounds measure was used as the criterion in the ROC analysis, it is excluded from these analyses. With the exception of the Phoneme Segmenting

measure when administered in the fall of the Kindergarten year, all measures demonstrated strong classification accuracy, with sensitivity and specificity well within accepted ranges. AUC plots for the kindergarten, grade 1, and grade 2 measures are presented in Figures, 1, 2, and 3, respectively.

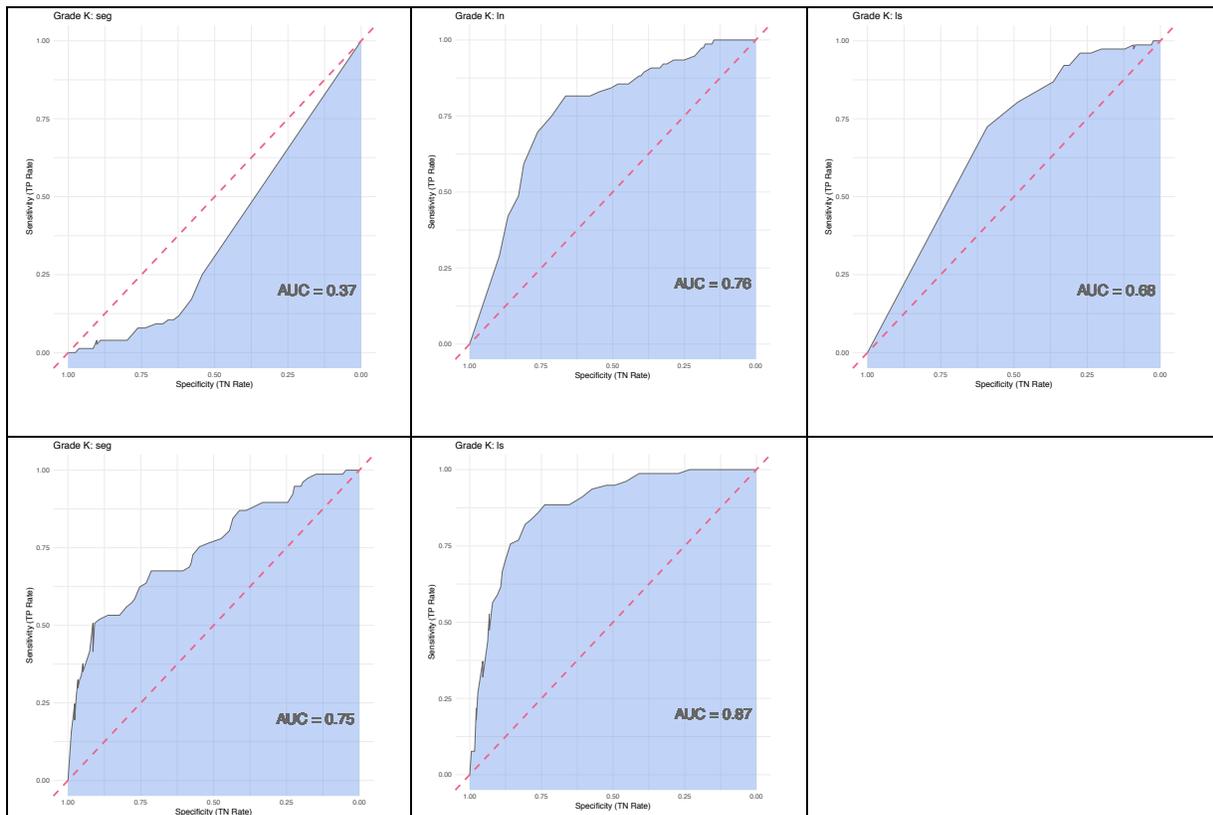


Figure 1
ROC plots, Kindergarten measures.

