

**Technical Report 2603-TDK8R**

**Test Development for easyCBM® in Grades K-8: Reading**

Gerald Tindal, 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®. The report does not reflect any endorsement by any of these organizations.

Copyright© 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. & McCaslin, S. (2026). *Test Development for easyCBM® in Grades K-8: Reading (Technical Report 2603-TDK8R)*. 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-TDK8R\_ItemDevReading

<b>Introduction and Overview</b>	<b>4-13</b>
<b>Summary of Technical Reports</b>	
0039: The Development of Early Literacy Measures for Use in a Progress Monitoring Assessment System: Letter Names, Letter Sounds, and Phoneme Segmenting	13
0040: The Development of Word and Passage Reading Fluency Measures in a Progress Monitoring Assessment System	15
0043: The Development of Fifth-grade Passage Reading Fluency Measures for Use in a Progress Monitoring Assessment System	17
0046: The Development of Middle School Passage Reading Fluency Measures for Use in a Progress Monitoring Assessment System	19
0041: Examining the Technical Adequacy of Reading Comprehension Measures in a Progress Monitoring Assessment System	21
0807: Examining the Technical Adequacy of Fifth Grade Reading Comprehension Measures in a Progress Monitoring Assessment System	22
0808: Examining the Technical Adequacy of Second Grade Reading Comprehension Measures in a Progress Monitoring Assessment System	24
1102: The Development and Technical Adequacy of Seventh-Grade Reading Comprehension Measures in a Progress Monitoring Assessment	28
1209: The Development of CBM Vocabulary Measures: Grade 2	30
1210: The Development of CBM Vocabulary Measures: Grade 3	32
1211: The Development of CBM Vocabulary Measures: Grade 4	34
1212: The Development of CBM Vocabulary Measures: Grade 5	35
1213: The Development of CBM Vocabulary Measures: Grade 6	37
1214: The Development of CBM Vocabulary Measures: Grade 7	38
1215: The Development of CBM Vocabulary Measures: Grade 8	40
1221: The Development of the easyCBM® CCSS Reading Assessments: Grade 3	41
1222: The Development of the easyCBM® CCSS Reading Assessments: Grade 4	42
1223: The Development of the easyCBM® CCSS Reading Assessments: Grade 5	43
1224: The Development of the easyCBM® CCSS Reading Assessments: Grade 6	44
1225: The Development of the easyCBM® CCSS Reading Assessments: Grade 7	45
1226: The Development of the easyCBM® CCSS Reading Assessments: Grade 8	46
1110: Progress Monitoring Instrument Development: Silent Reading Fluency, Vocabulary, and Reading Comprehension	47
1301: The Development of easyCBM® Spanish Literacy Assessments for Use in Grades K–2	47
1901: Initial Norms for the easyCBM® Spanish Language Vocabulary Assessments (Grades 2 – 8)	49
<b>Appendix A: Technical Report Table Titles</b>	<b>52-53</b>
<b>Technical Report References</b>	<b>54-55</b>

### Abstract

This document summarizes item development and key findings across a set of Behavioral Research and Teaching technical reports supporting easyCBM<sup>®</sup> reading-related measures (and selected supporting work). Development followed explicit test specifications, multi-stage review for alignment and accessibility, pilot testing with standardized administration, and psychometric evaluation using methods matched to measure structure. Rasch calibration was commonly used for early literacy and vocabulary, while passage reading fluency relied on classical form comparisons and comprehension measures used IRT testlet modeling to address passage effects. Across reports, most items and forms demonstrated acceptable fit and targeting, and multiple comparable alternate forms were assembled for benchmarking and progress monitoring. Spanish literacy findings underscore the importance of language-specific constructs, with syllable-level measures showing strong relations to fluency. Results in this summary are illustrative of those reported in each technical report; readers should consult the original documents for full analytic detail and context. for K-8 instructional decision making.

### The Development of easyCBM

easyCBM<sup>®</sup> was created by the Behavioral Research and Teaching (BRT) research unit in the College of Education at the University of Oregon. Development began with a grant from the federal Office of Special Education Programs in 2006, bolstered by subsequent grants from the Institute of Education Sciences (IES). In the spring of 2011, the University of Oregon partnered with Riverside Insights to expand easyCBM<sup>®</sup> to support the needs of school- and district-wide implementations. Because of the dynamic nature of the system, information derived from easyCBM<sup>®</sup> reflects the most current research and practice for schools.

### Basic Tenets of Curriculum Based Measurement

easyCBM<sup>®</sup> assessments are Curriculum Based Measures (CBMs), which are standardized measures that sample from a year's worth of curriculum to assess the degree to which students have mastered the skills and knowledge deemed critical at each grade level. They are also known as 'general outcome measures.' Curriculum Based Measurement (CBM) has a long research history, beginning with Stanley Deno and colleagues at the University of Minnesota. CBM was originally created to assist special education teachers in developing individual education plans and monitoring student progress. The use of these measures soon expanded to include general education, as they provide reliable and valid assessments of student progress in reading and mathematics (Tindal, 2013)<sup>1</sup>. In particular, the measures can be used for universal screening (benchmark testing) and progress monitoring, as they are sensitive to small, incremental changes in performance and are efficient to administer and score. Curriculum Based Measures such as Oral Reading Fluency correlate highly with state standardized reading tests and reading comprehension.

The measures that are part of the easyCBM<sup>®</sup> system are often referred to as 'next-generation CBMs,' as an advanced form of statistics, Item Response Theory (IRT), was used during development to increase the consistency of the alternate forms of each measure and to increase the sensitivity of the measures to monitor growth. At each grade level, alternate forms of each measure are designed to be of equivalent difficulty, so as teachers monitor student progress over time, changes in score reflect changes in student skill rather than variations in the difficulty level of the test forms.

