

**Technical Report 2603-RK8R**

**Reliability Analyses for easyCBM<sup>®</sup> in Grades K – 8: Reading**

Gerald Tindal, PhD – University of Oregon

Joseph F. T. Nese, PhD – University of Oregon

Sara McCaslin, PhD – [mccaslinwordsmithing.wordpress.com](http://mccaslinwordsmithing.wordpress.com)



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>

Note: This technical report was supported in part by Riverside Insight, the exclusive distributor of easyCBM<sup>®</sup>. The report does not reflect any endorsement by any of these organizations.

Copyright<sup>©</sup> 2026. Behavioral Research and Teaching. All rights reserved. This publication or parts thereof, may not be used or reproduced in any manner without written permission.

APA Reference: Tindal, G. and Nese, J. F. T. (2026). *Reliability Analyses for easyCBM<sup>®</sup> in Grades K-8: Reading (Technical Report 2603-RK8R)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

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.

### Table of Contents for 2603-RK8R\_ReliabilityReading

<b>Introduction and Overview<sup>1</sup></b>	<b>4-6</b>
<b>Summary of Technical Reports</b>	
<b>0906:</b> Alternate Form and Test–Retest Reliability of easyCBM <sup>®</sup> Reading Measures	7
<b>1003:</b> Technical Adequacy of the easyCBM <sup>®</sup> Primary-Level Reading Measures (Grades K–1), 2009–2010 Version	8
<b>1004:</b> Technical Adequacy of the easyCBM <sup>®</sup> Grade 2 Reading Measures (2009–2010 version)	9
<b>1403:</b> easyCBM <sup>®</sup> Beginning Reading Measures: Grades K-1 Alternate Form Reliability and Criterion Validity with the SAT-10	10
<b>1201:</b> Analyzing the Reliability of the easyCBM <sup>®</sup> Reading Comprehension Measures: Grade 2	16
<b>1202:</b> Analyzing the Reliability of the easyCBM <sup>®</sup> Reading Comprehension Measures: Grade 3	18
<b>1203:</b> Analyzing the Reliability of the easyCBM <sup>®</sup> Reading Comprehension Measures: Grade 4	20
<b>1204:</b> Analyzing the Reliability of the easyCBM <sup>®</sup> Reading Comprehension Measures: Grade 5	22
<b>1205:</b> Analyzing the Reliability of the easyCBM <sup>®</sup> Reading Comprehension measures: Grade 6	23
<b>1206:</b> Reliability of the easyCBM <sup>®</sup> Reading Comprehension: Grade 7	25
<b>1216:</b> An Examination of Test-Retest, Alternate Form Reliability, and Generalizability Theory Study of the easyCBM <sup>®</sup> Reading Assessments: Grade 1	27
<b>1217:</b> An Examination of Test–retest, Alternate Form Reliability, and Generalizability Theory Study of the easyCBM <sup>®</sup> Reading Assessments: Grade 2	29
<b>1218:</b> An Examination of Test–retest, Alternate Form Reliability, and Generalizability Theory Study of the easyCBM <sup>®</sup> Word and Passage Reading Fluency Assessments: Grade 3	31
<b>1219:</b> An Examination of Test–retest, Alternate Form Reliability, and Generalizability Theory Study of the easyCBM <sup>®</sup> Word and Passage Reading Fluency Assessments: Grade 4	33
<b>1220:</b> Test–retest, Alternate Form Reliability: Grade 5	35
<b>1406:</b> Internal Consistency of easyCBM <sup>®</sup> Vocabulary Measures: Grades 2–8	38
<b>1407:</b> Internal Consistency of the easyCBM <sup>®</sup> CCSS Reading Measures Grades K-8	39
<b>1807:</b> Supplementary Report on easyCBM MCRC Measures: A Follow-up	40
<b>1803:</b> Reliability of the Slope of the easyCBM <sup>®</sup> Reading Measures	40
<b>1111:</b> easyCBM <sup>®</sup> Slope Reliability: Letter Names, Word Reading Fluency, and Passage Reading Fluency	42
<b>2602:</b> Reliability Analyses for easyCBM <sup>®</sup> Measures in Grades 3-8: Total Scores for Vocabulary, Proficient Reading, and Proficient Math	43
<b>1701:</b> Oral reading fluency: Outcomes from 30 years of research	44
<b>Appendix A: Technical Report Table Titles</b>	<b>46</b>
<b>Appendix B: Guide to Spreadsheet Technical Report Value Displays</b>	<b>47</b>
<b>Technical Report References</b>	<b>48-49</b>

---

<sup>1</sup> For individual technical reports, see <https://brtprojects.org>: BRT Tech Reports / Search here...

## Reliability

Reliability refers to the degree to which an assessment yields scores that are consistent, precise, and reproducible for a defined purpose, population, and set of testing conditions. In educational measurement, reliability is not a fixed property of a test “in general”; it is an empirical characteristic of scores produced in a particular administration and interpreted for decisions. When reliability is high, observed score differences are more likely to reflect true differences in student performance on the construct being measured rather than random fluctuations in testing conditions, item sampling, or scoring.

Reliability is often described using observed-score theory, where an observed score is viewed as a combination of a true score plus measurement error. Error can arise from many sources: sampling a limited set of items from a broad content domain; day-to-day variability in student attention, fatigue, or motivation; differences in administration conditions (time limits, directions, access to tools, testing environment); and differences in scoring procedures. Because no single study isolates every potential source of error at once, educational assessment programs typically assemble a “reliability argument” by reporting multiple indices that are well matched to how scores will be used (e.g., screening classification, progress monitoring growth, or program evaluation).

### Types of Reliability

Alternate-form reliability (parallel-forms reliability) evaluates the consistency of scores across two equivalent forms designed to measure the same content and skills. Alternate forms are constructed from a common blueprint—content balance, difficulty range, item formats, time limits, and scoring rules—and administered under comparable conditions, often close in time. The correlation between Form A and Form B scores provides evidence that students would obtain similar results regardless of which form they received. In progress monitoring, alternate-form evidence is especially important because students are tested repeatedly and score change should reflect learning rather than differences among forms.

Test–retest reliability evaluates score stability over time by administering the same form (or a highly similar form) to the same students on two occasions separated by a specified interval. The resulting correlation estimates the extent to which students maintain their relative standing over that time window. Test–retest evidence is most informative when the construct is expected to be relatively stable across the retest interval and when administration conditions are held constant. If the interval is long enough for meaningful learning to occur, stability can decrease for substantive reasons; therefore, test–retest designs require careful interpretation and an interval aligned to the intended use (e.g., short-term stability vs. sensitivity to growth).

Internal consistency reliability addresses the extent to which items within a single form function together to measure a common underlying construct. For dichotomously scored items common in mathematics screeners, Cronbach’s alpha is frequently reported as an index of how well items cohere and how much precision is gained by aggregating across items. Internal consistency tends to increase when a test includes more informative items, when item difficulties are appropriately distributed (not all too easy or too hard), and when items discriminate among students in the target achievement range. Internal consistency is typically strongest for the total score and lower for short subtests, because fewer items provide less opportunity to average out item-sampling error.

Inter-judge (inter-rater) reliability applies when human judgment contributes to scoring—for example, scoring constructed responses, applying rubric criteria, or recording errors during performance tasks. It quantifies the extent to which different scorers assign the same score to the same student work under standardized scoring rules. Depending on the score scale and scoring design, inter-judge reliability may be estimated using percent agreement, Cohen’s kappa, intraclass correlation coefficients, or generalizability approaches. Strong inter-judge reliability is essential because inconsistent scoring introduces error that can overwhelm the precision gained from well-designed items.

These types of reliability are complementary rather than competing. Benchmark screening often prioritizes internal consistency to support dependable total scores at a single time point, while progress monitoring prioritizes alternate-form evidence to ensure equivalence across repeated administrations. Test–retest evidence clarifies short-term stability when the same form is reused or when interpreting change over brief windows. When scoring involves judgment, inter-judge evidence becomes essential because it can limit the maximum reliability attainable by any other design.

### Why Reliability is Important

Reliable scores are essential because educational decisions require separating true differences in student performance from random error. When reliability is low, screening cut scores and growth targets become unstable, confidence intervals widen, and students may be misclassified—either missing needed support or being placed in interventions unnecessarily. Reliable measurement reduces overreaction to chance score swings, strengthens progress-monitoring decisions, and improves fairness by providing consistent information across classrooms, schools, and testing occasions. It also increases statistical power for evaluating programs and policies and supports more credible validity arguments about what scores mean and how they should be used<sup>2</sup>.

This technical report focuses on reliability of outcomes from different easyCBMs<sup>®</sup> used to document early reading performance: The summaries and findings represent compilations from previously published technical reports so that reviewers can use this document as a ‘quick guide’ and then access the full report at the website for Behavioral Research and Teaching (BRT: <https://brtprojects.org>). Several reliability estimates are provided for different easyCBMs<sup>®</sup> used to document early reading performance: **Note:** All tables and figures in this summary are examples of those presented in full within the individual Technical Reports but are not exhaustive, just illustrative: The summaries and findings represent compilations from previously published technical reports so that reviewers can use this document as a ‘quick guide’ and then access the full report at the website for Behavioral Research and Teaching (BRT: <https://brtprojects.org>). The measures included Letter Names (LN), Letter Sounds (LS), Phoneme Segmenting (PS), Word Reading Fluency (WRF), and Passage Reading Fluency (PRF). Reliability analyses included documenting internal consistency, alternate form, test-retest, and generalizability. The results indicate positive results for the various measures when administered at different benchmark seasons.

### Preview of Technical Report Findings

The studies summarized in this document provide an integrated picture of score dependability for easyCBM<sup>®</sup> reading measures across grades and measure types. These findings are illustrative: this report synthesizes results from primary studies conducted by Behavioral Research and Teaching (BRT) researchers, with the full set of technical reports available at <https://brtprojects.org>. Across the body of work, reliability is consistently strongest for individually administered, fluency-based measures (Word Reading Fluency and Passage Reading Fluency), while measures that rely on fewer items, group administration, or broader constructs (e.g., multiple-choice reading comprehension) show more modest consistency and benefit from additional design and analytic supports.

A central cross-cutting finding is that fluency measures tend to yield highly stable scores across both alternate forms and short retest intervals. Across grades and studies, correlations among parallel passages or word lists are typically high, indicating that forms behave comparably and that observed differences are largely attributable to student performance rather than passage difficulty. Similarly, one-week test-retest estimates are generally strong, supporting the interpretation that scores are sufficiently stable to support progress-monitoring decisions under standard conditions. Where coefficients are lower, results suggest that reliability is sensitive to the form set, sample size, and classroom-level testing context—factors managed through careful form construction and administration guidance.

A second integrated finding is that generalizability theory analyses reinforce and sharpen the classical reliability results. When persons, forms, and occasions are modeled simultaneously, the largest share of variance is typically attributable to differences among students, with relatively small contributions from forms and occasions. This is desirable: it indicates that the assessments are primarily measuring stable differences in reading skill rather than being driven by the specific passage or testing day. Decision studies extend this insight by showing how precision improves with additional observations. For many fluency applications, one form on one occasion can meet common reliability thresholds for relative decisions, but dependability increases meaningfully with multiple forms and/or occasions, particularly when the goal is an absolute decision (e.g., classification against a benchmark) or when local samples are small.

A third theme concerns how construct breadth and administration mode interact with reliability. Fluency measures involve timed performance on constrained tasks (word reading or passage reading rate/accuracy) and are scored with relatively objective rules, which supports consistency. In contrast, multiple-choice reading comprehension (MCRC)

---

<sup>2</sup> American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

assesses a broader construct and is often group administered, introducing additional sources of error (e.g., differences in testing conditions, attention, and proctoring). Early evidence shows lower alternate-form correlations for comprehension relative to fluency, underscoring that reliability depends on the nature of the score and the conditions under which it is collected.

A fourth integrated finding is that item- and form-level analyses help explain and improve comprehension reliability. Across grade-specific MCRC studies, multiple methods are used to evaluate form equivalence and score consistency (e.g., repeated-measures ANOVA for form comparisons, split-half estimates, top-bottom group checks, and Rasch modeling). These analyses commonly identify that overall item functioning is acceptable, with most items showing reasonable fit and distractor behavior, while isolated items exhibit misfit or unexpected discrimination patterns. A recurring constraint is test length: when forms contain relatively few items, person reliability can be limited, especially when score distributions show ceiling effects or restricted variability. These patterns suggest that reliability for comprehension can be strengthened by optimizing item targeting, increasing information where score variability is limited, and maintaining consistent administration practices.

A fifth theme comes from internal consistency evidence for vocabulary. Using very large benchmark datasets across grades and seasons, internal consistency estimates (e.g., Cronbach's alpha) are stable and generally in the acceptable-to-strong range, supporting the use of vocabulary measures for screening and instructional planning. The consistency of estimates across fall and winter administrations suggests that the measures function reliably across typical seasonal contexts. Split-half evidence and top-bottom comparisons further indicate that items discriminate as expected, with higher-performing students selecting correct options at higher rates.

Taken together, these studies support the conclusion that easyCBM<sup>®</sup> reading measures can provide dependable scores for key assessment uses when aligned to appropriate designs. Fluency measures show strong evidence for alternate-form consistency and short-term stability, making them well suited to frequent progress monitoring. Comprehension and other item-based measures demonstrate acceptable reliability when evaluated with multiple complementary methods, but their precision is more sensitive to test length, score distributions, and administration conditions; accordingly, they benefit from continued refinement and from measurement designs that reduce error (e.g., additional items, multiple forms, or repeated occasions when needed). Overall, the integrated evidence base emphasizes a practical message: reliability is not only a coefficient, but a function of how measures are built and used, and these reports provide concrete guidance for selecting and administering measures to achieve dependable decisions.

Across the fluency studies that incorporate D-studies, a consistent implication is that measurement precision can be strategically increased by adding parallel forms and/or occasions, with diminishing returns after a small number of administrations. For relative (norm-referenced) decisions, many analyses show that one form on one occasion is often adequate for PRF, while WRF may benefit more from multiple forms, particularly in contexts where classroom samples are small or where the student-by-form interaction is nontrivial. For absolute (criterion-referenced) interpretations, Phi coefficients and associated SEMs highlight that additional administrations can meaningfully reduce error around cut scores, which is directly relevant when schools make risk classifications tied to interventions.

Another integrated pattern is that “problematic” results, when they occur, tend to be form-level or item-level rather than system-wide. For example, occasional lower correlations for a particular passage form suggest that a single passage may be unusually easy, hard, or sensitive to administration nuances. Similarly, isolated comprehension items may show misfit or unexpected top-bottom patterns, indicating the need for revision or replacement. The broader reliability picture remains positive because most forms and items behave as intended, and the analytic approaches used across the technical reports are designed to detect these exceptions so that form pools can be improved over time.

---

---

**Table 1. Grade-Level Benchmark Measures by Assessment Period**

Grade Level	Fall Benchmark	Winter Benchmark	Spring Benchmark
K	Letter Names	Letter Sounds	Letter Sounds
	Letter Sounds	Phoneme Segmenting	Phoneme Segmenting
	Phoneme Segmenting	Word Reading Fluency	Word Reading Fluency
1	Letter Sounds	Letter Sounds	Letter Sounds
	Phoneme Segmenting	Word Reading Fluency	Word Reading Fluency
	Word Reading Fluency	Passage Reading Fluency	Passage Reading Fluency
2	Passage Reading Fluency	Passage Reading Fluency	Passage Reading Fluency
	Vocabulary	Vocabulary	Vocabulary
	Proficient Reading	Proficient Reading	Proficient Reading

The measures in **Grades 3 through 8** consist of Vocabulary, Passage Reading Fluency, and two measures of comprehension: Basic and Proficient.

---

**Summary of Technical Report 0906:** Alternate Form and Test–Retest Reliability of easyCBM<sup>®</sup> Reading Measures (Alonzo & Tindal, 2009).

### Method

Approximately 50 students per grade participated in the study, all drawn from a mid-sized K–8 school in the Pacific Northwest. The school included diverse demographics and had reading performance levels slightly above district and state averages. All students present on testing days participated in the study. Measures examined varied by grade level. In Grade 1, the study included Phoneme Segmenting, Letter Names, Letter Sounds, Word Reading Fluency, and Passage Reading Fluency. In Grade 3, Word Reading Fluency, Passage Reading Fluency, and Multiple-Choice Reading Comprehension were examined. In Grades 5 and 8, Passage Reading Fluency and Reading Comprehension were included. Across grades, multiple alternate forms of each measure were selected, representing roughly 40% of the forms available in the easyCBM<sup>®</sup> assessment system.

Data collection occurred during two testing sessions administered one week apart. In the first session, students completed several forms of each measure type. In the second session, students repeated selected forms from the first session and completed additional alternate forms. Most measures were individually administered by trained research assistants, while reading comprehension tests were group administered by classroom teachers. Scores were recorded according to standardized scoring rules specific to each assessment type.