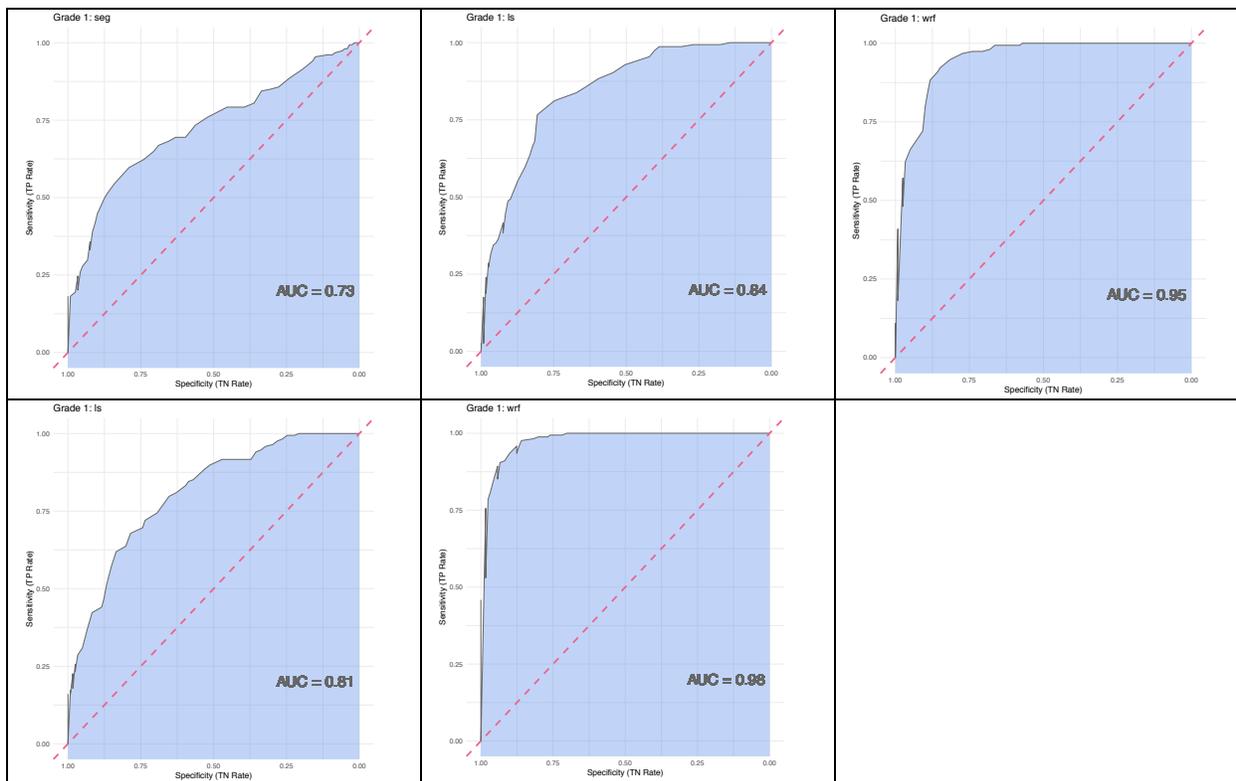


Figure 2
ROC plots, first-grade measures.