---

<sup>1</sup> Tindal, G. (2013). Curriculum-based measurement: A brief history of nearly everything from the 1970s to the present. *ISRN Education (International Scholarly Research Network)*, Volume 2013, Article ID 958530, 29 pages. <http://dx.doi.org/10.1155/2013/958530>

## easyCBM<sup>®</sup> K–8 Reading Blueprint using CCSS and National Reading Panel

The easyCBM<sup>®</sup> Reading Assessment System (K–8) is vertically articulated and grounded in the National Reading Panel (2000)<sup>2</sup> and the Common Core State Standards (2010)<sup>3</sup>. Foundational skills (K–1) measure letter names, letter sounds, phonemic awareness and phonics; Grades 2–8 integrate word and passage reading fluency, vocabulary, and comprehension. Passage Reading Fluency supports automaticity; Vocabulary and Proficient Reading assess literal, inferential, and evaluative comprehension. Alternate-form design, story-grammar consistency, and Rasch scaling support reliability and validity. The blueprint demonstrates construct coherence across grades and alignment to MTSS screening frameworks. This technical blueprint documents the easyCBM<sup>®</sup> Reading assessment system (Grades K–8) blueprint using the National Reading Panel (NRP, 2000) framework and the Common Core State Standards (CCSS). It integrates task descriptions from the easyCBM<sup>®</sup> User Manual<sup>4</sup> and a benchmark Student Reading Booklet and articulates a comprehensive construct representation of reading across grade levels.

The National Reading Panel (NRP, 2000) identified five essential components of reading: (1) Phonemic Awareness, (2) Phonics, (3) Fluency, (4) Vocabulary, and (5) Comprehension. These components represent a developmental progression in which early decoding skills support automaticity, which in turn supports comprehension and higher-order reasoning. The blueprint primarily used the NRP for Foundational skills to address phonemic awareness, phonics, and fluency in Grades Kindergarten, 1 and 2, while from Grade 3 through 8, Passage Reading Fluency was deployed with Vocabulary and both Basic and Proficient Reading Comprehension.

In Grades 3 and beyond, the Common Core State Standards (CCSS) for English Language Arts organize Reading Literature (RL) / Reading Informational Text (RI) anchor standards (K–12). RL/RI standards address comprehension processes: key ideas, craft and structure, integration of knowledge, and range of reading.

### Foundation for Reading Measurement

We have divided our blueprint into Early Reading Foundational Skills in Grades Kindergarten through Grade 2 and then Comprehensive Reading Skills in Grades 3 through 8 expressed through Word and Passage Fluency, Vocabulary, and finally Basic and Proficient Reading Comprehension.

#### Early Reading Foundational Skills Blueprint

**Letter Names (LN)** is a timed, individually administered measure for Kindergarten and Grade 1. It consists of both upper- and lower-case alphabet letters presented in an order based on empirical evidence of difficulty. The student is asked to name the letters presented by row. The student's score is the number of letter names identified correctly in one minute.

The English language contains 26 letters and with both upper and lower case considered, the total sample is 52 letters (though some letters look the same and differ only in size). The blueprint is based on each form presenting a random sample of letters in both upper (capitalized) and lower case with 10 letters per row and 10 rows. Students name letters from left to right and continue for 1 minute. Letters named incorrectly are scored 0 and letters named correctly are scored 1. At the end of one-minute, the test is stopped and scored for the number of letters named correctly and incorrectly with the difference providing the rate correct per minute.

The English language contains 26 letters but 44 sounds: 19 consonants, 7 digraphs, 5 'r-controlled' sounds, 5 long vowels, 5 short vowels, 2 'oo' sounds, 2 diphthongs. The blueprint is based on each form presenting a random sample of letter sounds with 10 per row and 10 rows. Students sound the letter(s) presented from left to right and continue for 1 minute. Letters sounded incorrectly are scored 0 and letters scored correctly

---

<sup>2</sup> National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature*. Washington, DC: National Institutes of Health.

<sup>3</sup> National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common Core State Standards*.

<sup>4</sup> easyCBM<sup>®</sup> District User Manual (2025). University of Oregon, Behavioral Research & Teaching.

are scored 1. At the end of one-minute, the test is stopped and scored for the number of letters sounded correctly and incorrectly with the difference providing the rate correct per minute.

**Phoneme Segmenting (PS)** is a timed, individually administered measure for Kindergarten and Grade 1. The student is asked to identify the individual phonemes in each word that is orally presented by the assessor. In contrast to letter sounds, this task is presented by having the administrator sound out a word and the students. For example, the administrator is directed to state: “I am going to say a word, and you will give me the sounds you hear in that word. If I say *cap*, you will say /c/ /a/ /p/. If I say *it*, you will say /i/ /t/. If I say *top*, you will say /t/ /o/ /p/. Let’s try.” The student's score is the number of correct phonemes or segments identified in one minute.

### **Reading Fluency Measure Blueprint**

**Word Reading Fluency (WRF)** is a timed, individually administered measure for Kindergarten to Grade 3 which assesses the fluency of words read in isolation. The student is given a list of words presented in order of increasing difficulty and is asked to read the words row by row. The student's score is the number of words read correctly in one minute.

Words for Word Reading Fluency were selected from a variety of sources, including Dolch word lists, online grade-level word lists, and Fry's 'instant 1000 words.' They include words with both regular and irregular sound patterns and in a variety of lengths. The words were piloted in a large multi-grade study in 2006; the difficulty of each word was then calculated, and test forms were constructed to be of equivalent difficulty within each given grade.