### Results

Analyses included descriptive statistics, bivariate correlations to estimate alternate-form reliability, and correlations across testing sessions to estimate test–retest reliability. Standard errors of measurement were also calculated using median reliability coefficients and average standard deviations. Results indicated strong alternate-form reliability across most reading fluency measures. In Grade 1, alternate-form correlations ranged from .76 to .97 across measures, with the highest coefficients observed for Word Reading Fluency and Passage Reading Fluency (.95–.97). In Grade 3, Word Reading Fluency correlations ranged from .87 to .93 and Passage Reading Fluency from .94 to .95. Passage Reading Fluency alternate-form correlations ranged from .87 to .96 in Grade 5 and .87 to .95 in Grade 8. Reading comprehension alternate forms produced lower correlations (.26–.59). Test–retest correlations ranged from .45 to .97 across measures and grades, with the highest values observed for Word Reading Fluency and Passage Reading Fluency. Overall, the findings demonstrate strong reliability for the easyCBM<sup>®</sup> fluency measures across grade levels.

**Table 2. Reliability for Various Measures Across Grades**

Grade	Measures Studied	Alternate Form Reliability (r)	Test–Retest Reliability (r)
Grade 1	SEG, LN, LS, WRF, PRF	.76 – .97	.45 – .97
Grade 3	WRF, PRF, MCRC	.87 – .95 (fluency) .26 (MCRC)	.92 – .94
Grade 5	PRF, MCRC	.87 – .96 (PRF) .59 (MCRC)	.92 – .94
Grade 8	PRF, MCRC	.87 – .95 (PRF) .35 (MCRC)	.91 – .91

### Reference

Alonzo, J., & Tindal, G. (2009). *Alternate form and test-retest reliability of easyCBM<sup>®</sup> reading measures (Technical Report No. 0906)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

---

**Summary of Technical Report 1003:** Technical Adequacy of the easyCBM<sup>®</sup> Primary-Level Reading Measures (Grades K–1), 2009–2010 Version (Lai et al., 2010).

### Methods

A set of studies examine technical adequacy for the easyCBM<sup>®</sup> early-reading progress monitoring measures used in Kindergarten and Grade 1 during the 2009–2010 school year. Data were drawn from a convenience sample of three Pacific Northwest districts (68 schools contributing Kindergarten data; 71 schools contributing Grade 1 data). All students present during testing windows participated. Sample sizes were large for each measure (Kindergarten:  $n \approx 1,984$ – $1,987$  for Segmenting and Letter Sounds;  $n \approx 1,985$  for Word Reading; Grade 1:  $n \approx 953$  for Segmenting;  $n \approx 2,179$ – $2,181$  for Letter Sounds and Word Reading;  $n \approx 2,180$  for Passage Reading). Measures were individually administered 60-second fluency probes, scored as correct phonemes, letter sounds, or words per minute, across fall, winter, and spring. Reliability of growth (slope) was evaluated using two-level hierarchical linear growth models (time at level 1, student at level 2). Growth reliability (for intercept and slope) was summarized by grade and by initial status groups, primarily quartiles based on fall performance (with Word Reading in Kindergarten summarized by lower vs upper 50th percentile).

### Results

**Kindergarten.** For the Kindergarten phoneme segmenting measure, the reliability of the growth slope for students in the first quartile was .67, the reliability for students in the second quartile was .66, the reliability for students in the third quartile was .46, and the reliability for students in the fourth quartile was .09. For the Kindergarten LS measure, the reliability of the growth slope for students in the first quartile was .76, the reliability for students in the second quartile was .76, the reliability for students in the third quartile was .68, and the reliability for students in the fourth quartile was .53. For the Kindergarten WRF measure, the reliability of the growth slope for students in the lower 50th percentile was .82, and the reliability for students in the upper 50th percentile was .71.

**Grade 1.** For the phoneme segmenting measure, the reliability of the growth slope for students in the first quartile was .28. No other reliability estimates were possible for this measure in grade 1. For the grade 1 LS measure, the reliability of the growth slope for students in the first quartile was .57, the reliability for students in the second quartile was .66, the reliability for students in the third quartile was .72, and the reliability for students in the fourth quartile was .42.

**Grade 1.** For the WRF measure, the reliability of the growth slope for students in the first quartile was .88, the reliability for students in the second quartile was .82, the reliability for students in the third quartile was .76, and the reliability for students in the fourth quartile was .63. For the grade 1 PRF measure, the reliability of the growth slope for students in the first quartile was .93, the reliability for students in the second quartile was .88, the reliability for students in the third quartile was .89, and reliability for students in the 4th quartile was .60.

## Reference

Lai, C.-F., Nese, J. F. T., Jamgochian, E. M., Alonzo, J., & Tindal, G. (2010). *Technical Adequacy of the easyCBM<sup>®</sup> Primary-Level Reading Measures (Grades K–1), 2009–2010 Version (Technical Report 1003)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

---

**Summary of Technical Report 1004:** Technical Adequacy of the easyCBM<sup>®</sup> Grade 2 Reading Measures (2009–2010 version) (Jamgochian et al., 2010).

## Methods

This report presents reliability and validity evidence for the easyCBM<sup>®</sup> Grade 2 reading measures from the 2009–2010 school year. Data were collected from 71 schools across three Pacific Northwest districts. All students in attendance during the assessment windows participated. Seasonal samples were large for the easyCBM<sup>®</sup> measures (approximately 2,144–2,301 students for multiple-choice reading comprehension; ~2,154–2,207 for word reading fluency; and ~2,205–2,236 for passage reading fluency across fall, winter, and spring). The Grade 2 measures included: (a) Word Reading Fluency (WRF; 60-second timed word list), (b) Passage Reading Fluency (PRF; 60-second oral passage reading with errors marked), and (c) Multiple-Choice Reading Comprehension (MCRC; a computer-delivered narrative passage with 12 multiple-choice items). Alternate-form reliability and content development were addressed in prior technical reports; this report focused on reliability of performance-level scores (evaluated for MCRC using Cronbach’s alpha and split-half methods) and reliability of growth (slope) for WRF, PRF, and MCRC using two-level hierarchical linear growth models across fall–winter–spring. Validity evidence included concurrent and predictive relations with SAT-10 (correlations and regression models), construct validity via confirmatory factor analysis (three-factor model: WRF, PRF, MCRC), and predictive validity of growth (correlating growth-model residuals with SAT-10).

## Results

Reliability evidence for performance-level scores was reported for MCRC (12 items). Across fall, winter, and spring, Cronbach’s alpha ranged from .679 to .748, indicating moderate internal consistency. Split-half analyses (first 6 vs. last 6 items) yielded correlations from .499 to .599, with Spearman–Brown/Guttman coefficients in the mid-.60s to mid-.70s, again suggesting moderate reliability.

Reliability of growth (slope) was generally moderate and varied by initial status (fall quartiles). For WRF, slope reliability declined from .77 (lowest quartile) to .50 (highest quartile). For PRF, slope reliability was .75, .70, and .68 in quartiles 1–3 but dropped to .25 in quartile 4. For MCRC, slope reliability was .48 (quartile 1) and .62 (quartile 2), could not be estimated in quartile 3 due to convergence problems, and was low in quartile 4 (.14). The discussion attributes the weakest slope reliability in the highest quartile to reduced growth/possible ceiling effects.

**Note:** Technical Report 1005 also reports on reliability for reading measures used in Grades 3–7: internal consistency for MCRC and slope for WRF, PRF, and MCRC. This report is summarized in 2603-VK8R\_VValidityReading with corresponding values in TR2603\_VK8R\_VValidityReadingTRs21-31.xlsx. The reason for this assignment is that the more important values focus on concurrent validity of these measures with the OAKS reading measures.

## Reference

Jamgochian, E. M., Park, B. J., Nese, J. F. T., Lai, C. F., Sáez, L., Anderson, D., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM<sup>®</sup> grade 2 reading measures (Technical Report No. 1004)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

---

**Summary of Technical Report 1403:** easyCBM<sup>®</sup> Beginning Reading Measures: Grades K-1 Alternate Form Reliability and Criterion Validity with the SAT-10 (Wray, Lai, et al., 2014).

Technical Report 1403 examined the alternate-form reliability and criterion-related validity of easyCBM<sup>®</sup> beginning reading measures for kindergarten and grade 1 students. Participants included 222 kindergarten students and 204 first-grade students from six kindergarten and eight first-grade classrooms across three K–5 schools in a semi-rural Pacific Northwest district. All students present during testing participated. Data collection was conducted by 15 trained research assistants who received structured training and periodic refresher sessions to ensure standardized administration and scoring procedures.

**Method**

The easyCBM<sup>®</sup> measures examined included Letter Names (LN), Letter Sounds (LS), Phoneme Segmenting (PS), and Word Reading Fluency (WRF). Students were assessed at five time points throughout the school year (November, December, February, April, and June). At each time point students completed two alternate forms of the same measure presented in counterbalanced order to estimate alternate-form reliability. Assessments were individually administered in quiet testing rooms and typically required about 10 minutes for kindergarten and 7 minutes for grade 1 students. Criterion measures were drawn from the Stanford Early School Achievement Test / Stanford Achievement Test–10 (SESAT/SAT-10). Kindergarten students completed the Sounds and Letters and Word Reading subtests, while grade 1 students completed Word Reading and Word Study Skills subtests in the spring administration.

Data were screened for outliers using boxplots, histograms, scatterplots, and z-score comparisons. Pearson correlations were used to estimate alternate-form reliability when assumptions of normality were met, while Spearman rank correlations were used when distributional assumptions were violated. Criterion validity was evaluated using standard multiple regression models predicting SAT-10 outcomes from the easyCBM<sup>®</sup> measures.

**Results**

Moderate to strong alternate-form reliability was found In kindergarten, reliability coefficients ranged from .31 to .94 across measures and time points, with stronger coefficients for Letter Names (.61–.90), Letter Sounds (.54–.92), and Word Reading Fluency (.74–.94). Phoneme Segmenting showed greater variability (.31–.90). In grade 1, alternate-form coefficients ranged from .27 to .96 with generally strong relations for Letter Names, Letter Sounds, and Word Reading Fluency. Regression analyses indicated that the combined easyCBM measures explained 35–58% of the variance in kindergarten SAT-10 outcomes and 14–32% of the variance in grade 1 SAT-10 outcomes. In addition, grade 1 early literacy measures explained 49–56% of the variance in end-of-year easyCBM Word Reading Fluency performance, supporting their usefulness for screening and progress monitoring.

- Kindergarten: LN coefficients ranged from .61–.90
- Kindergarten: LS coefficients ranged from .54–.92
- Kindergarten: PS coefficients ranged from .31–.90
- Kindergarten: WRF coefficients ranged from .74–.94
- Grade 1: These measures had coefficients that ranged from .26 – .96

**Reference**

Wray, K. A., Lai, C.-F., Sáez, L., Alonzo, J., & Tindal, G. (2013). *easyCBM<sup>®</sup> beginning reading measures: K-1 alternate form reliability and criterion validity with the SAT-10 (Technical Report No. 1403)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

---

**Table 3. Individual Coefficients from Technical Reports**

Grade	Measure	Reliability Type	Coefficients	TR#	Page	Median
K	LN	alternate_form	0.87, 0.82, 0.84, 0.76, 0.77, 0.66, 0.68, 0.61, 0.64, 0.84, 0.87, 0.77, 0.77, 0.69, 0.74, 0.62, 0.67, 0.89, 0.83, 0.81, 0.74, 0.74, 0.65, 0.71, 0.83, 0.81, 0.72, 0.7, 0.64, 0.68, 0.88, 0.8, 0.79, 0.73, 0.79, 0.78, 0.74, 0.74, 0.76, 0.9, 0.84, 0.85, 0.83, 0.84, 0.9	1403	17	0.77
K	LS	alternate_form	0.89, 0.81, 0.86, 0.75, 0.74, 0.64, 0.58, 0.6, 0.57, 0.78, 0.83, 0.72, 0.69, 0.61, 0.53, 0.55, 0.54, 0.92, 0.81, 0.82, 0.67, 0.61, 0.63, 0.62, 0.82, 0.82, 0.71, 0.67, 0.69, 0.66, 0.92, 0.79, 0.77, 0.77, 0.75, 0.83, 0.77, 0.79, 0.75, 0.87, 0.85, 0.88, 0.89, 0.87, 0.91	1403	18	0.77
K	PS	alternate_form	0.81, 0.42, 0.51, 0.4, 0.39, 0.42, 0.39, 0.48, 0.44, 0.49, 0.52, 0.37, 0.35, 0.39, 0.39, 0.49, 0.41, 0.9, 0.43, 0.45, 0.45, 0.41, 0.4, 0.31, 0.41, 0.52, 0.53, 0.48, 0.49, 0.43, 0.83, 0.34, 0.42, 0.38, 0.35, 0.46, 0.51, 0.54, 0.48, 0.83, 0.59, 0.51, 0.57, 0.5, 0.86	1403	19	0.45
K	RAN					
K	WRF	alternate_form	0.84, 0.81, 0.83, 0.75, 0.76, 0.77, 0.75, 0.75, 0.74, 0.77, 0.78, 0.7, 0.77, 0.8, 0.78, 0.79, 0.76, 0.91, 0.86, 0.84, 0.85, 0.84, 0.82, 0.81, 0.89, 0.87, 0.88, 0.85, 0.84, 0.82, 0.9, 0.86, 0.84, 0.85, 0.81, 0.86, 0.86, 0.86, 0.84, 0.92, 0.9, 0.89, 0.91, 0.9, 0.94	1403	20	0.84
1	LN	alternate_form	0.88, 0.84, 0.82, 0.75, 0.74, 0.69, 0.68, 0.67, 0.66, 0.85, 0.82, 0.75, 0.75, 0.69, 0.73, 0.7, 0.67, 0.9, 0.78, 0.77, 0.75, 0.78, 0.75, 0.73, 0.77, 0.75, 0.74, 0.76, 0.7, 0.7, 0.85, 0.81, 0.81, 0.81, 0.8, 0.82, 0.81, 0.81, 0.76, 0.89, 0.84, 0.82, 0.86, 0.85, 0.87	1403	21	0.77
1	LN	alternate_form	0.89, 0.82, 0.85	0906	10	0.85
1	LS	alternate_form	0.87, 0.89, 0.83, 0.82, 0.89, 0.85	1216	12	0.86
1	LS	alternate_form	0.76, 0.76, 0.88	0906	10	0.76
1	LS	alternate_form	0.80, 0.64, 0.40, 0.49, 0.71, 0.49, 0.55, 0.71, 0.75, 0.83	1403	22	0.68
1	LS	g coefficient	0.87, 0.95	1216	12	0.91

1	LS	test_retest	0.77, 0.83, 0.86, 0.87	1216	12	0.85
1	PRF	alternate_form	0.98, 0.98, 0.98, 0.96, 0.95, 0.93	1216	49	0.97
1	PRF	alternate_form	0.95, 0.96, 0.97 0.95, 0.95, 0.95, 0.96, 0.96, 0.96, 0.96, 0.96, 0.95, 0.95, 0.96, 0.93, 0.93, 0.92, 0.94, 0.94, 0.93, 0.92, 0.91, 0.93, 0.94, 0.91, 0.95, 0.95, 0.95, 0.96, 0.96, 0.93, 0.94, 0.92, 0.92, 0.92, 0.93, 0.93, 0.91, 0.91, 0.93, 0.93, 0.93, 0.93, 0.94, 0.93, 0.91, 0.91, 0.94, 0.92, 0.93, 0.93, 0.93, 0.93, 0.93, 0.92, 0.9, 0.93, 0.91, 0.96, 0.93, 0.94, 0.92, 0.93, 0.92, 0.92, 0.89, 0.93, 0.92, 0.97, 0.96, 0.92, 0.92, 0.92, 0.92, 0.92, 0.91, 0.91, 0.92, 0.91, 0.95, 0.95, 0.96, 0.91, 0.91, 0.91, 0.92, 0.91, 0.91, 0.88, 0.91, 0.9, 0.95, 0.94, 0.96, 0.94, 0.92, 0.93, 0.92, 0.94, 0.93, 0.92, 0.91, 0.93, 0.91, 0.96, 0.95, 0.97, 0.96, 0.96, 0.94, 0.94, 0.93, 0.94, 0.93, 0.92, 0.91, 0.94, 0.92, 0.96, 0.96, 0.96, 0.95, 0.96, 0.97, 0.91, 0.92, 0.93, 0.93, 0.92, 0.89, 0.9, 0.93, 0.91, 0.93, 0.94, 0.95, 0.96, 0.95, 0.95, 0.95, 0.91, 0.9, 0.91, 0.9, 0.91, 0.89, 0.89, 0.91, 0.88, 0.93, 0.93, 0.93, 0.94, 0.93, 0.94, 0.94, 0.92, 0.95, 0.94, 0.94, 0.94, 0.94, 0.92, 0.92, 0.94, 0.92, 0.96, 0.96, 0.96, 0.95, 0.96, 0.97, 0.97, 0.95, 0.94, 0.94, 0.94, 0.94, 0.94, 0.94, 0.93, 0.92, 0.94, 0.92, 0.96, 0.96, 0.96, 0.96, 0.95, 0.96, 0.96, 0.94, 0.94, 0.97	0906	11	0.96
1	PRF	alternate_form	0.91, 0.88, 0.93, 0.93, 0.93, 0.94, 0.93, 0.94, 0.94, 0.92, 0.95, 0.94, 0.94, 0.94, 0.94, 0.92, 0.92, 0.94, 0.92, 0.96, 0.96, 0.96, 0.95, 0.96, 0.97, 0.97, 0.95, 0.94, 0.94, 0.94, 0.94, 0.94, 0.94, 0.93, 0.92, 0.94, 0.92, 0.96, 0.96, 0.96, 0.96, 0.95, 0.96, 0.96, 0.94, 0.94, 0.97	40	39	0.93
1	PRF	g_coefficient	0.99, 0.91	1216	50	0.95
1	PRF	test_retest	0.98, 0.98, 0.83, 0.95	1216	49	0.97
1	PS	alternate_form	0.89, 0.82, 0.78, 0.62, 0.67	1216	24	0.78
1	PS	alternate_form	0.42, 0.41, 0.14, 0.27, 0.38, 0.31, 0.45, 0.33, 0.41, 0.45	1403	22	
1	PS	alternate_form	0.91, 0.86, 0.91	0906	10	0.91
1	PS	g_coefficient	0.83, 0.50, 0.72	1216	24-35	0.72
1	PS	test_retest	0.56, 0.50, 0.81, 0.58, 0.72, 0.32	1216	23	0.57
1	WRF	alternate_form	0.97, 0.91, 0.95, 0.89	1216	38	0.93
1	WRF	alternate_form	0.96, 0.89, 0.79, 0.73, 0.94, 0.85, 0.80, 0.92, 0.87, 0.94	1403	23	0.88
1	WRF	alternate_form	0.96, 0.95, 0.95	0906	11	
1	WRF	g_coefficient	0.98, 0.96	1216	38	0.97
1	WRF	test_retest	0.93, 0.95, 0.87, 0.91	1216	38	0.92