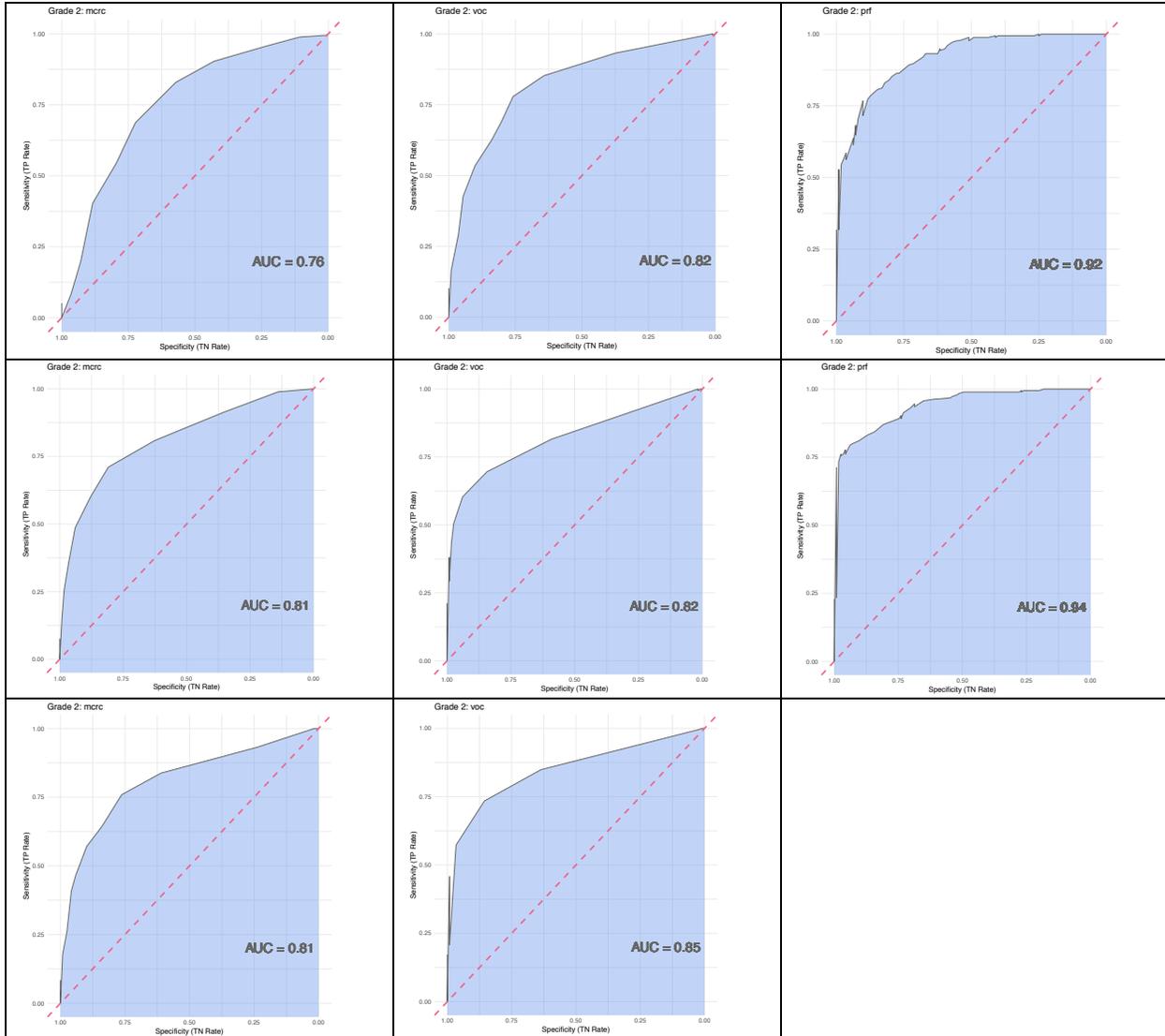


Figure 3
ROC plots, second-grade measures.

References

- Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012). *The development of CBM vocabulary measures: Grade 2* (Technical Report No. 1209). Eugene, OR: Behavioral Research and Teaching, University of Oregon.
[TechRpt 1209](#)
- Alonzo, J., Liu, K., & Tindal, G. (2008). *Examining the technical adequacy of second-grade reading comprehension measures in a progress monitoring assessment system* (Technical Report No. 0808). Eugene, OR: Behavioral Research and Teaching, University of Oregon.
[TechRpt 0808](#)
- Alonzo, J., & Tindal, G. (2007). *Examining the technical adequacy of early literacy measures in a progress monitoring assessment system: Letter names, letter sounds, and phoneme segmenting* (Technical Report No. 39). Eugene, OR: Behavioral Research and Teaching, University of Oregon. [TechRpt 39](#)
- Alonzo, J., & Tindal, G. (2007). *The development of word and passage reading fluency measures in a progress monitoring assessment system* (Technical Report No. 40). Eugene, OR: Behavioral Research and Teaching, University of Oregon. [TechRpt 40](#)
- Alonzo, J., Ulmer, K., Tindal, G., & Glasgow, A. (2006). easyCBM online assessment system. <http://easycbm.com>. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Silberglitt, B., & Hintze, J. (2005). Formative assessment using CBM-R cut scores to track progress toward success on state-mandated achievement tests: A comparison of methods. *Journal of Psychoeducational Assessment*, 23, 304-325. doi: 10.1177/073428290502300402