**Passage Reading Fluency (PRF)** is a timed, individually administered measure for Grades 1 to 8 which assesses the fluency of words read in context. Each passage consists of a brief work of original narrative fiction with content and passage complexity suitable for a given grade level. The student's score is the number of words read correctly in one minute.

For passage construction, the passages maintain a consistent discourse profile, so performance differences are not driven by shifts in text type or structure, and that forms maintain stable reading demands and comparable overall difficulty. These measures were created to follow the same story grammar across passages (introduction, rising action, climax, resolution) to reduce variation in difficulty across the different forms. The plots and characters were selected to be appropriate for the grade level written.

For both fluency measures (words and passages), performance is scored with incorrectly pronounced words (substitutions or omissions) as 0 points and correctly pronounced words as 1 point. The total is the number of words correct, incorrect, and words correct per minute. Within the broader literacy construct progression, fluency is an intermediate outcome that draws on earlier alphabetic-principle and phonological-awareness skills and serves as a strong predictor of early comprehension. Fluency measures are therefore intended to capture a student’s ability to decode and recognize words rapidly and accurately in connected text (passage reading fluency) and in word lists (word reading fluency), with sensitivity to small changes over short instructional intervals (about 1–2 weeks).

### **Vocabulary Measure Blueprint**

Vocabulary is an untimed assessment that measures vocabulary proficiency appropriate for a student's grade level, designed for students in Grades 2 to 8. The student's score is the number of correct responses. The words included in the Vocabulary measure were selected from a variety of content materials and were extensively field-tested. The bank of items represents a wide range of difficulty aligned to grade level content standards.

The Vocabulary CBM blueprint (Grades 2–8) is built around a large, standards-referenced item bank used to assemble benchmark and progress-monitoring forms. Item creation begins with explicit targets: writers use the Oregon State Standards for Vocabulary, and each item is tagged to a specific sub-domain within a standard in the item database. Writers are recruited for English language arts expertise and experience with struggling learners (including special education), and they are asked to produce items spanning very low to grade-level performance ranges.

Item format is tightly specified for computer delivery. Each item presents a single sentence with the target word highlighted, followed by three multiple-choice options. Writers were trained to reduce construct-irrelevant complexity while preserving standards alignment: use simple, direct language, avoid idioms and unnecessary clauses, focus attention on one idea, and keep grammar parallel between the stem and each option. Options are designed to be similar in length and complexity, mutually exclusive, and differentiated primarily by meaning. Writers use grade-appropriate vocabulary and contexts (supported by reference lists) and craft distractors that are plausible and relevant to the sentence.

Procedures support alternate-form comparability. Items were authored in structured templates, reviewed for clarity, alignment, formatting, and keying accuracy, and then piloted online using a design that combines anchor items with randomly selected items from the bank. Online delivery presents one item per screen, captures responses automatically, and randomizes option order to reduce cheating. Item statistics from piloting (difficulty and appropriateness) are then used to assemble multiple comparable benchmark and progress-monitoring forms from the calibrated bank.

### **Basic Comprehension Measure Blueprint**

Basic Reading is an untimed assessment that measures student comprehension utilizing a variety of text, designed for students in Grades 3 to 8. The Basic Reading measures include Read to Perform a Task, Informational Text, and Short Literary Text passages. Each of the passages is about 200 words with 5-10 questions presented after each passage. The Basic Reading Comprehension blueprint uses short sets of text-dependent questions linked explicitly to grade-level standards. Item writers (educators with grade-level expertise and experience with struggling readers) were trained in Universal Design for Assessment, item development conventions, and the content standards. Writers created literary stories and informational passages for Literary and Informational Text subtests and designed graphics for “Read to Perform a Task” subtests; a graphic artist finalized visuals for computer delivery. For each prompt (text or graphic), writers produced five selected-response questions, each aligned to a specific grade-level standard. Every question includes a stem plus three options (one correct answer and two plausible distractors), supporting efficient scoring and reduced construct-irrelevant variance.

Students complete a total of 25 items organized as five subtests (each subtest = one passage/graphic plus five questions). Materials were developed for online delivery, stored in a secure database, and subjected to multi-reviewer quality checks (3–5 trained reviewers) for bias, grade appropriateness, and adherence to Universal Design principles, with revisions made directly in the database. The blueprint anticipates large item pools: approximately 900 items per grade (Grades 3–8) are available for empirical testing.

Common items across test takers enable concurrent scaling of the full grade-level pool even though each student sees only a small sample. The computer system randomly assigns subtests to reduce class effects and broaden sampling. Overall, the blueprint yields a standards-mapped, computer-scored comprehension system with consistent prompt-question structure, systematic review, and piloting designed to calibrate a large bank into usable forms.

### **Proficient Comprehension Measure Blueprint**

Proficient Reading is an untimed assessment that measures student comprehension of written text, designed for students in Grades 2 to 8. Students read an original narrative fiction story and answer multiple-choice questions that are text dependent. Passages contain approximately 700 words in Grade 2 to around 1,300-1,500 words in Grades 3-8. Multiple choice questions at the end of the passage assess students' literal (approximately 8-10), inferential (approximately 6-8), and (in Grades 3 to 8) evaluative comprehension of text (approximately 4-6). The Proficient Reading Comprehension is a group-administered, passage-based measure for Grades 2–8. Each item includes a stem plus three answer choices (one correct, two plausible distractors) and is scored 1 point per correct response (12 points in Grade 2; 20 points in Grades 3–8). The selected-response format was chosen to resemble classroom reading tasks while improving reliability, ease of use, and cost-effectiveness relative to open-ended formats.