2	PRF	alternate_form	0.93, 0.95, 0.92, 0.94, 0.92, 0.91 0.92, 0.93, 0.94, 0.93, 0.94, 0.94, 0.91, 0.93, 0.93, 0.94, 0.92, 0.94, 0.95, 0.95, 0.94, 0.93, 0.92, 0.93, 0.95, 0.92, 0.95, 0.93, 0.93, 0.94, 0.96, 0.93, 0.96, 0.95, 0.92, 0.93, 0.94, 0.95, 0.93, 0.95, 0.94, 0.94, 0.92, 0.92, 0.93, 0.94, 0.92, 0.94, 0.94, 0.94, 0.95, 0.91, 0.91, 0.92, 0.94, 0.92, 0.94, 0.94, 0.94, 0.94, 0.95, 0.92, 0.92, 0.93, 0.95, 0.92, 0.94, 0.94, 0.95, 0.94, 0.96, 0.95, 0.9, 0.9, 0.91, 0.91, 0.9, 0.92, 0.91, 0.92, 0.91, 0.91, 0.91, 0.91, 0.91, 0.93, 0.94, 0.94, 0.91, 0.94, 0.93, 0.94, 0.94, 0.94, 0.93, 0.93, 0.93, 0.93, 0.93, 0.95, 0.95, 0.92, 0.94, 0.94, 0.95, 0.95, 0.94, 0.94, 0.95, 0.93, 0.96, 0.91, 0.92, 0.94, 0.94, 0.92, 0.94, 0.93, 0.94, 0.93, 0.92, 0.93, 0.93, 0.92, 0.95, 0.95, 0.9, 0.91, 0.92, 0.93, 0.91, 0.94, 0.93, 0.95, 0.92, 0.94, 0.94, 0.95, 0.91, 0.93, 0.94, 0.93, 0.91, 0.91, 0.91, 0.93, 0.89, 0.93, 0.92, 0.93, 0.91, 0.92, 0.92, 0.93, 0.89, 0.91, 0.93, 0.91, 0.94, 0.92, 0.93, 0.93, 0.93, 0.92, 0.94, 0.93, 0.94, 0.93, 0.94, 0.93, 0.93, 0.91, 0.94, 0.94, 0.92, 0.94, 0.93, 0.93, 0.92, 0.92, 0.94, 0.92, 0.95, 0.96, 0.95, 0.94, 0.94, 0.95, 0.95, 0.92, 0.93, 0.95, 0.93, 0.95, 0.94, 0.94	1217	23	0.925
2	PRF	alternate_form	0.95, 0.94, 0.94, 0.95, 0.93, 0.96, 0.91, 0.92, 0.94, 0.94, 0.92, 0.94, 0.93, 0.94, 0.93, 0.92, 0.93, 0.93, 0.92, 0.95, 0.95, 0.9, 0.91, 0.92, 0.93, 0.91, 0.94, 0.93, 0.95, 0.92, 0.94, 0.94, 0.95, 0.91, 0.93, 0.94, 0.93, 0.91, 0.91, 0.91, 0.93, 0.89, 0.93, 0.92, 0.93, 0.91, 0.92, 0.92, 0.93, 0.89, 0.91, 0.93, 0.91, 0.94, 0.92, 0.93, 0.93, 0.93, 0.92, 0.94, 0.93, 0.94, 0.93, 0.94, 0.93, 0.93, 0.91, 0.94, 0.94, 0.92, 0.94, 0.93, 0.93, 0.92, 0.92, 0.94, 0.92, 0.95, 0.96, 0.95, 0.94, 0.94, 0.95, 0.95, 0.92, 0.93, 0.95, 0.93, 0.95, 0.94, 0.94	40	41	0.93
2	PRF	g_coefficient	0.97, 0.98	1217	24	0.98
2	PRF	test_retest	0.88, 0.96, 0.93, 0.90, 0.94, 0.95	1217	23	0.95
2	WRF	alternate_form	0.95, 0.92, 0.95, 0.94, 0.92, 0.92	1217	12	0.93
2	WRF	g_coefficient	0.99, 0.96	1217	12	0.98
2	WRF	test_retest	0.94, 0.95, 0.93, 0.89, 0.92, 0.87	1217	11	0.93

**Table 4. Median Coefficients from Technical Reports**

Grade	Measure	Reliability_type	Median	Tech Report	Page
K	LN	alternate_form	0.768	1403	17
K	LS	alternate_form	0.746	1403	18
K	PS	alternate_form	0.490	1403	19
K	WRF	alternate_form	0.830	1403	20
<hr/>					
1	LN	alternate_form	0.777	1403	21
1	LN	alternate_form	0.853	0906	10
1	LS	alternate_form	0.858	1216	12
1	LS	alternate_form	0.800	0906	10
1	LS	alternate_form	0.637	1403	22
1	LS	g_coefficient		1216	12
1	LS	test_retest	0.833	1216	12
1	PRF	alternate_form	0.963	1216	49
1	PRF	alternate_form	0.960	0906	11
1	PRF	alternate_form	0.933	40	39
1	PRF	g_coefficient		1216	50
1	PRF	test_retest	0.935	1216	49
1	PS	alternate_form	0.756	1216	24
1	PS	alternate_form	0.357	1403	22
1	PS	alternate_form	0.893	0906	10
1	PS	g_coefficient		1216	24-35
1	PS	test_retest	0.582	1216	23
1	WRF	alternate_form	0.930	1216	38
1	WRF	alternate_form	0.869	1403	23
1	WRF	alternate_form	0.953	0906	11
1	WRF	g_coefficient		1216	38
1	WRF	test_retest	0.915	1216	38
<hr/>					
2	PRF	alternate_form	0.928	1217	23
2	PRF	alternate_form	0.931	40	
2	PRF	g_coefficient		1217	24
2	PRF	test_retest	0.927	1217	23
2	WRF	alternate_form	0.933	1217	12
2	WRF	g_coefficient		1217	12
2	WRF	test_retest	0.917	1217	11

**Table 5. Grade Group Summary of Coefficients from Technical Reports**

<b>Grade</b>	<b>Measure</b>	<b>Reliability</b>	<b>Value</b>
K	LN	alternate_form	0.768
K	LS	alternate_form	0.746
K	PS	alternate_form	0.490
K	WRF	alternate_form	0.830
<hr/>			
1	LN	alternate_form	0.777
1	LN	alternate_form	0.853
1	LS	alternate_form	0.858
1	LS	alternate_form	0.800
1	LS	alternate_form	0.673
1	LS	test_retest	0.832
1	PRF	alternate_form	0.963
1	PRF	alternate_form	0.960
1	PRF	alternate_form	0.933
1	PRF	test_retest	0.935
1	PS	alternate_form	0.756
1	PS	alternate_form	0.357
1	PS	alternate_form	0.893
1	PS	test_retest	0.582
1	WRF	alternate_form	0.930
1	WRF	alternate_form	0.869
1	WRF	alternate_form	0.953
1	WRF	test_retest	0.915
<hr/>			
2	PRF	alternate_form	0.928
2	PRF	alternate_form	0.931
2	PRF	test_retest	0.927
2	WRF	alternate_form	0.933
2	WRF	test_retest	0.917

### Summary

Kindergarten has **one** type of reliability above lower bound of median (.70) for Letter Names (LN), Letter Sounds (LS), and Word Reading Fluency (WRF).

Grade 1 has **two** types of reliability above lower bound of median (.70) for Letter Sounds (LS), Passage Reading Fluency (PRF), and Word Reading Fluency (WRF). **One** type of reliability exists above lower bound of median (.70) for Letter Names and Phoneme Segmenting.

Grade 2 has **two** types of reliability above lower bound of median (.70) for both Passage Reading Fluency (PRF) and Word Reading Fluency (WRF).

**Summary of Technical Report 1201:** Analyzing the Reliability of the easyCBM<sup>®</sup> Reading Comprehension Measures: Grade 2 (C. F. Lai, P. S. Irvin, et al., 2012).

This report describes the reliability and technical performance of curriculum-based measurement (CBM) assessments used for screening and progress monitoring within a Response to Intervention (RTI) framework. The study emphasizes methodological design, statistical modeling, and empirical findings related to score consistency and measurement precision.

### Methods

Participants were drawn from an extant database of students who completed CBM assessments during a single academic year. The sample included students across multiple grade levels and represented a range of instructional needs, including general education and intervention populations. Students were included in analyses only if they had multiple observations on the same measure, ensuring adequate data for reliability estimation. Assessments were administered under standardized conditions using multiple equivalent forms designed to support repeated measurement.

Data preparation involved screening for incomplete records, removing off-grade testing, and addressing anomalous scores. Scores occurring within short time windows were averaged to reduce measurement error, and observations were organized into evenly spaced time segments to support growth-based analyses. Analytical methods combined descriptive statistics with reliability modeling. Internal consistency and stability of scores over time were examined using correlation-based approaches and growth modeling techniques. When longitudinal data were available, individual growth parameters were estimated and compared across parallel forms to evaluate the reliability of observed change. Model estimation relied on maximum likelihood procedures with robust standard errors to ensure accurate parameter estimation.

### Results

Results indicated that the CBM measures demonstrated acceptable to strong reliability across grade levels and assessment types. Score distributions showed sufficient variability to differentiate student performance, and reliability coefficients generally met or exceeded recommended thresholds for screening and progress monitoring. Growth estimates were stable, suggesting that observed changes in scores reflected true changes in student performance rather than measurement error. Overall, the findings support the technical adequacy of the measures for use in RTI decision-making and instructional planning.

**Table 6. Tests of Within-Subject Effects and Contrasts from Technical Report 1201**

Table 2

*Tests of Within-Subjects Effects for Grade 2 Measures 8 to 10*

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
forms	Sphericity Assumed	14.111	2	7.056	4.040	.021
	Greenhouse-Geisser	14.111	1.856	7.602	4.040	.024
	Huynh-Feldt	14.111	1.941	7.272	4.040	.022
	Lower-bound	14.111	1.000	14.111	4.040	.051
Error(forms)	Sphericity Assumed	143.222	82	1.747		
	Greenhouse-Geisser	143.222	76.105	1.882		
	Huynh-Feldt	143.222	79.563	1.800		
	Lower-bound	143.222	41.000	3.493		

Note. Mauchly's Test of Sphericity: The assumption of sphericity was not violated, Mauchly's W was 0.92,  $\chi^2(2) = 3.23, p > .05$ .

Table 3

*Tests of Within-Subjects Contrasts for Grade 2 Measures 8 to 10*

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
forms	Level 1 vs. Level 3	14.881	1	14.881	4.916	.032
	Level 2 vs. Level 3	25.929	1	25.929	5.807	.021
Error(forms)	Level 1 vs. Level 3	124.119	41	3.027		
	Level 2 vs. Level 3	183.071	41	4.465		

**Table 7. Sample Split-Half Coefficients from Technical Report 1201**

Table 3  
*Grade 2 Split-Half Coefficients for MCRC Form 9 with N = 12 Items*

Cronbach's Alpha	Part 1	Value	.667
		N of Items	6 <sup>a</sup>
	Part 2	Value	.520
		N of Items	6 <sup>b</sup>
	Total N of Items		12
Correlation Between Forms			.775
Spearman-Brown Coefficient	Equal Length		.873
	Unequal Length		.873
Guttman Split-Half Coefficient			.872

**Table 8. Distractor Analysis from Technical Report 1201**

Table 10  
*Distractor Analysis, Grade 2, Form 12*

Entry #	Data Code	Score Value	Count	%	Average Measure	S.E. Mean
1	A	1	48	94	3.00	0.22
	B	0	3	6	-0.18	0.54
	C	0	0	0	0.00	0.00
	Missing	**				
2	A	0	10	20	0.91	0.47
	B	0	2	4	0.20	0.68
	C	1	39	76	3.43	0.18
	Missing	**				
3	A	1	46	90	3.14	0.20
	B	0	4	8	-0.14	0.65
	C	0	1	2	-0.48	0.00
	Missing	**				
4	A	0	1	2	-1.46	0.00
	B	0	6	12	0.27	0.60
	C	1	44	86	3.26	0.18
	Missing	**				
5	A	1	50	98	2.87	0.23
	B	0	0	0	0	0
	C	0	1	2	-0.04	0
	Missing	**	0	0	0	0
6	A	0	2	4	1.09	2.03
	B	1	47	92	3.02	0.22
	C	0	2	4	-0.26	0.22
	Missing	**				
7	A	1	48	94	3.05	0.20
	B	0	2	4	-0.97	0.49
	C	0	1	2	-0.94	0.00
	Missing	**				
8	A	0	23	45	2.17	0.28
	B	1	27	53	3.48	0.30
	C	0	1	2	-0.48	0.00
	Missing	**				
9	A	0	3	6	0.75	0.42
	B	0	4	8	-0.25	0.60
	C	1	44	86	3.23	0.20
	Missing	**				
10	A	0	1	2	1.42	0.00
	B	1	43	84	3.11	0.23
	C	0	4	8	1.54	0.76
	Missing	**	3	6	0.69	1.09

**Table 9. Summary of Key Findings from Technical Report 1201**

Category	Summary
<b>Sample</b>	715 students from two Pacific Northwest public school districts participated. Students were primarily from Grade 2, with broader recruitment across Grades 2–8 to support form comparisons. Grade 2 analyses focused on approximately 44–52 students per form.
<b>Assessment Forms</b>	easyCBM <sup>®</sup> multiple-choice reading comprehension (MCRC) measures, Grade 2. Nine forms were analyzed (Forms 8–16). Each form included 12 items assessing literal and inferential comprehension.
<b>Analysis Method</b>	One-way repeated measures ANOVA (form equivalence), split-half reliability using the Guttman formula, top–bottom reliability analyses (23rd vs. 78th percentile groups), and Rasch modeling (item difficulty, fit statistics, person and item reliability, distractor analysis).
<b>Items Analyzed</b>	12 items per form (11 items for one form due to low variance). Items were designed with a range of difficulties and ordered from easiest to most difficult.
<b>Problematic Items</b>	A small number of items across forms showed misfit (mean square outfit > 1.5 or < 0.5). Some items exhibited unexpected performance in top–bottom analyses, including one item where lower-performing students outperformed higher-performing students. Most issues were isolated and identified via Rasch and distractor analyses.
<b>Item Fit</b>	Overall item fit was acceptable. Most items fell within the recommended Rasch outfit range (0.5–1.5). Distractor analyses generally showed appropriate functioning, with correct answers selected by higher-ability students.
<b>Overall Conclusion</b>	The Grade 2 easyCBM <sup>®</sup> MCRC measures demonstrated adequate reliability for screening and progress monitoring. While person reliability was sometimes limited due to short test length and ceiling effects, results support the technical adequacy of the measures, with recommendations for continued refinement and future research using larger samples.

### Reference

Lai, C.-F., Irvin, P. S., Alonzo, J., Park, B. J., & Tindal, G. (2012). *Analyzing the reliability of the easyCBM<sup>®</sup> reading comprehension measures: Grade 2 (Technical Report 1201)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1202:** Analyzing the Reliability of the easyCBM<sup>®</sup> Reading Comprehension Measures: Grade 3 (C. F. Lai et al., 2012).

Technical Report 1202 presents an evaluation of curriculum-based measurement (CBM) tools designed for use in screening and progress monitoring within a Response to Intervention (RTI) framework. The report focuses on the methodological procedures used to collect longitudinal student performance data and the statistical analyses conducted to determine the reliability and usefulness of the resulting scores.