Passage construction was controlled through written story specifications to increase comparability and minimize construct-irrelevant variance across forms. Story writers received guidance on length (from 700 words in Grade 2 to approximately 1,300-1,500 words in Grades 3 through 8) and key narrative/discourse

features—characters, setting, and plot—so forms are similar in structure and demand. Stories were drafted by trained educators/graduate students and screened against criteria before piloting, with quality control embedded in both story and item development.

Question design followed a deliberate comprehension blueprint. Across grades, items assess literal and inferential comprehension; evaluative comprehension is added for Grades 3–8. In Grade 2, the specified distribution is 7 literal and 5 inferential items. In Grades 3 through 8, Specifications require a range of difficulty within each comprehension type and guide sequencing so forms progress from easier literal items toward the most challenging items, ending with the highest comprehension level assessed in that grade. The result is a set of comparable narrative passages paired with systematically balanced question types, difficulty, and ordering across alternate forms.

### Summary of Grade and Season Tasks for Benchmarks

- Kindergarten students take LN, LS, and PS (fall), LS, PS, WRF (winter and spring)
- Grade 1 students take LS, PS, and WRF (fall) and LS, WRF, and PRF (winter and spring)
- Grade 2-8 students take PRF, Vocabulary, and Proficient Reading Comprehension (fall, winter, and spring)

### Grade Level Summaries of Blueprints

#### Kindergarten Blueprint

Kindergarten students complete Letter Sounds (LS), Phoneme Segmenting (PS), and Word Reading Fluency (WRF) in winter and spring. Fall additionally includes Letter Names (LN). These tasks directly measure NRP foundational components.

**Table 1. Tasks for Kindergarten Blueprint**

Task	Construct (NRP)	CCSS Alignment	Administration	Score
Letter Names (LN)	Alphabetic Knowledge	RF.K.1d	1-minute timed	Letters correct/min
Letter Sounds (LS)	Phonics	RF.K.3a–c	1-minute timed	Sounds correct/min
Phoneme Segmenting (PS)	Phonemic Awareness	RF.K.2d	1-minute timed	Phonemes correct/min
Word Reading Fluency (WRF)	Fluency/Decoding	RF.K.3–4	1-minute timed word list	Words correct/min

Kindergarten measures focus exclusively on foundational skills. No comprehension task is administered. The blueprint emphasizes phonemic awareness and alphabetic principle as precursors to decoding fluency.

#### Grade 1 Blueprint

Grade 1 students transition from foundational skills toward connected-text fluency. Winter tasks include Letter Sounds (LS), Word Reading Fluency (WRF), and Passage Reading Fluency (PRF).

**Table 2. Tasks for Grade 1 Blueprint**

Task	Construct (NRP)	CCSS Alignment	Administration	Score
Letter Sounds (LS)	Phonics	RF.1.3	1-minute timed	Sounds correct/min
Word Reading Fluency (WRF)	Fluency	RF.1.4a–c	1-minute timed word list	Words correct/min
Passage Reading Fluency (PRF)	Fluency	RF.1.4	1-minute oral passage	Words correct/min

PRF introduces connected-text reading with consistent story grammar structure. This supports development of automaticity and begins linking fluency with comprehension.

### Grades 2–8 Blueprint

Beginning in Grade 2, the blueprint integrates three major tasks each benchmark period: (1) Passage Reading Fluency (PRF), (2) Vocabulary, and (3) Proficient Reading.

**Table 3. Tasks for Grades 2–8 Blueprint**

Task	Construct (NRP)	CCSS Alignment	Structure	Score
Passage Reading Fluency (PRF)	Fluency	RF.2–5.4	1-minute oral passage	Words correct/min
Vocabulary	Vocabulary	RL/RI.4; L.4–6	Untimed; ~20–25 items	Number correct
Proficient Reading	Comprehension	RL/RI.1–10	2 passages × 20 items (40)	Number correct

Each Proficient Reading form contains two grade-level passages, each followed by 20 multiple-choice items (40 total). Items represent three comprehension types: literal recall, inferential reasoning, and evaluative analysis. Comprehension Blueprint Analysis: Comprehension questions align to CCSS Reading Anchor Standards. Typical distribution per passage (approximate):

- Literal (Standards 1–3): 8–10 items
- Inferential (Standards 1, 3–6): 6–8 items
- Evaluative/Integrative (Standards 6–9): 4–6 items

This distribution supports increasing cognitive demand across grades.

Vocabulary tasks measure depth of word knowledge and context-based meaning, supporting RL/RI.4 and Language standards (L.4–6).

### Vertical Construct Progression

The blueprint demonstrates a coherent progression:

- Kindergarten–Grade 1: Phonemic Awareness → Phonics → Fluency
- Grades 2–3: Fluency → Vocabulary → Basic Comprehension
- Grades 4–5: Vocabulary depth + Analytical Comprehension
- Grades 6–8: Advanced Comprehension, Text Integration, Evaluation

This progression reflects both NRP evidence and CCSS grade-level expectations. Timed CBM measures provide growth-sensitive screening data, while untimed comprehension measures capture higher-order literacy skills. Together, the tasks represent the full construct of reading. The easyCBM® K–8 Reading Assessment Blueprint demonstrates strong alignment to the National Reading Panel framework and the Common Core State Standards. Foundational skills are measured directly and efficiently in early grades, while later grades emphasize vocabulary and comprehension complexity. The design ensures construct coverage, vertical coherence, and psychometric defensibility for screening and benchmark purposes.

**Table 4. Standards Crosswalk with CCSS, NRP, and easyCBM®**

CCSS Anchor Standard	NRP Component	easyCBM® Task	Construct Evidence
RF.K–2 Foundational Skills	Phonemic Awareness / Phonics	PS, LS, LN	Timed segmentation and sound production
RF.1–5.4 Fluency	Fluency	WRF, PRF	Words correct per minute
RL/RI.1–8.1–3 Key Ideas	Comprehension	Proficient Reading	Literal & inferential questions
RL/RI.4 Vocabulary	Vocabulary	Vocabulary Task	Context-based meaning
RL/RI.8–9 Integration	Comprehension	Proficient Reading	Evaluative reasoning

The easyCBM® reading assessment system aligns with evidence-based reading instruction frameworks (National Reading Panel, 2000) and the Common Core State Standards (NGA & CCSSO, 2010). It operationalizes phonemic awareness, phonics, fluency, vocabulary, and comprehension across K–8.

In the technical documents included in the remainder of this report, scoring rules are presented for all tasks. The early reading skills (LN, LS, PS) and the fluency measures (WRF and PRF) contain explicit administration and scoring rules in the User Manual. These measures all require production responses so standardized scoring is critical. The other measures (Vocabulary, and both Basic and Proficient Reading Comprehension) use selected responses so scoring rules are built into the computer-based administration. These latter three measures are also the most important tasks requiring a blueprint related to the CCR expectations. These measures also contain explicit references in the count of items related to various types of comprehension. Processes for ensuring content accuracy and avoiding technical/editorial flaws are not only referenced, but a QA workflow is also provided in the individual technical reports. The expected distribution of content within and across ELA domains and the distribution of score points by domain/subdomain is specified in the technical documents on test development. In the early foundation skills of LN, LS, and PS, the distribution is normal but changes over the year. All other measures present normal distributions that show increases in median and mean performance but no changes in variation.

### **Item Development**

Item development is the foundation of any assessment system that intends to support high-stakes decisions, frequent progress monitoring, or both. In curriculum-based measurement (CBM) systems, item development must accomplish two objectives that can be in tension: (a) create content that is instructionally interpretable and aligned to the skills educators teach, and (b) produce scores that are psychometrically defensible across students, forms, and administrations. The technical reports summarized in this document illustrate how these objectives were addressed across a sequence of easyCBM® reading-related measures spanning early literacy (Kindergarten to Grade 2), passage reading fluency (Grades 5 through 8), vocabulary (Grades 2 through 8), CCSS-aligned reading comprehension (Grades 3 through 8), and Spanish literacy measures (Kindergarten to Grade 2). Together, they represent an integrated approach to designing brief, repeatable measures that can be used for screening and progress monitoring, while also supporting the creation of multiple alternate forms to minimize practice effects and preserve score meaning over time.

Across the reports, item and form development followed a common set of design principles. First, developers began with explicit test specifications that defined the construct, targeted grade range, intended use (benchmarking versus progress monitoring), timing and administration conditions, and the kinds of student responses required. For early literacy measures, specifications focused on foundational skills underlying the alphabetic principle: letter identification, letter-sound correspondence, and oral segmentation of phonemes. For fluency measures, specifications addressed passage characteristics such as length, genre, readability constraints, formatting, and scoring rules aligned with correct words per minute conventions. For vocabulary and CCSS reading comprehension, specifications addressed the breadth of grade-level standards, item formats (primarily multiple-choice), and cognitive demands to ensure the measures captured targeted knowledge and processes without introducing avoidable barriers.

Second, content was drafted by trained item writers or educators with relevant classroom expertise and then refined through multi-stage review. Reviews typically included checks for alignment to the construct and grade-level expectations, readability and linguistic accessibility, bias and sensitivity, and adherence to formatting and presentation constraints required for standardized delivery (paper-based for some fluency tasks, computer-based for multiple-choice tasks). Universal Design for Assessment principles were emphasized for item sets delivered online, with attention to simplified language where appropriate, clear stems and options, and avoidance of unnecessary complexity that could introduce construct-irrelevant variance. In Spanish literacy development, the review process also reflected the linguistic structure of Spanish, recognizing that syllable-level regularities and orthographic transparency change both the nature of early reading development and the indicators that best represent it.

Third, item pools and draft forms were evaluated empirically through pilot testing. Pilot designs varied by measure type but consistently aimed to collect sufficient data to evaluate item functioning, identify problematic content, and assemble comparable alternate forms. Many reports used overlapping anchor sets (often around 20% common content) to support calibration and equating across forms when groups were nonequivalent or when forms were administered across different samples. Administration procedures were standardized and used trained assessors for individually administered early literacy and fluency tasks, while computer-based platforms enabled efficient capture of large samples for multiple-choice measures. The goal was not only to evaluate average performance, but to obtain evidence about how individual items behaved across a range of student ability.

Fourth, analyses combined classical test theory (CTT) and item response theory (IRT), selected to match the structure of the measure. For early literacy and vocabulary measures, one-parameter logistic (1PL) Rasch models were frequently used to estimate item difficulty, standard errors, and fit statistics such as outfit mean square. These analyses supported decisions about retaining, revising, or removing items and helped assemble forms with comparable difficulty. For reading comprehension measures built around passages with clustered items, testlet-based IRT models were used to acknowledge local dependence among items tied to the same text. For passage reading fluency, each passage served as the scored unit, so CTT-based comparisons of means and correlations across passages provided the most direct evidence for form equivalence and the need for targeted revisions.

Finally, item development was treated as iterative. Empirical findings informed revision cycles, including the removal of items with unestimable parameters or poor distractor functioning, refinement of passages that were too easy or too difficult, and the selection of anchor forms to stabilize interpretation across administrations. Across measures, a key outcome of the development work was the creation of multiple equivalent alternate forms (often 10 to 20 forms) that support repeated measurement while controlling for unwanted variability. This emphasis on comparability reflects the central requirement for progress monitoring: change in scores should be interpretable as change in student skill, not change in form difficulty.