### Methods

Participants were drawn from an extant dataset consisting of students across multiple grade levels who were administered CBM assessments during a single academic year. The sample included both general education students and students receiving supplemental or intensive instructional support. Inclusion criteria required students to have multiple observations on the same measure to allow for reliable estimation of score stability and growth. Assessments were administered under standardized conditions using alternate forms intended to be of equivalent difficulty. These forms allowed for repeated measurement without substantial practice effects. Data preparation procedures included removal of incomplete records, exclusion of off-grade administrations, and screening for anomalous scores. When multiple scores occurred within a short time window, values were averaged to reduce random measurement error. Observations were organized into consistent time intervals to support longitudinal

analyses. Analyses combined descriptive statistics with reliability-focused modeling techniques. Internal consistency and score stability were evaluated using correlation-based approaches. Where longitudinal data permitted, growth parameters were estimated to examine the reliability of change over time. Statistical models were estimated using maximum likelihood methods with robust standard errors to ensure stable and unbiased parameter estimates.

## Results

Results indicated that the CBM measures demonstrated acceptable to strong reliability across grades and assessment types. Score distributions showed sufficient variability to distinguish student performance levels, and reliability coefficients generally met recommended thresholds for screening and progress monitoring. Growth estimates were stable across repeated administrations, suggesting that observed score changes primarily reflected true changes in student performance rather than measurement error. Collectively, the findings support the technical adequacy of the measures for instructional decision-making within RTI systems.

**Table 10. Example Tests of Within-Subject Effect from Technical Report 1202**

Table 2

*Tests of Within-Subjects Effects for Grade 3 Measures 8 to 10*

		Type III Sum of				
Source		Squares	df	Mean Square	F	Sig.
forms	Sphericity Assumed	13.282	2	6.641	1.249	.293
	Greenhouse-Geisser	13.282	1.983	6.698	1.249	.292
	Huynh-Feldt	13.282	2.000	6.641	1.249	.293
	Lower-bound	13.282	1.000	13.282	1.249	.271
Error(forms)	Sphericity Assumed	404.051	76	5.316		
	Greenhouse-Geisser	404.051	75.357	5.362		
	Huynh-Feldt	404.051	76.000	5.316		
	Lower-bound	404.051	38.000	10.633		

Note. Mauchly's Test of Sphericity: The assumption of sphericity was not violated. Mauchly's W was 0.99,  $\chi^2(2) =$

0.32,  $p > .05$ .

**Table 11. Summary of Outcomes from Technical Report 1202**

Category	Summary
<b>Sample</b>	Approximately 700 students drawn from two public school districts in the Pacific Northwest. Grade 3 analyses focused on samples of roughly 44–52 students per form, with additional students included for form comparisons.
<b>Assessment Forms</b>	Grade 3 easyCBM <sup>®</sup> multiple-choice reading comprehension (MCRC) assessments. Nine forms were analyzed, each designed to be parallel in structure and difficulty for repeated administration.
<b>Analysis Method</b>	One-way repeated measures ANOVA to evaluate form equivalence; split-half reliability using the Guttman formula; top–bottom group comparisons (upper and lower performance percentiles); and Rasch modeling to examine item difficulty, fit statistics, person reliability, and distractor functioning.
<b>Items Analyzed</b>	12 multiple-choice reading comprehension items per form, ordered from easiest to most difficult and designed to assess literal and inferential comprehension skills.
<b>Problematic Items</b>	A small number of items showed misfit (outfit mean square values outside the recommended 0.5–1.5 range) or unexpected performance in top–bottom analyses. These items were flagged for potential revision.
<b>Item Fit</b>	Overall item fit was acceptable. Most items demonstrated appropriate Rasch fit and effective distractor functioning, with higher-performing students consistently selecting correct responses.
<b>Overall Conclusion</b>	Results supported the technical adequacy of the Grade 3 easyCBM <sup>®</sup> reading comprehension measures for screening and progress monitoring. Although person reliability was somewhat constrained by short test length, the measures demonstrated acceptable reliability and item functioning, with recommendations for continued refinement and larger-sample validation.

## Reference

Lai, C.-F., Irvin, P. S., Alonzo, J., Park, B. J., & Tindal, G. (2012). *Analyzing the reliability of the easyCBM<sup>®</sup> reading comprehension measures: Grade 3 (Technical Report 1202)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

---

### Summary of Technical Report 1203: Analyzing the Reliability of the easyCBM<sup>®</sup> Reading Comprehension Measures: Grade 4 (Park, Irvin, Alonzo, et al., 2012).

Technical Report 1203 investigates the technical adequacy of curriculum-based measurement (CBM) assessments used for screening and progress monitoring within a Response to Intervention (RTI) framework. The report focuses on the methodological procedures used to collect longitudinal student performance data and the statistical analyses conducted to evaluate score reliability and interpretability.

## Methods

Participants were drawn from a large, extant dataset of students who completed CBM assessments during a single academic year. The sample included students across multiple grade levels and represented both general education populations and students receiving supplemental or intensive instructional services. To ensure sufficient data for reliability and growth analyses, students were included only if they had multiple administrations of the same CBM measure.

Assessments were administered under standardized conditions using multiple alternate forms designed to be of comparable difficulty. These alternate forms allowed for repeated measurement while minimizing practice effects. Data preparation procedures included screening for incomplete records, excluding off-grade administrations, and removing anomalous scores. When multiple scores occurred within short time windows, values were averaged to reduce random measurement error. Observations were then organized into consistent time intervals to support longitudinal analyses. Analytical methods included descriptive statistics and reliability-focused modeling approaches. Internal consistency and score stability were evaluated using correlation-based methods. Where longitudinal data were available, growth parameters were estimated to examine the reliability of change over time. Statistical models were estimated using maximum likelihood procedures with robust standard errors to ensure stable parameter estimation.

## Results

Results indicated that the CBM measures demonstrated acceptable to strong reliability across grades and assessment types. Score distributions showed adequate variability to differentiate student performance levels, and reliability coefficients generally met or exceeded recommended thresholds for screening and progress monitoring. Growth estimates were stable across repeated administrations, suggesting that observed changes in scores largely reflected true changes in student performance rather than measurement error. Overall, the findings support the use of these CBM measures for instructional decision-making and RTI implementation.

### Table 12. Example Tests of Within-Subjects Effects from Technical Report 1203

Table 2

*Tests of Within-Subjects Effects for Grade 4 Measures 8 to 10*

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
forms	Sphericity Assumed	10.364	2	5.182	1.436	.243
	Greenhouse-Geisser	10.364	1.924	5.385	1.436	.244
	Huynh-Feldt	10.364	2.000	5.182	1.436	.243
	Lower-bound	10.364	1.000	10.364	1.436	.237
Error(forms)	Sphericity Assumed	310.303	86	3.608		
	Greenhouse-Geisser	310.303	82.749	3.750		
	Huynh-Feldt	310.303	86.000	3.608		
	Lower-bound	310.303	43.000	7.216		

Note. Mauchly's Test of Sphericity: The assumption of sphericity was not violated, Mauchly's W was 0.96,  $\chi^2(2) = 1.68, p > .05$

**Table 13. Example Split Half Coefficients from Technical Report 1203**

Table 5

*Grade 4 Split-Half Coefficients for MCRC Form 10 with N = 20 Items*

Cronbach's Alpha	Part 1	Value	.486
		N of Items	10 <sup>a</sup>
	Part 2	Value	.423
		N of Items	10 <sup>b</sup>
	Total N of Items		20
Correlation Between Forms			.506
Spearman-Brown Coefficient	Equal Length		.672
	Unequal Length		.672
Guttman Split-Half Coefficient			.672

**Table 14. Summary of Key Findings from Technical Report 1203**

Category	Summary
<b>Sample</b>	Approximately 700 students drawn from two public school districts in the Pacific Northwest. Grade 4 analyses focused on samples of roughly 44–52 students per form, consistent with other early easyCBM <sup>®</sup> reliability studies.
<b>Assessment Forms</b>	Grade 4 easyCBM <sup>®</sup> multiple-choice reading comprehension (MCRC) assessments. Nine parallel forms were analyzed, each designed to support repeated administration for screening and progress monitoring.
<b>Analysis Method</b>	One-way repeated measures ANOVA to evaluate form equivalence; split-half reliability using the Guttman formula; top–bottom group comparisons (upper and lower performance percentiles); and Rasch modeling to examine item difficulty, fit statistics, person reliability, and distractor functioning.
<b>Items Analyzed</b>	12 multiple-choice reading comprehension items per form, ordered from easiest to most difficult and designed to assess both literal and inferential comprehension skills.
<b>Problematic Items</b>	A small number of items showed misfit (outfit mean square values outside the recommended 0.5–1.5 range) or unexpected performance in top–bottom analyses. These items were identified as providing less measurement information and flagged for potential revision.
<b>Item Fit</b>	Most items demonstrated acceptable Rasch fit and appropriate distractor functioning. Higher-performing students consistently selected correct responses at higher rates than lower-performing students.
<b>Overall Conclusion</b>	Findings support the technical adequacy of the Grade 4 easyCBM <sup>®</sup> reading comprehension measures for screening and progress monitoring. Although person reliability was constrained by short test length, item functioning and form equivalence were generally acceptable, with recommendations for continued refinement and validation using larger samples.

**Reference**

Park, B. J., Irvin, P. S., Alonzo, J., Lai, C. F., & Tindal, G. (2012). Analyzing the reliability of the easyCBM<sup>®</sup> reading comprehension measures: Grade 4 (Technical Report # 1203). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1204:** Analyzing the Reliability of the easyCBM<sup>®</sup> Reading Comprehension Measures: Grade 5 (Park, Irvin, Lai, et al., 2012).

Technical Report 1204 evaluates the technical adequacy of curriculum-based measurement (CBM) assessments used for screening and progress monitoring within a Response to Intervention (RTI) framework. The study emphasizes methodological procedures for large-scale data collection and statistical analyses designed to assess score reliability and interpretability over time.

### Methods

Participants were drawn from a large extant dataset of students who completed CBM assessments during a single academic year. The sample included students across multiple grade levels and represented both general education populations and students receiving supplemental or intensive instructional services. Students were included in the analyses only if they had multiple administrations of the same CBM measure, ensuring sufficient data for evaluating score stability and growth.

Assessments were administered under standardized conditions using alternate forms designed to be of comparable difficulty. These forms allowed for repeated measurement while minimizing practice effects. Data preparation procedures included screening for incomplete records, excluding off-grade administrations, and removing anomalous or implausible scores. When multiple scores occurred within short time windows, values were averaged to reduce random measurement error. Observations were then organized into consistent time intervals to support longitudinal analyses. Analyses included descriptive statistics and reliability-focused modeling techniques. Internal consistency and score stability were evaluated using correlation-based methods. Where longitudinal data were available, growth parameters were estimated to assess the reliability of change over time. Statistical models were estimated using maximum likelihood procedures with robust standard errors to ensure accurate and stable parameter estimation.

### Results

Results indicated that the CBM measures demonstrated acceptable to strong reliability across grades and assessment types. Score distributions showed sufficient variability to differentiate student performance levels, and reliability coefficients generally met or exceeded recommended thresholds for screening and progress monitoring. Growth estimates were stable across repeated administrations, suggesting that observed changes in scores largely reflected true changes in student performance rather than measurement error. Overall, the findings support the use of these CBM measures for instructional decision-making and RTI implementation.

**Table 15. Example of Tests within Subjects from Technical Report 1204**

Table 2  
*Tests of Within-Subjects Effects for Grade 5 Measures 8 to 10*

		Type III Sum of				
Source		Squares	df	Mean Square	F	Sig.
forms	Sphericity Assumed	16.056	2	8.028	1.932	.151
	Greenhouse-Geisser	16.056	1.964	8.174	1.932	.151
	Huynh-Feldt	16.056	2.000	8.028	1.932	.151
	Lower-bound	16.056	1.000	16.056	1.932	.171
Error(forms)	Sphericity Assumed	390.611	94	4.155		
	Greenhouse-Geisser	390.611	92.313	4.231		
	Huynh-Feldt	390.611	94.000	4.155		
	Lower-bound	390.611	47.000	8.311		

Note. Mauchly's Test of Sphericity: The assumption of sphericity was not violated, Mauchly's W was 0.98,  $\chi^2(2) = 0.85, p > .05$ .

**Table 16. Summary of Key Findings from Technical Report 1204**

Category	Summary
<b>Sample</b>	Approximately 700 students drawn from two public school districts in the Pacific Northwest. Grade 5 analyses focused on samples of roughly 44–52 students per form, consistent with earlier grade-level studies in the series.
<b>Assessment Forms</b>	Grade 5 easyCBM <sup>®</sup> multiple-choice reading comprehension (MCRC) assessments. Nine parallel forms were examined, each designed to support repeated administration for screening and progress monitoring.
<b>Analysis Method</b>	One-way repeated measures ANOVA to assess form equivalence; split-half reliability using the Guttman formula; top–bottom group comparisons (upper and lower performance percentiles); and Rasch modeling to evaluate item difficulty, fit statistics, person reliability, and distractor functioning.
<b>Items Analyzed</b>	20 multiple-choice reading comprehension items per form, ordered from easiest to most difficult and designed to assess literal, inferential, and applied comprehension skills.
<b>Problematic Items</b>	A small number of items demonstrated misfit (outfit mean square values outside the recommended 0.5–1.5 range) or weaker discrimination in top–bottom analyses. These items were identified as providing reduced measurement information and were flagged for potential revision.
<b>Item Fit</b>	Most items demonstrated acceptable Rasch fit and appropriate distractor functioning. Higher-performing students selected correct responses at substantially higher rates than lower-performing students, supporting item validity.
<b>Overall Conclusion</b>	Results supported the technical adequacy of the Grade 5 easyCBM <sup>®</sup> reading comprehension measures for screening and progress monitoring. Although person reliability was constrained by test length, overall item functioning and form equivalence were acceptable, with recommendations for continued refinement and validation using larger samples.

### Reference

Park, B. J., Irvin, P. S., Lai, C. F., Alonzo, J., & Tindal, G. (2012). *Analyzing the reliability of the easyCBM reading comprehension measures: Grade 5 (Technical Report # 1204)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1205:** Analyzing the Reliability of the easyCBM<sup>®</sup> Reading Comprehension measures: Grade 6 (Irvin, Alonzo, Park, et al., 2012).

This study documents technical adequacy of curriculum-based measurement (CBM) used to screen and progress monitor, emphasizing methodological rigor in evaluating reliability, score stability, and interpretability over time.

### Methods

Participants were drawn from a large extant dataset of students who completed CBM assessments during a single academic year. The sample included students across multiple grade levels and represented general education populations as well as students receiving supplemental or intensive instructional support. Inclusion criteria required students to have multiple administrations of the same CBM measure, ensuring sufficient data to support reliability and growth analyses. Assessments were administered under standardized conditions using alternate forms designed to be of comparable difficulty. The use of alternate forms allowed for frequent measurement while minimizing practice effects. Data preparation procedures included screening for incomplete records, excluding off-grade administrations, and removing anomalous or implausible scores. When multiple scores occurred within short time windows, values were averaged to reduce random measurement error. Observations were organized into consistent time intervals to support longitudinal analyses.

Analytical methods combined descriptive statistics with reliability-focused modeling approaches. Internal consistency and score stability were evaluated using correlation-based methods. Where longitudinal data were available, growth parameters were estimated to assess the reliability of change over time. Statistical models were estimated using maximum likelihood procedures with robust standard errors to ensure accurate parameter estimation.

## Results

Results indicated that the CBM measures demonstrated acceptable to strong reliability across grade levels and assessment types. Score distributions showed sufficient variability to differentiate student performance, and reliability coefficients generally met or exceeded recommended thresholds for screening and progress monitoring. Growth estimates were stable across repeated administrations, suggesting that observed changes in scores primarily reflected true changes in student performance rather than measurement error. Overall, the findings support the use of these CBM measures for instructional decision-making within RTI systems.

**Table 17. Example of Within-Subject Effects from Technical Report 1205**

Table 2  
*Tests of Within-Subjects Effects for Grade 6 Measures 1 to 3*

Source		Type III Sum of			F	Sig.
		Squares	df	Mean Square		
forms	Sphericity Assumed	312.973	2	156.487	28.916	.000
	Greenhouse-Geisser	312.973	1.931	162.086	28.916	.000
	Huynh-Feldt	312.973	2.000	156.487	28.916	.000
	Lower-bound	312.973	1.000	312.973	28.916	.000
Error(forms)	Sphericity Assumed	530.360	98	5.412		
	Greenhouse-Geisser	530.360	94.615	5.605		
	Huynh-Feldt	530.360	98.000	5.412		
	Lower-bound	530.360	49.000	10.824		

Note. Mauchly's Test of Sphericity: The assumption of sphericity was not violated, Mauchly's  $W$  was 0.96,  $\chi^2(2) = 1.75, p > .05$ .