The technical reports also demonstrate that “item development” in CBM systems extends beyond writing good items. It includes creating a coherent assessment framework in which forms, scoring rules, administration procedures, and psychometric models work together to support specific uses. In early literacy, the emphasis is on efficiency and sensitivity to emerging skills. In fluency, the emphasis is on passage comparability and stable scoring. In vocabulary and comprehension, the emphasis is on aligning to standards while maintaining accessibility and reliable scaling. In Spanish literacy, the emphasis is on selecting indicators that reflect the linguistic properties of the language of instruction. Taken together, the reports illustrate a development approach designed to yield measures that are practical for educators and technically defensible for reporting and decision-making.

In addition, the development work emphasizes the link between item content and the decisions educators make. CBM measures are intended to be actionable, so item development prioritized representativeness (sampling skills students are expected to demonstrate), transparency (clear scoring rules and limited construct-irrelevant complexity), and efficiency (brief administration times). These considerations appear in the early literacy and fluency work, where timing and standardized prompts are essential, and in computer-based measures, where consistent item presentation supports comparability.

The reports also highlight practical trade-offs. Overly strict fit rules can remove items that help content coverage, while overly permissive rules can introduce noise. Likewise, maximizing form equivalence can reduce content variation across forms, but too little variation can increase practice effects. The development approach reflected here treats these issues as iterative design decisions supported by empirical evidence rather than as one-time judgments.

### Highlights of Findings from Technical Reports

Across the technical reports, the most consistent finding is that carefully specified development procedures can yield forms that are sufficiently comparable for repeated use in screening and progress monitoring. While the details vary by measure type, the reports collectively show that a disciplined sequence of specification, expert review, piloting, and psychometric evaluation supports the creation of item banks and forms that function as intended across grades and contexts. At the same time, it is important to interpret these findings as illustrative of those reported in each of the individual technical reports. **Note:** The summaries and tables highlight recurring patterns and representative outcomes, but they do not reproduce all analyses, nuances, or caveats from the full reports.

Early literacy development (Kindergarten through Grade 2) provides an example of item-bank based development with Rasch calibration. Letter Names, Letter Sounds, and Phoneme Segmenting items largely exhibited acceptable fit to a 1PL model, with outfit statistics typically within preferred ranges and difficulty estimates that supported differentiation among students. Items that showed extreme misfit or unestimable parameters were flagged for review and, when appropriate, removed. A practical implication is that even when a subset of items performs less ideally, the broader item pool can still support strong form construction if decisions incorporate both fit and measurement contribution (e.g., standard errors and coverage of the ability continuum). Using calibrated pools and anchor designs, the development work supported the assembly of many equivalent alternate forms, which is essential for frequent monitoring.

For passage reading fluency in Grade 5 and in Grades 6 to 8, the central psychometric question shifted from “How do items function?” to “How comparable are passages when each passage is the unit score?” The reports show that repeated measures designs and inter-form correlations are effective for identifying passages that align in difficulty and those requiring revision. In both the Grade 5 and middle school development work, correlations among passages were generally strong, supporting the conclusion that multiple forms can be used interchangeably for progress monitoring with minimal distortion. Mean differences across passages were used diagnostically: passages that consistently deviated from the target difficulty band were revised or set aside. A notable outcome is the deliberate use of anchor passages or benchmark forms that provide stable reference points within a larger set of progress-monitoring passages.

Reading comprehension development and evaluation, including the earlier comprehension adequacy report and the CCSS-aligned Grade 3 through Grade 8 assessments, illustrates the importance of matching analytic methods to test structure. When comprehension items are nested within passages, local dependence is expected; testlet-based IRT modeling provides a principled way to estimate difficulty and fit while accounting for passage effects. Across grades, the reports indicate that most items and testlets demonstrated acceptable model fit, with difficulty values spread to capture meaningful variation among students. Where items were problematic, the issues often involved limited variability, extreme difficulty, or distractor patterns that reduced discrimination. These findings point to the value of including distractor analyses alongside difficulty and fit indices, because a multiple-choice item can appear to fit an IRT model yet fail to provide useful diagnostic information if distractors don’t attract responses by lower-performing students.

The sequence of vocabulary development reports (Grades 2 through 8) provides one of the clearest examples of iterative item refinement under a Rasch framework. Across grades, the general pattern was stable calibration, acceptable fit, and the successful assembly of both benchmark and progress-monitoring forms with closely aligned mean measures. Items with underfit or non-functioning distractors were removed, while overfitting items were sometimes retained when distractor functioning indicated that the item behaved predictably rather than misleadingly. The reports also emphasize that universal design considerations are not merely a fairness add-on; they support technical quality by reducing construct-irrelevant variance that can inflate error and reduce interpretability. The practical result is a set of short vocabulary measures that can be administered repeatedly, scaled consistently, and interpreted within an RTI decision-making framework.

The Spanish literacy development report adds an important linguistic and conceptual finding: the strongest early indicators of Spanish reading fluency were often syllable-level measures rather than phoneme-level measures. This is consistent with Spanish orthographic transparency and highlights a key lesson for assessment development: constructs must be operationalized in ways that reflect how skills develop in the language and instructional context. The technical adequacy evidence for the Spanish measures included strong alternate-form correlations for syllable reading and reading fluency measures, reinforcing the feasibility of building reliable progress monitoring tools when development aligns with language structure. Taken together, the findings support three high-level conclusions. **First**, systematic development procedures can yield technically adequate item pools and forms across diverse reading constructs and grade levels. **Second**, combining CTT and IRT evidence strengthens decision-making because different measures require different kinds of comparability evidence. **Third**, the reports collectively demonstrate that progress monitoring systems benefit from an explicit commitment to form equivalence, iterative refinement, and alignment to both developmental theory and instructional practice. These conclusions, while summarized here, should be read as representative patterns drawn from the set of technical reports rather than as a substitute for the full technical documentation in each report.

The reports also clarify that “acceptable” psychometric evidence is used as a decision framework. Fit thresholds, difficulty targets, and distractor patterns guide concrete actions: retain, revise, replace, or remove. This decision orientation matters in operational settings because it connects technical evidence to the ongoing maintenance of item pools and forms.

An important caveat for this summary is that the results presented here are illustrative of those reported in each of the technical reports. The narrative and tables emphasize recurring patterns (for example, generally acceptable fit, strong inter-form correlations, and successful assembly of multiple alternate forms), but they do not reproduce every statistic, subgroup analysis, or design nuance from the originals. Readers should consult the full reports for details on sampling, administration conditions, and measure-specific decision rules.

---

**Summary of Technical Report 0039:** The Development of Early Literacy Measures for Use in a Progress Monitoring Assessment System: Letter Names, Letter Sounds, and Phoneme Segmenting (Alonzo & Tindal, 2007a).

### **Methods**

This report describes the development of three early literacy progress monitoring measures designed for Kindergarten and first-grade students: Letter Names, Letter Sounds, and Phoneme Segmenting. The goal was to create calibrated item banks that could support the construction of multiple equivalent forms for frequent progress monitoring of foundational reading skills.

Item development began with large pools of candidate items. The Letter Names and Letter Sounds measures included uppercase and lowercase letters, while the Phoneme Segmenting measure used words that students segmented into individual phonemes. Pilot testing was conducted in spring 2006 with Kindergarten and first-grade students in a suburban Pacific Northwest school district. Response data were collected from several hundred to more than two thousand students depending on the item.

A common-item nonequivalent groups design was used so that items across different test forms could be calibrated on a single scale. Item parameters were estimated using a one-parameter Rasch model, which provided estimates of item difficulty, standard errors, and item fit statistics. These analyses were used to determine whether items functioned appropriately and should be retained in the item bank.

### **Results**

Rasch analyses produced difficulty estimates for each item across the three measures. For the Letter Names measure, most letters produced stable estimates, although three letters were removed due to unreliable parameters. The remaining items were grouped into easy, moderate, and difficult categories to guide the construction of alternate forms.

For the Letter Sounds measure, several letter-sound items with unstable estimates were excluded. Remaining items were categorized by difficulty and used to assemble equivalent forms. For the Phoneme Segmenting measure, words with poor model fit were removed, leaving 181 items in the final item bank. These items were organized into fourteen difficulty categories based on Rasch difficulty estimates. Using these calibrated item pools, the researchers constructed twenty alternate forms of each measure for Kindergarten and first grade. Items were distributed across forms according to difficulty levels to ensure comparable measurement for progress monitoring.

**Table 5. Example of IRT Analysis for Technical Report 0039**

Table 2  
*Results of IRT Analysis of Letter Names Test, Spring Pilot, 2006*

Letter	Upper or Lower Case	Count	Measure	Mean Square Outfit
o	Lower case	306	-2.55	.77
X	Upper case	297	-2.14	1.65
A	Upper case	297	-1.67	.24
s	Lower case	1036	-1.64	2.36
O	Upper case	433	-1.64	.39
B	Upper case	306	-1.42	1.29
E	Upper case	433	-1.18	1.70
a	Lower case	433	-1.18	.46
T	Upper case	433	-1.18	.33
x	Lower case	297	-1.17	.55
e	Lower case	297	-.99	.86
r	Lower case	306	-.96	.73
Z	Upper case	306	-.90	1.40
S	Upper case	306	-.77	.81
L	Upper case	297	-.72	.62
t	Lower case	306	-.71	1.17
R	Upper case	433	-.65	.96
N	Upper case	1036	-.60	.41
p	Lower case	1036	-.60	.65
C	Upper case	297	-.54	1.45
m	Lower case	433	-.52	.31
D	Upper case	306	-.49	.67
P	Upper case	297	-.40	.82
n	Lower case	306	-.39	.40
F	Upper case	433	-.39	.32
f	Lower case	306	-.34	.33
I	Upper case	433	-.27	.85
K	Upper case	306	-.20	.54
k	Lower case	297	.05	.98
M	Upper case	297	.22	1.10
i	Lower case	306	.24	.77
c	Lower case	433	.26	.73
G	Upper case	1036	.31	.73
v	Lower case	297	.51	.96
z	Lower case	433	.65	.60
W	Upper case	306	.69	1.33
U	Upper case	297	.81	.80
h	Lower case	1036	.85	.77
Q	Upper case	297	.86	.98
u	Lower case	306	.88	.69
w	Lower case	433	.92	1.54
y	Lower case	730	.98	.62
l	Lower case	433	1.06	1.80

**Table 6. Summary Table for Technical Report 0039**

Category	Summary
<b>Purpose</b>	Develop early literacy progress monitoring measures for <b>Letter Names, Letter Sounds, and Phoneme Segmenting</b> .
<b>Target Population</b>	Kindergarten and Grade 1 students learning foundational reading skills.
<b>Item Development</b>	Large item pools were created using <b>uppercase/lowercase letters</b> and <b>phoneme segmentation words</b> .