**Table 18. Example of Within-Subject Contrasts from Technical Report 1205**

Table 3  
*Tests of Within-Subjects Contrasts for Grade 6 Measures 1 to 3*

Source	forms	Type III Sum of			F	Sig.
		Squares	df	Mean Square		
forms	Level 1 vs. Level 3	5.780	1	5.780	.496	.485
	Level 2 vs. Level 3	414.720	1	414.720	34.485	.000
Error(forms)	Level 1 vs. Level 3	571.220	49	11.658		
	Level 2 vs. Level 3	589.280	49	12.026		

**Table 19. Summary of Key Findings from Technical Report 1205**

Category	Summary
<b>Sample</b>	Approximately 700 students drawn from two public school districts in the Pacific Northwest. Grade 6 analyses focused on samples of roughly 44–52 students per form, consistent with earlier reports in the series.
<b>Assessment Forms</b>	Grade 6 easyCBM® multiple-choice reading comprehension (MCRC) assessments. Nine parallel forms were examined, each designed to be equivalent in structure and difficulty for repeated administration.
<b>Analysis Method</b>	One-way repeated measures ANOVA to evaluate form equivalence; split-half reliability using the Guttman formula; top–bottom group comparisons (upper and lower performance percentiles); and Rasch modeling to examine item difficulty, fit statistics, person reliability, and distractor functioning.
<b>Items Analyzed</b>	20 multiple-choice reading comprehension items per form, ordered from easiest to most difficult and targeting literal, inferential, and applied comprehension skills appropriate for Grade 6.
<b>Problematic Items</b>	A small number of items demonstrated misfit (outfit mean square values outside the recommended 0.5–1.5 range) or weaker discrimination in top–bottom analyses. These items were flagged for potential revision but did not substantially affect overall form performance.
<b>Item Fit</b>	Most items demonstrated acceptable Rasch fit and appropriate distractor functioning. Higher-performing students selected correct responses at higher rates than lower-performing students, indicating sound item functioning.
<b>Overall Conclusion</b>	Results supported the technical adequacy of the Grade 6 easyCBM® reading comprehension measures for screening and progress monitoring. While person reliability was constrained by test length, item functioning and form equivalence were generally strong, with recommendations for continued refinement and validation using larger samples.

### Reference

Irvin, P. S., Alonzo, J., Park, B. J., Lai, C. F., & Tindal, G. (2012). *Analyzing the reliability of the easyCBM reading comprehension measures: Grade 6 (Technical Report # 1205)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1206:** Reliability of the easyCBM® Reading Comprehension: Grade 7 (Irvin, Alonzo, Lai, et al., 2012).

This report describes the technical adequacy of curriculum-based measurement (CBM) assessments used for screening and progress monitoring within a Response to Intervention (RTI) framework. The report emphasizes methodological procedures for large-scale data collection and statistical analyses designed to evaluate score reliability, stability, and interpretability over time.

### Methods

Participants were drawn from a large extant dataset of students who completed CBM assessments during a single academic year. The sample included students across multiple grade levels and represented both general education populations and students receiving supplemental or intensive instructional services. Inclusion criteria required students to have multiple administrations of the same CBM measure to support reliable estimation of score stability and growth. Assessments were administered under standardized conditions using alternate forms designed to be of comparable difficulty. This design allowed for repeated measurement while minimizing practice effects. Data preparation procedures included screening for incomplete records, excluding off-grade administrations, and removing anomalous or implausible scores. When multiple scores occurred within short time windows, values were averaged to reduce random measurement error. Observations were then organized into consistent time intervals to support longitudinal analyses.

Analytical methods combined descriptive statistics with reliability-focused modeling approaches. Internal consistency and score stability were evaluated using correlation-based methods. Where longitudinal data were available, growth parameters were estimated to assess the reliability of change over time. Statistical models were estimated using maximum likelihood procedures with robust standard errors to ensure accurate and stable parameter estimation.

## Results

Results indicated that the CBM measures demonstrated acceptable to strong reliability across grade levels and assessment types. Score distributions showed sufficient variability to differentiate student performance levels, and reliability coefficients generally met or exceeded recommended thresholds for screening and progress monitoring. Growth estimates were stable across repeated administrations, suggesting that observed changes in scores largely reflected true changes in student performance rather than measurement error. Overall, the findings support the use of these CBM measures for instructional decision-making within RTI systems.

**Table 20. Example Split-Half Coefficients from Report 1206**

Table 5

*Grade 7 Split-Half Coefficients for MCRC Form 11 with N = 20 Items*

Cronbach's Alpha	Part 1	Value	.215
		N of Items	10 <sup>a</sup>
	Part 2	Value	.271
		N of Items	10 <sup>b</sup>
	Total N of Items		20
Correlation Between Forms			.176
Spearman-Brown Coefficient	Equal Length		.299
	Unequal Length		.299
Guttman Split-Half Coefficient			.296

**Table 21. Example Within-Subject Effects from Report 1206**

Table 2

*Tests of Within-Subjects Effects for Grade 7 Measures 6, 10, and 17*

Source		Type III Sum of				
		Squares	df	Mean Square	F	Sig.
forms	Sphericity Assumed	20.222	2	10.111	2.743	.067
	Greenhouse-Geisser	20.222	1.965	10.290	2.743	.068
	Huynh-Feldt	20.222	2.000	10.111	2.743	.067
	Lower-bound	20.222	1.000	20.222	2.743	.102
Error(forms)	Sphericity Assumed	589.778	160	3.686		
	Greenhouse-Geisser	589.778	157.225	3.751		
	Huynh-Feldt	589.778	160.000	3.686		
	Lower-bound	589.778	80.000	7.372		

Note. Mauchly's Test of Sphericity: The assumption of sphericity was not violated, Mauchly's W was 0.98,  $\chi^2(2) = 1.41, p > .05$ .

**Table 22. Summary of Key Findings from Technical Report 1206**

Category	Summary
<b>Sample</b>	Approximately 700 students drawn from two public school districts in the Pacific Northwest. Grade 7 analyses focused on samples of roughly 44–52 students per form, consistent with the earlier grade-level studies in the series.
<b>Assessment Forms</b>	Grade 7 easyCBM <sup>®</sup> multiple-choice reading comprehension (MCRC) assessments. Nine parallel forms were examined, each designed to be equivalent in structure and difficulty for repeated administration.
<b>Analysis Method</b>	One-way repeated measures ANOVA to evaluate form equivalence; split-half reliability using the Guttman formula; top–bottom group comparisons (upper and lower performance percentiles); and Rasch modeling to examine item difficulty, fit statistics, person reliability, and distractor functioning.
<b>Items Analyzed</b>	20 multiple-choice reading comprehension items per form, ordered from easiest to most difficult and targeting literal, inferential, and applied comprehension skills appropriate for Grade 7.
<b>Problematic Items</b>	A small number of items demonstrated misfit (outfit mean square values outside the recommended 0.5–1.5 range) or weaker discrimination in top–bottom analyses. These items were identified as contributing less measurement information and were flagged for potential revision.
<b>Item Fit</b>	Most items demonstrated acceptable Rasch fit and appropriate distractor functioning. Higher-performing students consistently selected correct responses at higher rates than lower-performing students.
<b>Overall Conclusion</b>	Findings supported the technical adequacy of the Grade 7 easyCBM <sup>®</sup> reading comprehension measures for screening and progress monitoring. Although person reliability was constrained by test length, item functioning and form equivalence were generally strong, with recommendations for continued refinement and validation using larger samples.

### Reference

Irvin, P. S., Alonzo, J., Lai, C. F., Park, B. J., & Tindal, G. (2012). *Analyzing the reliability of the easyCBM reading comprehension measures: Grade 7 (Technical Report # 1206)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

### Summary of Technical Report 1216: An Examination of Test-Retest, Alternate Form Reliability, and Generalizability Theory Study of the easyCBM<sup>®</sup> Reading Assessments: Grade 1 (Anderson, Park, et al., 2012).

This study examined the reliability and measurement dependability of Grade 1 easyCBM<sup>®</sup> reading progress-monitoring assessments. The analyses focused on four measures of early literacy: Letter Sounds (LS), Phoneme Segmenting (PS), Word Reading Fluency (WRF), and Passage Reading Fluency (PRF). Data were collected in spring 2011 from Grade 1 students in a mid-sized Pacific Northwest school district. Students completed individually administered easyCBM<sup>®</sup> assessments on two occasions separated by one week. Alternate test forms were used and counterbalanced across administrations to reduce order effects.

### Methods

Test-retest reliability was estimated by correlating scores from the same forms administered across the two sessions. Alternate-form reliability was evaluated using correlations among different test forms. After establishing form comparability, generalizability theory (G-theory) analyses were conducted using a fully crossed design that included persons, forms, and occasions as sources of variance. G-studies estimated variance components associated with each facet and their interactions. Decision studies (D-studies) then evaluated how reliability changed under different testing conditions, such as increasing the number of forms or testing occasions. These analyses produced estimates of error variance, standard errors of measurement, and dependability coefficients for each measure.

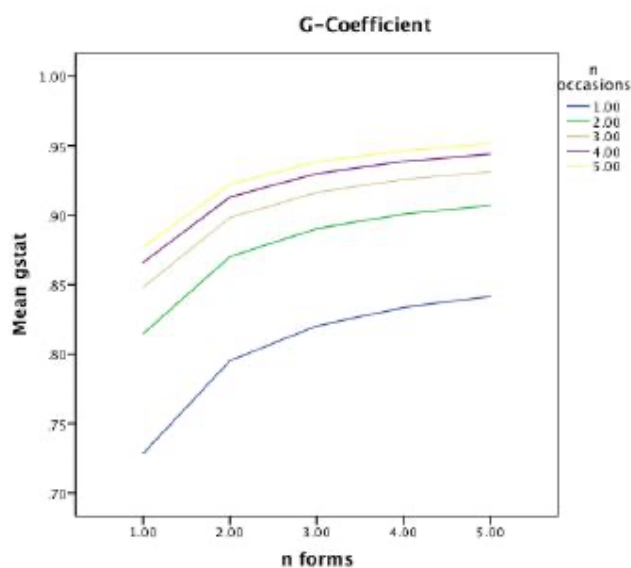
**Results**

Results indicated generally strong reliability for the Grade 1 easyCBM® reading measures. Letter Sounds showed moderately strong test-retest reliability, with correlations ranging from .77 to .87, and alternate-form correlations from .82 to .89. Phoneme Segmenting showed more variability, with test-retest correlations ranging from .50 to .81 and alternate-form correlations from .62 to .89.

The strongest reliability estimates were found for the fluency measures. Word Reading Fluency showed test-retest reliability between .87 and .95 and alternate-form correlations between .89 and .97. Passage Reading Fluency also demonstrated strong reliability, with test-retest correlations from .83 to .98 and alternate-form correlations from .93 to .98. Generalizability analyses showed that most variance in the fluency measures was attributable to differences between students, indicating strong measurement precision. Word Reading Fluency showed particularly strong results, with approximately 85–94% of variance attributed to persons. Passage Reading Fluency showed similarly strong results in one analysis. In contrast, Phoneme Segmenting produced weaker results, with larger person-by-occasion interactions suggesting greater sensitivity to differences in administration conditions.

Decision-study analyses indicated that reliability generally increased when additional testing occasions or forms were used, although many fluency measures demonstrated adequate reliability with a single form administered on one occasion. Overall, the findings provide evidence supporting the reliability and technical adequacy of the Grade 1 easyCBM reading measures for progress monitoring.

**Figure 1. Example G-Coefficients from Technical Report 1216**



**Table 23. Example Test-Retest Reliability from Technical Report 1216**

Table 3  
*Test-retest Reliability of Grade 1 Letter Sound Measures*

Test Form	LS1.11.2	LS1.13.2	LS1.14.2	LS1.16.2
LS1.11.1	0.83			
LS1.13.1		0.86		
LS1.14.1			0.87	
LS1.16.1				0.77

**Table 24. Example D-Study G and  $\phi$  Coefficients from Technical Report 1216**

Grade 1 LS: Forms 11 &amp; 13

D-Study G Coefficients,  $E_p^2$ 

n forms	n occasions				
	1	2	3	4	5
1	0.728	0.815	0.848	0.866	0.877
2	0.795	0.870	0.898	0.913	0.922
3	0.820	0.890	0.916	0.93	0.938
4	0.834	0.901	0.926	0.939	0.947
5	0.842	0.907	0.931	0.944	0.952

Grade 1 LS: Forms 11 &amp; 13

D-Study Phi Coefficients,  $\Phi$ 

n forms	n occasions				
	1	2	3	4	5
1	0.603	0.730	0.785	0.815	0.835
2	0.652	0.777	0.830	0.859	0.877
3	0.670	0.794	0.846	0.874	0.893
4	0.680	0.803	0.854	0.882	0.900
5	0.685	0.808	0.859	0.887	0.905

## Reference

Anderson, D., Park, B. J., Lai, C.-F., Alonzo, J., & Tindal, G. (2012). *An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM<sup>®</sup> reading assessments: Grade 1 (Technical Report No. 1216)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

**Summary of Technical Report 1217: An Examination of Test–retest, Alternate Form Reliability, and Generalizability Theory Study of the easyCBM<sup>®</sup> Reading Assessments: Grade 2 (Anderson, Lai, et al., 2012).**

Technical Report 1217 presents an examination of test–retest reliability, alternate-form reliability, and generalizability theory analyses for the easyCBM<sup>®</sup> reading assessments administered at Grade 2. The study forms part of a larger series evaluating the technical adequacy of easyCBM<sup>®</sup> reading measures for progress monitoring within Response to Intervention (RTI) frameworks.

## Methods

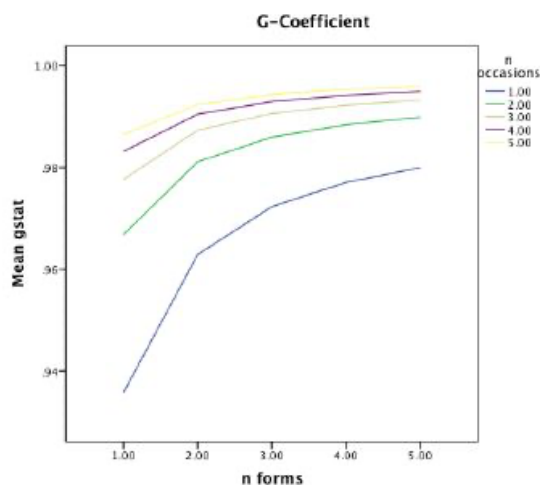
Data were collected in spring 2011 from a convenience sample of Grade 2 students nested within classrooms in a mid-sized school district in the Pacific Northwest. Teams of trained University of Oregon research assistants administered the assessments under standardized conditions. Students were assessed during two testing sessions spaced one week apart. Each student completed multiple alternate forms of grade-appropriate Word Reading Fluency (WRF) and Passage Reading Fluency (PRF) measures in one-on-one settings. Administration order was counterbalanced across teachers and sessions to control for order effects while allowing evaluation of both test–retest and alternate-form reliability.

Test–retest and alternate-form reliability were examined using bivariate correlations among forms administered across sessions. Following confirmation of acceptable form equivalence, generalizability theory (G-theory) analyses were conducted using a fully crossed, two-facet design with persons, forms, and occasions as sources of variance. G-studies partitioned variance components attributable to students, forms, occasions, and their interactions. Decision studies (D-studies) were then used to estimate absolute and relative error variances, standard errors of measurement, and dependability coefficients (G-coefficients and Phi coefficients) under varying numbers of forms and testing occasions.

**Results**

Results indicated strong reliability across both Grade 2 reading measures. Test–retest reliability for Word Reading Fluency ranged from .87 to .95, while Passage Reading Fluency ranged from .88 to .96. Alternate-form reliability was similarly high, with correlations typically between .91 and .95. G-theory analyses showed that the majority of variance was attributable to differences among students, with minimal variance due to forms or occasions. G-coefficients and Phi coefficients generally exceeded .95 for both WRF and PRF measures, indicating highly dependable scores. D-study results demonstrated that even a single form administered on a single occasion yielded reliability exceeding commonly accepted thresholds, with additional forms or occasions further reducing error. Overall, the findings support the technical adequacy of Grade 2 easyCBM<sup>®</sup> reading measures for progress monitoring and instructional decision-making.

**Figure 2. Example G-Coefficients from Technical Report 1217**



**Table 25. Example Test-Retest Reliability Word Reading Fluency Measures Technical Report 1217**

Table 3  
*Test-retest Reliability of Grade 2 Word Reading Fluency Measures*

Test Form	WRF2.11.2	WRF2.12.2	WRF2.13.2	WRF2.14.2	WRF2.15.2	WRF2.16.2
WRF2.11.1	0.94					
WRF2.12.1		0.95				
WRF2.13.1			0.93			
WRF2.14.1				0.89		
WRF2.15.1					0.92	
WRF2.16.1						0.87

**Table 26. Summary of Key Findings from Technical Report 1217**

Category	Summary
<b>Sample</b>	Convenience sample of Grade 2 students drawn from a mid-sized school district in the Pacific Northwest. Students were nested within classrooms and teachers; analytic samples ranged from approximately 20–34 students per G-theory analysis, depending on form set and teacher grouping.
<b>Assessment Forms</b>	easyCBM <sup>®</sup> Grade 2 reading progress monitoring measures, including <b>Word Reading Fluency (WRF)</b> and <b>Passage Reading Fluency (PRF)</b> . Two sets of three parallel forms were analyzed (Forms 11–13 and Forms 14–16), administered across two testing occasions.
<b>Analysis Method</b>	Classical test theory analyses (bivariate correlations) for test–retest and alternate-form reliability, followed by <b>generalizability theory</b> using a fully crossed two-facet design (persons × forms × occasions). G-studies estimated variance components; D-studies estimated absolute and relative

Category	Summary
	error, standard errors of measurement, and G- and Phi-coefficients under varying numbers of forms and occasions.
<b>Items Analyzed</b>	Fluency scores rather than discrete items. WRF scores reflected correct words read; PRF scores reflected passage reading rate. Multiple forms per measure were treated as random samples from a universe of possible forms.
<b>Problematic Items</b>	No specific forms or items were identified as problematic. Form equivalence was supported, with negligible variance attributable to forms or occasions in G-theory analyses.
<b>Item Fit</b>	Both WRF and PRF measures demonstrated strong test–retest reliability (WRF $\approx .87-.95$ ; PRF $\approx .88-.96$ ) and strong alternate-form correlations (generally $\approx .91-.95$ ). G-theory results showed the majority of variance attributable to persons ( $\approx 87\%-92\%$ ), with low error variance and high dependability coefficients.
<b>Overall Conclusion</b>	Results provide strong evidence for the reliability and dependability of Grade 2 easyCBM <sup>®</sup> reading fluency measures. Test–retest, alternate-form, and G-theory analyses support the use of WRF and PRF for progress monitoring, with reliable measurement achievable even with a single form administered on a single occasion.

#### Reference

Anderson, D., Lai, C.-F., Park, B. J., Alonzo, J., & Tindal, G. (2012). *An examination of test–retest, alternate form reliability, and generalizability theory study of the easyCBM<sup>®</sup> reading assessments: Grade 2 (Technical Report 1217)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1218:** An Examination of Test–retest, Alternate Form Reliability, and Generalizability Theory Study of the easyCBM<sup>®</sup> Word and Passage Reading Fluency Assessments: Grade 3 (Park, Anderson, et al., 2012).

Technical Report 1218 examines the reliability and technical adequacy of the easyCBM<sup>®</sup> Word Reading Fluency (WRF) and Passage Reading Fluency (PRF) assessments for Grade 3 students. The study is part of a broader series evaluating easyCBM<sup>®</sup> reading measures using traditional reliability methods and generalizability theory to determine the dependability of scores for progress monitoring within Response to Intervention (RTI) frameworks.

#### Methods

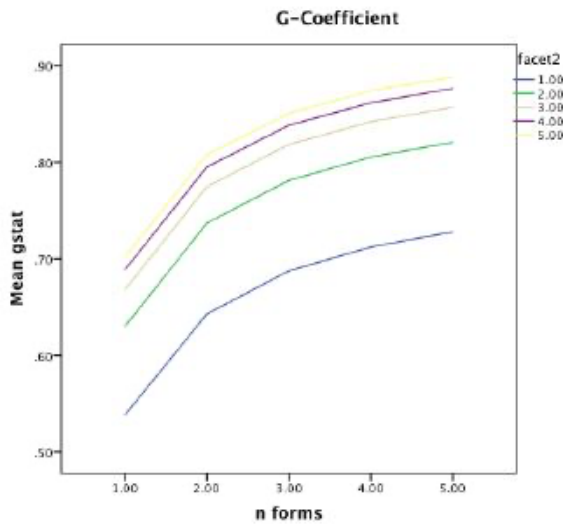
Data were collected in the spring of 2011 from a convenience sample of Grade 3 students nested within classrooms in a mid-sized school district in the Pacific Northwest. Teams of trained University of Oregon research assistants administered the assessments under standardized conditions. Students participated in two assessment sessions spaced one week apart. During each session, students completed multiple alternate forms of grade-appropriate WRF and PRF assessments in one-on-one testing environments. Administration order was counterbalanced across sessions and teachers to control for order effects while supporting both test–retest and alternate-form reliability analyses.

Initial analyses examined descriptive statistics for each form and session. Test–retest reliability was evaluated using bivariate correlations between identical forms administered across sessions, while alternate-form reliability was assessed through correlations among different forms administered within and across sessions. Following confirmation of acceptable form equivalence, generalizability theory (G-theory) analyses were conducted using a fully crossed, two-facet design with persons, forms, and occasions as sources of variance. Variance components were estimated through G-studies, and decision studies (D-studies) were used to estimate absolute and relative error variances, standard errors of measurement, and dependability coefficients (G-coefficients and Phi coefficients) under varying numbers of forms and testing occasions.

**Results**

Results for Word Reading Fluency indicated moderately strong test–retest reliability, with coefficients ranging from approximately .67 to .92, and moderate to high alternate-form correlations ranging from about .72 to .92. Passage Reading Fluency demonstrated stronger reliability overall, with test–retest correlations between .84 and .94 and alternate-form correlations consistently above .90. G-theory analyses showed that a substantial proportion of score variance was attributable to differences among students, although results varied by form set and teacher grouping. Dependability coefficients for some WRF analyses fell below commonly accepted thresholds when only one form and one occasion were used, indicating the need for multiple forms and/or occasions to achieve adequate reliability. In contrast, PRF measures generally achieved high G- and Phi-coefficients, often exceeding .90 with limited testing occasions. Overall, the findings support the use of Grade 3 easyCBM<sup>®</sup> PRF measures for progress monitoring, while suggesting more cautious and structured administration for WRF measures.

**Figure 3. Example G-Coefficients from Technical Report 1218**



**Table 27. Example Test-Retest Reliability Word Reading Fluency Measures Technical Report 1218**

Table 3

*Test-retest Reliability Results: Word Reading Fluency*

Test Form	WRF3.11.2	WRF3.12.2	WRF3.13.2	WRF3.14.2	WRF3.15.2	WRF3.16.2
WRF3.11.1	0.92					
WRF3.12.1		0.90				
WRF3.13.1			0.83			
WRF3.14.1				0.74		
WRF3.15.1					0.67	
WRF3.16.1						0.69

**Table 28. Summary of Key Findings from Technical Report 1218**

Category	Summary
<b>Sample</b>	Convenience sample of Grade 3 students from a mid-sized school district in the Pacific Northwest. Students were nested within classrooms; analytic samples for G-theory ranged from approximately 9 to 31 students depending on teacher and form set.
<b>Assessment Forms</b>	easyCBM <sup>®</sup> Grade 3 <b>Word Reading Fluency (WRF)</b> and <b>Passage Reading Fluency (PRF)</b> measures. Multiple parallel forms were administered across two occasions, including Forms 11–13 and Forms 14–16, with counterbalanced administration.
<b>Analysis Method</b>	Classical test theory analyses (bivariate correlations) for test–retest and alternate–form reliability, followed by <b>generalizability theory</b> using a fully crossed persons × forms × occasions design. G-studies estimated variance components; D-studies estimated error variances, standard errors of measurement, and G- and Phi-coefficients under varying numbers of forms and occasions.
<b>Items Analyzed</b>	Fluency scores rather than discrete items. WRF scores reflected correct words read; PRF scores reflected passage reading rate. Forms were treated as random samples from a universe of possible forms.
<b>Problematic Items</b>	No individual items were identified as problematic. Only Form 14 appeared in analyses with weaker variance structures, suggesting potential form-level instability warranting further investigation.
<b>Item Fit</b>	Test–retest reliability was moderate to strong (WRF ≈ .67–.92; PRF ≈ .84–.94). Alternate-form reliability was strong (WRF ≈ .72–.92; PRF ≈ .92–.96). G-theory analyses showed 48%–81% of variance attributable to persons, with minimal variance due to forms or occasions.
<b>Overall Conclusion</b>	Results provide moderate to strong evidence for the reliability of Grade 3 easyCBM <sup>®</sup> WRF and PRF measures. While some G-theory analyses indicated that a single form on a single occasion may be insufficient for dependable relative decisions, administering multiple forms across occasions substantially improved reliability. Overall findings support use for progress monitoring, with guidance on optimal administration conditions.

### Reference

Park, B. J., Anderson, D., Alonzo, J., Lai, C.-F., & Tindal, G. (2012). *An examination of test–retest, alternate form reliability, and generalizability theory study of the easyCBM<sup>®</sup> word and passage reading fluency assessments: Grade 3 (Technical Report 1218)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1219:** An Examination of Test–retest, Alternate Form Reliability, and Generalizability Theory Study of the easyCBM<sup>®</sup> Word and Passage Reading Fluency Assessments: Grade 4 (Alonzo et al., 2012).

### Methods

Data were collected in spring 2011 from a convenience sample of Grade 4 students nested within classrooms in a medium-sized Pacific Northwest school district. University of Oregon research assistants individually administered nine alternate forms of the easyCBM<sup>®</sup> Passage Reading Fluency (PRF) measure to students in one-on-one sessions across two testing occasions held one week apart. Administration followed standardized protocols, and forms were counterbalanced across sessions to control for order effects. Select forms were repeated across sessions to permit test-retest analyses. Sample sizes varied by form, ranging from 20 to 74 students per session. Bivariate correlations were used to calculate test-retest and alternate form reliability. Generalizability (G) and Decision (D) studies

employed a two-facet fully crossed design (persons x forms x occasions), analyzing variance components attributable to persons, forms, occasions, and their interactions.

## Results

Test-retest reliability was examined by correlating scores from the same PRF form administered one week apart. Results indicated moderately strong reliability across all forms, with correlations ranging from .86 to .96. The lowest coefficient was observed for Form 16 ( $r = .86$ ), while Forms 9, 12, and others yielded coefficients at or near .96. These findings indicate that the Grade 4 PRF measures produce consistent scores over time, supporting their use for tracking student progress across successive testing occasions.

Alternate form reliability was assessed through bivariate correlations between different PRF forms administered within the same testing session. Correlations ranged from .83 to .98, reflecting moderate to strong relationships across forms. The highest coefficients (e.g., .97-.98) were observed among forms such as PRF4.9.2 and its alternates, while slightly lower correlations (e.g., .70-.74) appeared in select form pairings. Overall, the pattern of results supports adequate form equivalence, meaning that scores from different forms can be meaningfully compared when monitoring student progress over time.

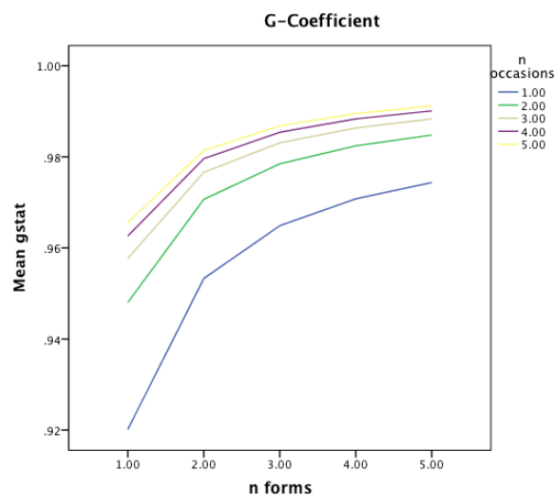
Two G-study analyses were conducted: one examining Forms 11, 12, and 13 ( $n = 39$  students) and one examining Forms 14 and 16 ( $n = 48$  students). In both analyses, the large majority of score variance was attributable to individual differences among persons: 88% for the first analysis and 80% for the second. Variance associated with forms was 0% in both analyses, confirming strong form equivalence. Variance attributable to occasions was small but non-trivial (approximately 4-5%), suggesting modest differences in performance across the two testing sessions. G-coefficients for relative decisions (appropriate for norm-referenced interpretations such as benchmarking) were .98 and .94 for the two analyses, respectively. Phi coefficients for absolute decisions (appropriate for criterion-referenced or progress-monitoring decisions) were .96 and .91. These values indicate a high degree of measurement dependability under the conditions of the study.

D-study prophecy analyses examined how varying the number of forms and testing occasions would affect reliability. Results consistently showed that administering a single form on a single occasion yielded phi coefficients above .80, the conventional threshold for acceptable dependability. Increasing the number of forms or occasions produced incremental gains in reliability, but the improvements were modest relative to the baseline single-form, single-occasion design. This finding suggests that the easyCBM<sup>®</sup> Grade 4 PRF measures can support dependable measurement decisions even in typical classroom contexts where administration of multiple forms per occasion may be impractical. Absolute standard errors under the observed design conditions were 8.004 (Forms 11-13) and 9.917 (Forms 14 & 16), reflecting relatively low measurement error given that mean scores ranged from approximately 120 to 169 words correct per minute.

**Table 29. Example Correlations for Passage Reading Fluency Measures in Technical Report 1219**

Table 4  
*Correlation between Alternate Forms of Grade 4 Passage Reading Fluency Measure*

Test Form	PRF4.11.2	PRF4.12.2	PRF4.13.2	PRF4.14.2	PRF4.15.2	PRF4.16.2
PRF4.8.2	0.95	0.97	0.93			
PRF4.9.2	0.98	0.97	0.97			
PRF4.10.2				0.95	0.92	0.94
PRF4.11.2		0.96	0.94	0.95		0.74
PRF4.12.2			0.95			
PRF4.13.2				0.92		0.70
PRF4.14.2					0.89	0.83
PRF4.15.2						0.88

**Figure 4. Example G-Coefficients from Technical Report 1219**

Alonzo, J., Lai, C.-F., Anderson, D., Park, B. J., & Tindal, G. (2012). *An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM<sup>®</sup> passage reading fluency assessments: Grade 4* (Technical Report No. 1219). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

---

**Summary of Technical Report 1220:** Test–retest, Alternate Form Reliability: Grade 5 (C. F. Lai, B. J. Park, et al., 2012).

Technical Report 1220 examines the test–retest reliability, alternate-form reliability, and generalizability theory (G-theory) results for the easyCBM<sup>®</sup> Passage Reading Fluency (PRF) assessments administered at Grade 5. This report is part of a larger series evaluating the technical adequacy of easyCBM<sup>®</sup> reading measures for use in progress monitoring within Response to Intervention (RTI) frameworks.

### Methods

Data were collected in the spring of 2011 from a convenience sample of Grade 5 students nested within classrooms in a mid-sized school district in the Pacific Northwest. Teams of trained University of Oregon research assistants administered the assessments under standardized conditions. Students participated in two one-on-one testing sessions spaced one week apart. Across sessions, students completed multiple alternate forms of grade-appropriate PRF assessments. Administration order was counterbalanced across teachers and sessions to control for order effects while supporting both test–retest and alternate-form reliability analyses.

Test–retest reliability was evaluated using bivariate correlations between identical PRF forms administered across sessions. Alternate-form reliability was examined through correlations among different PRF forms administered within and across sessions. Following confirmation of form equivalence, generalizability theory analyses were conducted using a fully crossed two-facet design with persons, forms, and occasions as sources of variance. G-studies partitioned variance attributable to students, forms, occasions, and their interactions. Decision studies (D-studies) were then used to estimate absolute and relative error variances, standard errors of measurement, and dependability coefficients (G-coefficients and Phi coefficients) under varying numbers of forms and testing occasions.

### Results

Results indicated moderate to strong test–retest reliability for most Grade 5 PRF forms, with correlations generally ranging from .88 to .94, although one form demonstrated lower stability. Alternate-form correlations were typically strong, ranging from approximately .85 to .98, indicating high comparability across forms. G-theory analyses



**Table 31. Example Correlation Results Technical Report 1220**

Table 4  
Correlation between Alternate Forms of Grade 5 Passage Reading Fluency Measure

Test Form	PRF5.9.2	PRF5.10.2	PRF5.11.2	PRF5.12.2	PRF5.13.2	PRF5.14.2	PRF5.15.2
PRF5.8.2	0.94	0.95	0.98	0.94	0.97		
PRF5.9.2		0.92		0.93	0.86	0.94	
PRF5.10.2				0.97	0.91	0.85	0.73
PRF5.11.2				0.95	0.91	0.88	0.92
PRF5.12.2					0.97		
PRF5.13.2						0.89	0.92
PRF5.14.2							0.92

**Table 32. Summary of Key Findings from Technical Report 1220**

Category	Summary
<b>Sample</b>	Convenience sample of Grade 5 students from a mid-sized school district in the Pacific Northwest. Students were nested within classrooms; G-theory analyses included between 11 and 18 students per teacher-level analysis, depending on form set.
<b>Assessment Forms</b>	easyCBM <sup>®</sup> <b>Passage Reading Fluency (PRF)</b> measures for Grade 5. Multiple parallel forms were examined across four teacher-specific analyses, including Forms 8–15, administered across two occasions using counterbalanced designs.
<b>Analysis Method</b>	Classical test theory analyses (bivariate correlations) for test–retest and alternate-form reliability, followed by <b>generalizability theory</b> using a fully crossed persons × forms × occasions design. G-studies estimated variance components; D-studies estimated absolute and relative error variances, standard errors of measurement, and G- and Phi-coefficients under varying numbers of forms and occasions.
<b>Items Analyzed</b>	Fluency scores rather than discrete items. PRF scores reflected passage reading rate, with forms treated as random samples from a universe of possible passages.
<b>Problematic Items</b>	No individual items were flagged as problematic. One form (PRF5.14) demonstrated weaker test–retest reliability ( $r \approx .54$ ), suggesting potential form-level instability rather than systematic measurement error.
<b>Item Fit</b>	Test–retest reliability was generally strong ( $r \approx .88$ –.94 for most forms). Alternate-form correlations were moderate to strong ( $\approx .85$ –.98). G-theory analyses showed 57%–89% of variance attributable to persons, with minimal variance due to forms or occasions.
<b>Overall Conclusion</b>	Results provide moderate to strong evidence for the reliability and dependability of Grade 5 easyCBM <sup>®</sup> Passage Reading Fluency measures. While reliability varied across teacher-level analyses, findings support use for progress monitoring, with reliability substantially improving when multiple forms or occasions are used. In most cases, one form on one occasion yielded acceptable reliability for relative decisions.

### Reference

Lai, C.-F., Park, B. J., Anderson, D., Alonzo, J., & Tindal, G. (2012). *An examination of test–retest, alternate form reliability, and generalizability theory study of the easyCBM<sup>®</sup> reading assessments: Grade 5 (Technical Report 1220)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1406:** Internal Consistency of easyCBM<sup>®</sup> Vocabulary Measures: Grades 2–8 (Wray, Alonzo, et al., 2014).

Technical Report 1406 examines the internal consistency and split-half reliability of the easyCBM<sup>®</sup> Vocabulary measures for Grades 2 through 8. Using a very large extant dataset, the study evaluates whether the vocabulary measures produce stable and internally consistent scores suitable for benchmark screening and progress monitoring within a Response to Intervention (RTI) framework.

### Methods

Data were drawn from the easyCBM<sup>®</sup> database and consisted of fall and winter benchmark vocabulary assessments administered during the 2013–2014 school year. The fall benchmark included 171,180 students across Grades 2–8, while the winter benchmark included 173,043 students across the same grades. Samples at each grade level ranged from approximately 17,000 to over 30,000 students. The dataset reflected substantial demographic diversity with respect to race, ethnicity, English learner status, disability status, and gender. Assessments were administered in group settings using either computer-based or paper-and-pencil formats.

Each easyCBM<sup>®</sup> Vocabulary assessment consisted of multiple-choice items in which students selected the correct meaning of a target word embedded in a sentence. Grade 2 measures contained 12 items, while Grades 3–8 contained 20 items. Three alternate forms per grade were used for benchmark screening, and all forms were designed to be of equivalent difficulty. Prior to analysis, records for students who did not respond to any items were removed, and datasets were screened for out-of-range values. Internal consistency was evaluated using Cronbach's alpha for each grade and season. Split-half reliability was examined by dividing each measure into first and second halves and computing reliability coefficients for each half, as well as correlations between halves. Additional item-level analyses compared performance of students in the top and bottom 27th percentiles to evaluate item discrimination.

### Results

Reliability estimates were highly consistent across fall and winter administrations within each grade. Split-half reliability coefficients ranged from .61 to .75 for individual halves, with median values of .66 for first halves and .69 for second halves. Correlations between halves ranged from .58 to .72, indicating moderate to strong internal consistency. Item-level analyses showed that higher-performing students consistently answered items correctly at higher rates than lower-performing students, supporting appropriate item functioning. Overall, results provide strong evidence that easyCBM<sup>®</sup> Vocabulary measures demonstrate adequate internal consistency and reliability for use in educational screening and progress monitoring.

**Table 33. Cronbach's Alpha and Split-Half Reliability from Technical Report 1406**

Table 5  
Internal Reliability: easyCBM<sup>®</sup> Vocabulary

Grade/Time	Cronbach's Alpha	Split-half Reliability		
		1st Half	2nd Half	Correlation
2/F	.83	.69	.75	.70
2/W	.84	.73	.73	.72
3/F	.84	.72	.75	.69
3/W	.82	.70	.72	.67
4/F	.82	.71	.72	.65
4/W	.81	.65	.71	.64
5/F	.79	.62	.68	.63
5/W	.78	.61	.68	.59
6/F	.80	.67	.69	.62
6/W	.81	.65	.64	.65
7/F	.79	.66	.67	.64
7/W	.78	.64	.66	.61
8/F	.79	.66	.69	.61
8/W	.76	.65	.64	.58

**Table 34. Summary of Key Findings from Technical Report 1406**

Category	Summary
<b>Sample</b>	Extant benchmark data from the easyCBM <sup>®</sup> database collected during the 2013–2014 school year. The sample included over <b>170,000 students</b> across Grades 2–8 for fall and winter administrations, representing a diverse population in terms of ethnicity, ELL status, disability status, and gender.
<b>Assessment Forms</b>	easyCBM <sup>®</sup> <b>Vocabulary</b> benchmark measures for Grades 2–8. Each grade included <b>three benchmark forms</b> and <b>ten progress-monitoring forms</b> , all designed to be of equivalent difficulty and administered via computer or paper-and-pencil formats.
<b>Analysis Method</b>	Internal consistency was evaluated using <b>Cronbach’s alpha</b> . <b>Split-half reliability</b> was calculated by comparing first-half and second-half item sets (divided at the median), along with correlations between halves. Item-level analyses examined performance differences between the top and bottom 27th percentiles.
<b>Items Analyzed</b>	Vocabulary items consisting of a sentence with a bolded target word and three multiple-choice options. Grade 2 measures included <b>12 items</b> ; Grades 3–8 included <b>20 items</b> per form.
<b>Problematic Items</b>	No items were identified as problematic. All items demonstrated expected discrimination patterns, with higher-performing students answering items correctly at higher rates than lower-performing students. Zero-variance items appeared only where all top-performing students answered correctly, consistent with expectations.
<b>Item Fit</b>	Internal consistency was strong across grades, with <b>Cronbach’s alpha ranging from .76 to .84</b> and a median of <b>.81</b> . Split-half reliability ranged from <b>.61 to .75</b> , with median correlations between halves of <b>.64</b> , indicating acceptable internal structure across forms and grades.
<b>Overall Conclusion</b>	Results provide strong evidence for the internal consistency and reliability of the easyCBM <sup>®</sup> Vocabulary measures across Grades 2–8. Findings support their use for benchmark screening and progress monitoring, with consistent performance across grades, seasons, and forms indicating sound measurement quality and form equivalence.

### Reference

Wray, K. A., Alonzo, J., & Tindal, G. (2013). *Internal consistency of the easyCBM<sup>®</sup> vocabulary measures: Grades 2–8 (Technical Report 1406)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1407:** Internal Consistency of the easyCBM<sup>®</sup> CCSS Reading Measures Grades K-8 (Guerreiro et al., 2014).

Technical Report 1407 examines the internal consistency and split-half reliability of the easyCBM<sup>®</sup> CCSS Reading benchmark assessments for Grades 3 through 8. The purpose of the study was to evaluate the reliability of the reading measures used for benchmarking and progress monitoring, particularly given their role in high-frequency instructional decision making. The CCSS Reading measures were designed to assess comprehension across literary text, informational text, and read-to-perform-a-task formats aligned with Common Core State Standards.

### Methods

Data for the study were drawn from an extant dataset collected during the 2013–2014 school year and included fall and winter benchmark scores from more than 150,000 students across grades 3 through 8. Sample sizes varied by grade and benchmark period, with fall administrations including approximately 75,900 students and winter administrations including approximately 81,200 students. The dataset reflected substantial demographic diversity, including variation in race, ethnicity, English learner status, disability status, and gender, although some demographic variables were missing for grades and testing windows.

The easyCBM<sup>®</sup> CCSS Reading measures consist of five passage-based testlets per form, each containing five multiple-choice items, for a total of 25 items per assessment. Students receive one point for each correct response, yielding a maximum possible score of 25. Measures are administered online in group settings, with automated

scoring and data capture. All benchmark and progress monitoring forms within each grade were developed concurrently and designed to be of equivalent difficulty. Prior to analysis, records with no item responses and out-of-range values were removed. Internal consistency was evaluated using Cronbach's alpha, while split-half reliability was examined by comparing first-half and second-half item sets and computing correlations between halves. Item-level analyses comparing performance of students in the upper and lower 27th percentiles were also conducted to examine discrimination patterns.

## Results

Results indicated strong internal consistency across grades and benchmark periods. Cronbach's alpha values ranged from .83 to .90, with a median of .87 across all measures. Split-half reliability coefficients ranged from .73 to .86, with median values of .76 for first halves and .82 for second halves. Item-level analyses demonstrated consistent discrimination, with higher-performing students answering items correctly more frequently than lower-performing students. Overall findings provide strong evidence supporting the reliability and technical adequacy of the easyCBM CCSS Reading measures for use in benchmarking and progress monitoring.

**Table 35. Summary of Results for Technical Report 1407**

Category	Summary
Grades	Grades 3–8
Assessment Focus	CCSS-aligned reading comprehension
Sample Size	Over 150,000 students across fall and winter benchmarks
Reliability Evidence	Median Cronbach's alpha = .87; strong split-half reliability
Statistical Methods	Cronbach's alpha and split-half reliability
Item Performance	Consistent discrimination across upper and lower ability groups
Intended Use	Benchmarking and progress monitoring

**Summary of Technical Report 1807:** Supplementary report on easyCBM MCRC Measures: A Follow-up to previous technical report (Alonzo & Anderson, 2018).

A large extant data set was used. Reliability results demonstrated adequate internal consistency across grades. Cronbach's alpha values ranged from approximately .66 (Grade 8) to .79 (Grade 4), with most grades clustering between .71 and .79. Split-half reliability coefficients showed similar patterns, ranging from approximately .60 to .80. These values indicate acceptable reliability for screening and benchmarking purposes, particularly given the time-limited nature of the assessments. Subgroup analyses further supported reliability, with higher internal consistency observed in general education and special education subgroups, often exceeding .80. Gender and ethnicity subgroup analyses demonstrated consistent reliability estimates across groups, supporting fairness and generalizability of the measures

## References

Guerreiro, M., Alonzo, J., & Tindal, G. (2014). *Internal consistency of the easyCBM<sup>®</sup> CCSS reading measures, Grades 3–8 (Technical Report No. 1407)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., & Anderson, D. (2018). *Supplementary report on easyCBM MCRC Measures: A Follow-up to previous technical report (Technical Report # 1807)*. Eugene OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1803:** Reliability of the Slope of the easyCBM<sup>®</sup> Reading Measures (Nese et al., 2018).

## Methods

The analytic sample comprised students in Grades K–8 who were identified by their school districts as needing intensive intervention and who had completed easyCBM<sup>®</sup> reading progress monitoring assessments during the 2014–2015, 2015–2016, and 2016–2017 school years. Inclusion required a minimum of 10 assessment scores per measure with at least 20 weeks between the first and last administration, resulting in sample sizes that varied by grade and measure. Six reading measures were examined: Letter Names (LN), Phoneme Segmenting (SEG), Letter Sounds (LS),

Word Reading Fluency (WRF), Passage Reading Fluency (PRF), Vocabulary (VOC), Multiple Choice Reading Comprehension (MCRC), and CCSS Reading (CCSS). Slope reliability was evaluated using two approaches: (1) Pearson split-test correlation analysis, in which each student's assessments were split into odd- and even-numbered subsets and OLS growth slopes were correlated, and (2) a reliability-of-slope approach, defined as the ratio of true score variance (random slope variance from a mixed-effects growth model) to total variance, computed in R.

## Results

Results varied systematically by measure, grade level, and analytic approach, with the most consistent and favorable reliability estimates observed for early literacy measures in the lower grades. For the Pearson split-test correlation approach, the early/emergent literacy measures—Letter Names, Phoneme Segmenting, and Letter Sounds—demonstrated reasonable reliability ( $r \geq .60$ , with lower confidence interval bounds meeting or exceeding  $.40$ ) in Kindergarten and Grade 1. Similarly strong results were obtained for Word Reading Fluency in Grade 1 ( $r = .88$ ) and Passage Reading Fluency in Grade 1 ( $r = .85$ ) and ranged from  $.25$  to  $.42$  in Grades 2–8. Vocabulary and MCRC measures showed largely weak and inconsistent correlations across grades.

The reliability-of-slope approach yielded somewhat more favorable results overall, particularly for the fluency-based measures. Word Reading Fluency demonstrated strong reliability through Grade 2 ( $r = .75$ – $1.00$ ), and Passage Reading Fluency showed acceptable reliability through Grade 3 ( $r = .56$ – $.88$ ). The early literacy measures again performed well in Grades K–1, with coefficients ranging from  $.72$  to  $.87$ . In contrast, Vocabulary, MCRC, and CCSS Reading showed weak reliability across grades under this approach as well, with many coefficients below  $.40$  and wide confidence intervals, reflecting high variability and small sample sizes for several grade levels. Across both analytic approaches, a clear developmental pattern emerged: slope reliability was strongest in the earliest grades and for measures targeting foundational reading skills, and weakest in the later grades (Grade 4 and above) and for higher-order comprehension and vocabulary measures.

**Table 36. Illustrative Results from Technical Report 1803**

Table 2  
Reliability of Slope Analyses Results

Measure	Grade	<i>n</i>	Correlation coefficient ( <i>r</i> )	Confidence Interval	
				Lower	Upper
Letter Names (LN)	K	253	.87**	.74	1.00
	1	44	.79**	.50	1.00
Phoneme Segmenting (SEG)	K	42	.83**	.53	1.00
	1	91	.72**	.51	.98
Letter Sounds (LS)	K	237	.86**	.72	1.00
	1	199	.72**	.57	.89
Word Reading Fluency (WRF)	K	10	1.00**	.45	1.00
	1	348	.92**	.81	1.00
	2	155	.75**	.58	.94
Passage Reading Fluency (PRF)	3	60	.54	.29	.87
	1	75	.88**	.65	1.00
	2	412	.62**	.52	.74
	3	387	.56**	.45	.68
	4	163	.53	.37	.72
	5	263	.42	.29	.56
	6	64	.48	.23	.81
	7	11	.03	NA	NA
8	137	.40	.22	.62	
Vocabulary (VOC)	2	41	.07	.00	.42
	3	7	.47	.00	1.00
	4	22	.48	.08	1.00
	5	17	.77	.33	1.00
	6	7	.44	.00	1.00
	7	1	NA	NA	NA
	8	32	.08	.00	.55

Note. NA = Not analyzed. \*\*Lower bound of the confidence interval around the median correlation estimate falls below 0.50 but meets or exceeds 0.40.

## References

Nese, J. F. T., Anderson, D., Irvin, P. S., & Alonzo, J. (2018). *In-brief: Reliability of the slope of the easyCBM<sup>®</sup> reading measures (Technical Report No. 1803)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1111: easyCBM<sup>®</sup> Slope Reliability: Letter Names, Word Reading Fluency, and Passage Reading Fluency (Patarapichayatham et al., 2011).****Methods**

Participants were students in Kindergarten through Grade 8 drawn from an extant database of approximately 170,000 students per grade who used the easyCBM<sup>®</sup> system during the 2010–2011 school year. After data cleaning—excluding students with fewer than three scores, off-grade-level testing, scores at or above the 50th percentile on the first progress monitoring measure, and zero scores—final sample sizes ranged from 122 to 1,146. Three curriculum-based measures (CBMs) were examined: Letter Names (K–1), Word Reading Fluency (K–3), and Passage Reading Fluency (grades 1–8), each administered in 60-second timed trials across up to 20 alternate forms. Raw scores were averaged within weeks and then collapsed into bi-weekly segments. Using structural equation modeling (SEM) in Mplus, a parallel-process linear growth model was applied by alternating bi-weekly segments into two parallel series (Process A and Process B). Individual growth slopes were estimated as empirical Bayes factor scores within each process, and the Spearman-Brown formula was applied to the inter-process slope correlation to yield a reliability coefficient for the full measurement period.

**Results**

The results of this study provided strong evidence for the reliability of growth slope estimates across all three easyCBM<sup>®</sup> reading measures and the grades in which they were administered. Across all analyses, Spearman-Brown corrected slope reliability coefficients were universally above .80, with the majority exceeding .85, indicating that the rate of change captured by these measures is highly stable and replicable across equivalent time samples within a school year.

For Letter Names, reliability estimates were .859 for kindergarten and .753 for grade 1, indicating good-to-strong slope stability in the early literacy domain. Although the grade 1 coefficient was the lowest among Letter Names estimates, it still exceeded the conventional threshold of .70 generally considered acceptable for educational measurement purposes.

Word Reading Fluency yielded slope reliabilities of .812 (grade 1), .881 (grade 2), and .853 (grade 3), reflecting strong consistency across the three grades examined. These findings suggest that this measure effectively captures stable growth trajectories in students' word-level reading skills over the course of a school year.

Passage Reading Fluency demonstrated robust slope reliability across grades 1 through 7, with coefficients ranging from .711 to .869. Notably, grades 2 through 6 all produced estimates above .77, with several grades yielding reliabilities exceeding .82. Grade 1 produced the lowest estimate in this measure (.711), likely reflecting greater variability in early oral reading fluency growth trajectories. The one clear exception to the otherwise high reliability pattern was grade 8, which yielded a substantially lower coefficient of .495. This grade 8 estimate falls below acceptable thresholds and suggests that the Passage Reading Fluency measure may not reliably track individual growth slopes among older, more advanced readers who are closer to the ceiling of the measurement scale. Taken together, these findings support the use of easyCBM<sup>®</sup> measures as reliable tools for monitoring student learning trajectories within a response to intervention (RTI) framework, particularly for students in the earlier grades or those performing below grade-level expectations. The sole notable concern is the grade 8 Passage Reading Fluency reliability, which warrants caution when interpreting slope-based decisions for students at that level. The overall pattern reinforces the value of these measures for identifying whether individual students are responding to instructional interventions across the academic year.

**Table 37. Illustrative Results from Technical Report 1111**

Table 14

Passage Reading Fluency Descriptive Statistics: Grade 7

Time Point	N	Minimum	Maximum	Mean	Std. Deviation
1A	51	44	150	112.82	24.775
1B	43	85	175	132.47	20.536
2A	109	36	170	112.83	26.431
2B	175	30	175	116.14	27.849
3A	118	23	187	120.32	26.762
3B	81	28	203	121.96	27.377
<b>4A</b>	<b>192</b>	<b>4</b>	<b>185</b>	<b>108.18</b>	<b>34.597</b>
<b>4B</b>	<b>73</b>	<b>40</b>	<b>178</b>	<b>107.96</b>	<b>30.366</b>
<b>5A</b>	<b>86</b>	<b>44</b>	<b>177</b>	<b>108.37</b>	<b>29.759</b>
5B	106	16	170	115.05	31.199
6A	141	22	192	128.26	30.802
6B	140	19	183	114.52	28.089
7A	128	45	175	119.52	27.167
7B	91	34	189	116.29	31.463
8A	90	40	186	128.24	26.770
8B	107	63	185	125.76	27.152
9A	149	23	196	123.50	33.089
9B	67	24	180	110.44	31.384
<b>10A</b>	<b>14</b>	<b>60</b>	<b>154</b>	<b>110.29</b>	<b>30.603</b>
<b>10B</b>	<b>0</b>				
Valid N (listwise)	0				

## Reference

Patarapichayatham, C., Anderson, D., Irvin, P. S., Kamata, A., Alonzo, J., & Tindal, G. (2011). *easyCBM<sup>®</sup> slope reliability: Letter names, word reading fluency, and passage reading fluency (Technical Report No. 1111)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

---

**Summary of Technical Report 2602:** Reliability Analyses for easyCBM<sup>®</sup> Measures in Grades 3-8: Total Scores for Vocabulary, Proficient Reading, and Proficient Math (Tindal & Nese, 2026).

## Methods

TR2602 reports internal-consistency reliability for three easyCBM<sup>®</sup> benchmark measures in Grades 3–8: Vocabulary, Proficient Reading, and Proficient Math. Student records were drawn from districts that permitted research use of easyCBM<sup>®</sup> server data. Each record included grade and available demographics plus item-level 0/1 responses (0 = incorrect, 1 = correct).

## Methods

Analyses were limited to Benchmark administrations and split into Fall and Winter benchmark files. Results were reported by grade (3–8) and season, alongside demographics and descriptive statistics (n, mean, SD, range, median, skewness, kurtosis) to describe score distributions.

Reliability was estimated using two complementary approaches. First, Rasch marginal reliability (IRT) was computed from expected a posteriori (EAP) theta estimates and their standard errors. Rasch models were fit in R with the TAM package using marginal maximum likelihood; marginal reliability summarizes the proportion of variance in estimated ability attributable to true differences rather than expected error. Median coefficients and 95% confidence intervals were obtained using 500 bootstrap replications for each grade-by-season cell. Second, Cronbach's alpha (KR-20 for dichotomous items) was estimated in R with the psych package and similarly summarized with bootstrapped medians and 95% intervals. Findings are presented separately for each measure. Both indices were interpreted as evidence of score precision for screening uses.

**Results: Vocabulary**

Vocabulary (20 items) showed strong reliability across grades and both benchmark seasons. Fall means rose from Grade 3 (M = 14.30, SD = 4.49) to Grades 6–8 (M ≈ 16.40–16.51), with medians of 16–18 and negative skewness. Winter means were slightly higher for most grades (e.g., Grade 3 M = 15.16; Grade 6 M = 16.46), with medians typically 17–18.

Rasch marginal reliability medians ranged from 0.74 to 0.81. The highest estimate was Grade 3 Fall (median = 0.81; 95% CI 0.81–0.82). Most grade-by-season cells clustered between 0.74 and 0.76 (e.g., Grade 5 Winter median = 0.74; CI 0.73–0.75; Grade 8 Fall median = 0.75; CI 0.74–0.75).

Cronbach's alpha converged with the IRT results. Median alphas ranged from 0.80 to 0.86, highest in Grade 3 (Fall 0.85; Winter 0.86) and lowest in Grade 5 Winter (median = 0.80; lower bound = 0.79). All other alpha lower bounds were at least 0.80. The report concludes Vocabulary total scores provide dependable internal consistency for benchmark screening in Grades 3–8 in Fall and Winter. Confidence intervals were narrow because of large n. Both indices indicate consistent total-score measurement across grades and benchmark seasons overall.

**Results: Proficient Reading**

Proficient Reading (20 items) showed adequate-to-strong reliability, but more variability than Vocabulary. Fall means ranged from Grade 3 (M = 10.76, SD = 3.97) to Grade 6 (M = 13.70, SD = 3.53), with medians 11–15. Winter means ranged from Grade 3 (M = 9.87, SD = 3.21) to Grade 5 (M = 14.71, SD = 3.98), with medians 10–16.

Rasch marginal reliability medians ranged from 0.60 to 0.78. Grade 3 Winter was lowest (median = 0.60; 95% CI 0.59–0.61). Most other cells fell in the low-to-mid 0.70s (e.g., Grade 4 Fall 0.78; Grade 5 Winter 0.77; Grade 7 Fall 0.75), with lower values in Grade 6 (Fall 0.68; Winter 0.69) and Grade 8 Winter (0.69).

Cronbach's alpha mirrored this pattern (medians 0.60–0.81). Grade 3 Winter was again lowest (median = 0.60; lower bound = 0.59) and Grade 5 Winter highest (median = 0.81; 95% CI 0.80–0.81). Grade 8 Winter was the only grade/season with an alpha lower bound below 0.70 (0.69). The report concludes Proficient Reading generally supports benchmark screening, while highlighting Grade 3 Winter (and modestly lower Grade 6/8 winter bounds) for monitoring. Confidence intervals were similarly narrow for most grades.

Tindal, G., & Nese, J. F. T. (2026). *Reliability Analyses for easyCBM<sup>®</sup> Measures in Grades 3-8: Total Scores for Vocabulary, Proficient Reading, and Proficient Math (Technical Report 2602)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

---

**Summary of Technical Report 1701:** Oral reading fluency: Outcomes from 30 years of research (Tindal, 2017).

**Methods**

Technical Report 1701 synthesizes approximately three decades of empirical and methodological research on Oral Reading Fluency (ORF) as a curriculum-based measurement approach used for screening, benchmarking, and progress monitoring in reading. The report's purpose is to document how ORF has been operationalized across studies, summarize evidence regarding growth, reliability, and validity, and describe measurement design features that influence score interpretation for instructional decision making.

ORF is described as a standardized, brief passage-reading task in which students read aloud for a fixed duration, most commonly 60 seconds, and performance is scored as words correct per minute (WCPM). Scoring conventions emphasize accurate word identification, with errors recorded for mispronunciations, substitutions, omissions, and hesitations; self-corrections are typically not counted as errors. Passage development and selection are treated as central to score comparability. The report reviews approaches used to construct passage sets that approximate within-grade difficulty targets, including the use of readability indicators, empirical calibration, and procedures intended to manage passage-to-passage variation that can obscure true change.

Evidence summarized in the report draws from multiple datasets and published studies spanning elementary through secondary grades. The reviewed work includes small investigations of special populations as well as large, regionally or nationally stratified samples used to generate norms and seasonal growth expectations, commonly covering Grades 1 through 8 and extending to Grades 9 through 12 in later work. Data collection practices described across studies typically involved universal screening during fall, winter, and spring benchmark windows, with additional administrations at shorter intervals when ORF was used for progress monitoring. Administration procedures were generally standardized, including directions, timing rules, and scoring guidelines designed to support comparability across examiners and occasions.

Analytic methods highlighted in the synthesis include classical test theory (CTT) indices of reliability and stability, generalizability analyses used to partition error attributable to passages and occasions, and item response theory (IRT) models used in selected research to evaluate passage difficulty and growth. The report cites prior work using a Rasch model to analyze ORF growth and to support development of norms across grades. Validity evidence is summarized through correlational and predictive analyses linking ORF to broader reading outcomes, including associations with standardized reading tests and large-scale assessments, as well as studies examining sensitivity to instructional change.

## Results

Across the body of evidence, ORF scores exhibit systematic growth patterns, with growth rates varying by grade level and season. Reliability is generally sufficient for screening and progress monitoring when multiple passages or repeated administrations are used, whereas single-passage estimates are more sensitive to passage-to-passage variability. Overall, the report concludes that ORF provides an efficient indicator of reading proficiency and growth when passage sets, administration procedures, and interpretive benchmarks are aligned with intended uses in data-based decision-making frameworks.

**Table 38. Summary of Results from Technical Report 1701**

<b>Area</b>	Oral Reading Fluency (ORF) is positioned as a curriculum-based indicator for screening, benchmarking, and progress monitoring of reading proficiency.
<b>Item Functioning</b>	Passage selection and calibration are identified as primary determinants of score comparability; passage-to-passage variability can meaningfully affect WCPM unless managed through design and administration controls.
<b>Reliability</b>	Across the research base, ORF demonstrates adequate reliability for intended uses when multiple passages or repeated administrations are used; single-passage estimates are less stable.
<b>Validity Evidence</b>	ORF scores show moderate to strong relationships with broader reading outcomes and standardized measures, supporting convergent and predictive validity for decision making in tiered support systems.
<b>Population Coverage</b>	Evidence spans elementary through secondary grades and includes both general and special populations, with interpretive guidance emphasizing alignment between norms, grade level, and administration season.

## Reference

Tindal, G. (2017). *Oral reading fluency: Outcomes from 30 years of research (Technical Report No. 1701)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Appendix A: Technical Report Table and Figure Titles**

Table 1. Grade-Level Benchmark Measures by Assessment Period
Table 2. Reliability for Various Measures Across Grades
Table 3. Individual Coefficients from Technical Reports
Table 4. Median Coefficients from Technical Reports
Table 5. Grade Group Summary of Coefficients from Technical Reports
Table 6. Tests of Within-Subject Effects and Contrasts from Technical Report 1201
Table 7. Sample Split-Half Coefficients from Technical Report 1201
Table 8. Distractor Analysis from Technical Report 1201
Table 9. Summary of Key Findings from Technical Report 1201
Table 10. Example Tests of Within-Subject Effect from Technical Report 1202
Table 11. Summary of Outcomes from Technical Report 1202
Table 12. Example Tests of Within-Subjects Effects from Technical Report 1203
Table 13. Example Split Half Coefficients from Technical Report 1203
Table 14. Summary of Key Findings from Technical Report 1203
Table 15. Example of Tests within Subjects from Technical Report 1204
Table 16. Summary of Key Findings from Technical Report 1204
Table 17. Example of Within-Subject Effects from Technical Report 1205
Table 18. Example of Within-Subject Contrasts from Technical Report 1205
Table 19. Summary of Key Findings from Technical Report 1205
Table 20. Example Split-Half Coefficients from Report 1206
Table 21. Example Within-Subject Effects from Report 1206
Table 22. Summary of Key Findings from Technical Report 1206
Table 23. Example Test-Retest Reliability from Technical Report 1216
Table 24. Example D-Study G and $\phi$ Coefficients from Technical Report 1216
Table 25. Example Test-Retest Reliability Word Reading Fluency Measures Technical Report 1217
Table 26. Summary of Key Findings from Technical Report 1217
Table 27. Example Test-Retest Reliability Word Reading Fluency Measures Technical Report 1218
Table 28. Summary of Key Findings from Technical Report 1218
Table 29. Example Correlations for Passage Reading Fluency Measures in Technical Report 1219
Table 30. Example Test-Retest Reliability Results Technical Report 1220
Table 31. Example Correlation Results Technical Report 1220
Table 32. Summary of Key Findings from Technical Report 1220
Table 33. Cronbach's Alpha and Split-Half Reliability from Technical Report 1406
Table 34. Summary of Key Findings from Technical Report 1406
Table 35. Summary of Results for Technical Report 1407
Table 36. Illustrative Results from Technical Report 1803
Table 37. Illustrative Results from Technical Report 1111
Table 38. Summary of Results from Technical Report 1701

Figure 1. Example G-Coefficients from Technical Report 1216
Figure 2. Example G-Coefficients from Technical Report 1217
Figure 3. Example G-Coefficients from Technical Report 1218
Figure 4. Example G-Coefficients from Technical Report 1219
Figure 5. Example G-Coefficients from Technical Report 1220

### **Appendix B: Guide to Spreadsheet Technical Report Value Displays**

See Riverside Insights or BRT to access exact values for TR Summaries  
2603-RK8R\_ReliabilityReadingTables.xlsx

- TR0906
- TR1003
- TR1004
- TR1403
- TR1201
- TR1202
- TR1203
- TR1204
- TR1205
- TR1206
- TR1216
- TR1217
- TR1218
- TR1219
- TR1220
- TR1406
- TR1407
- TR1807
- TR1803
- TR1111
- TR2602
- TR1701...See Tech Report 2603-RK8R-Reliability Reading - Page 44

### Technical Report References

- Alonzo, J., & Anderson, D. (2018). *Supplementary report on easyCBM MCRC Measures: A Follow-up to previous technical report (Technical Report # 1807)*. Eugene OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Lai, C. F., Anderson, D., Park, B. J., & Tindal, G. (2012). *An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM reading assessments: Grade 4 (Technical Report # 1219)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., & Tindal, G. (2009). *Alternate form and test-retest reliability of easyCBM reading measures (Technical Report # 0906)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Anderson, D., Lai, C. F., Park, B. J., Alonzo, J., & Tindal, G. (2012). *An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM reading assessments: Grade 2 (Technical Report # 1217)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Anderson, D., Park, B. J., Lai, C., F., Alonzo, J., & Tindal, G. (2012). *An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM reading assessments: Grade 1 (Technical Report # 1216)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Guerreiro, M., Alonzo, J., & Tindal, G. (2014). *Internal consistency of the easyCBM CCSS reading measures grades K-8 (Technical Report # 1407)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Irvin, P. S., Alonzo, J., Lai, C. F., Park, B. J., & Tindal, G. (2012). *Analyzing the reliability of the easyCBM reading comprehension measures: Grade 7 (Technical Report # 1206)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Irvin, P. S., Alonzo, J., Park, B. J., Lai, C. F., & Tindal, G. (2012). *Analyzing the reliability of the easyCBM reading comprehension measures: Grade 6 (Technical Report # 1205)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Jamgochian, E. M., Park, B. J., Nese, J. F. T., Lai, C. F., Sáez, L., Anderson, D., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM Grade 2 reading measures (Technical Report # 1004)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Lai, C. F., Irvin, P. S., Alonzo, J., Park, B. J., & Tindal, G. (2012). *Analyzing the reliability of the easyCBM reading comprehension measures: Grade 2 (Technical Report # 1201)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Lai, C. F., Irvin, P. S., Park, B. J., Alonzo, J., & Tindal, G. (2012). *Analyzing the reliability of the easyCBM reading comprehension measures: Grade 3 (Technical Report # 1202)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Lai, C. F., Nese, J. F. T., Jamgochian, E. M., Kamata, A., Anderson, D., Park, B. J., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM primary-level reading measures (Grades K-1), 2009-2010 Version. (Technical Report # 1003)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Lai, C. F., Park, B. J., Anderson, D., Alonzo, J., & Tindal, G. (2012). *An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM reading assessments: Grade 5 (Technical Report # 1220)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Nese, J. F. T., Anderson, D., Irvin, P. S., & Alonzo, J. (2018). *In-Brief: Reliability of the slope of the easyCBM<sup>®</sup> reading measures (Technical Report # 1803)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Park, B. J., Anderson, D., Alonzo, J., Lai, C. F., & Tindal, G. (2012). *An examination of test-retest, alternate form reliability, and generalizability theory study of the easyCBM reading assessments: Grade 3 (Technical Report # 1218)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Park, B. J., Irvin, P. S., Alonzo, J., Lai, C. F., & Tindal, G. (2012). *Analyzing the reliability of the easyCBM reading comprehension measures: Grade 4 (Technical Report # 1203)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Park, B. J., Irvin, P. S., Lai, C. F., Alonzo, J., & Tindal, G. (2012). *Analyzing the reliability of the easyCBM reading comprehension measures: Grade 5 (Technical Report # 1204)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Patarapichayatham, C., Anderson, D., Irvin, P. S., Kamata, A., Alonzo, J., & Tindal, G. (2011). *easyCBM slope reliability: Letter names, word reading fluency, and passage reading fluency (Technical Report # 1111)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Tindal, G. (2017). *Oral reading fluency: Outcomes from 30 years of research (Technical Report # 1701)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

- Tindal, G., & Nese, J. F. T. (2026). *Reliability Analyses for easyCBM<sup>®</sup> Measures in Grades 3-8: Total Scores for Vocabulary, Proficient Reading, and Proficient Math (Technical Report # 2602)*. Eugene, OR.: Behavioral Research and Teaching, University of Oregon.
- Wray, K., Alonzo, J., & Tindal, G. (2014). *Internal consistency of the easyCBM vocabulary measures Grades 2-8 (Technical Report # 1406)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Wray, K., Lai, C. F., Sáez, L., Alonzo, J., & Tindal, G. (2014). *easyCBM beginning reading measures: Grades K-1 alternate form reliability and criterion validity with the SAT-10 (Technical Report # 1403)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.