Category	Summary
<b>Pilot Testing</b>	Spring 2006 pilot with <b>hundreds to over 2,000 student responses per item</b> in a suburban school district.
<b>Item Analysis</b>	<b>Rasch modeling</b> used to estimate item difficulty and evaluate item fit; poorly functioning items were removed.
<b>Final Forms</b>	Remaining items were organized by <b>difficulty levels</b> and used to construct <b>20 equivalent alternate forms</b> for progress monitoring.

## Reference

Alonzo, J., & Tindal, G. (2007). *The development of early literacy measures for use in a progress monitoring assessment system: Letter names, letter sounds, and phoneme segmenting (Technical Report 0039)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

---

**Summary of Technical Report 0040:** The Development of Word and Passage Reading Fluency Measures in a Progress Monitoring Assessment System (Alonzo & Tindal, 2007b).

## Methods

This report describes the development and initial validation of curriculum-based measures (CBM) for reading fluency, specifically Word Reading Fluency (WRF) and Passage Reading Fluency (PRF), intended for use in kindergarten through Grade 4 as progress-monitoring tools within Response to Intervention (RTI) systems. The goal was to create sets of alternate forms that could be administered repeatedly while maintaining comparable difficulty.

For Word Reading Fluency, researchers created a large bank of candidate words representing a wide range of reading difficulty. Individual words were analyzed using Rasch item-response modeling to estimate item difficulty and to place words on a common measurement scale. Based on the Rasch difficulty estimates, the words were organized into four difficulty strata, which were then used to construct equivalent alternate forms. Each form included a balanced mix of words from the four strata to ensure consistent difficulty across administrations. A total of 20 alternate WRF forms were constructed.

For Passage Reading Fluency, researchers wrote a large pool of candidate reading passages aligned with grade-level expectations for vocabulary, syntax, and readability. Each passage was reviewed by both the report's lead author and an experienced classroom teacher to ensure instructional appropriateness and comparable complexity. From this pool, 20 passages per grade level were selected to serve as alternate forms. Pilot testing data were collected from elementary students, and analyses focused on mean score comparisons and inter-form correlations to determine whether the passages functioned equivalently.

## Results

Results indicated that the development procedures produced strong alternate-form equivalence for both reading fluency measures. Rasch analyses confirmed that the word difficulty estimates used to construct the WRF forms produced balanced distributions of item difficulty across forms. Mean performance levels across WRF forms were similar, indicating that no individual form was substantially easier or harder than others.

For the Passage Reading Fluency measures, analyses of pilot data showed strong correlations among alternate passages, demonstrating that the passages measured the same underlying fluency construct. Comparisons of average student performance across passages showed only minor variation, suggesting that the development procedures successfully controlled passage difficulty.

Together, these findings indicate that the 20 alternate forms developed for both WRF and PRF provide technically adequate measures for repeated progress monitoring in early reading. The combination of Rasch modeling, expert review, and empirical form-equivalence analyses supported the conclusion that the measures can be administered frequently to monitor student growth without introducing substantial measurement error due to form difficulty differences.

**Table 7. Typical Results of IRT Analysis from Technical Report 0040**

Table 2  
*Results of IRT Analysis of Word Reading Fluency, Spring Pilot, 2006, Grades K-3*

Word	Count	Measure	Mean Square Outfit
the	1960	-6.21	1.10
or	1960	-2.81	2.72
will	2654	-3.78	0.85
number	2198	-1.55	0.66
of	1960	-2.93	9.90
about	2654	-1.54	0.58
remain	2654	2.13	3.14
no	2654	-3.71	4.43
and	2654	-5.20	1.15
little	1960	-3.03	4.45
much	1960	-1.80	1.69

**Table 8. Example of Descriptive Statistics from Technical Report 0040**

Table 4  
*Descriptive Statistics for Grade 1 Passage Reading Measures*

Passage	Title	n	M	SD
Gr1PR_1_C	The New Puppy	60.94	40.00	207
Gr1PR_2_C	Ice Cream Treat	62.98	37.31	205
Gr1PR_3_C	Patty the Pig	55.79	34.76	204
Gr1PR_4_C	The Dog Show	62.20	37.65	205
Gr1PR_5_C	Trains	57.73	36.96	206
Gr1PR_6_C	The Best Story Ever	63.44	42.30	206
Gr1PR_7_C	Making a Pizza	71.17	38.99	206
Gr1PR_8_C	Ocean Fish	59.60	35.70	206
Gr1PR_9_C	The Puppy	58.99	36.23	201
Gr1PR_10_C	The Airplane Ride	64.14	39.44	201
Gr1PR_11_C	The Story for Mom	60.97	40.87	201
Gr1PR_12_C	Loose Tooth	56.11	38.61	200
Gr1PR_13_C	A Prized Garden	62.40	35.52	203
Gr1PR_14_C	The Lake	55.44	36.52	204
Gr1PR_15_C	The Swing	67.68	40.42	204
Gr1PR_16_C	Play Ball	60.18	37.93	202
Gr1PR_17_C	Sick Day	61.14	39.07	195
Gr1PR_18_C	Hannah the Hamster	63.73	40.13	195
Gr1PR_19_C	Playful Sam	60.70	38.99	195
Gr1PR_20_C	The Fair	64.23	41.37	195

*Results of Repeated Measures Analysis Across All 20 Passages*

Grand Mean	Standard Error	95% Confidence Interval	
		Lower Bound	Upper Bound
63.43	2.89	57.73	69.12

**Table 9. Key Findings Summary from Technical Report 0040**

Category	Summary
<b>Grade Levels</b>	Kindergarten through Grade 4
<b>Assessment Focus</b>	Word Reading Fluency and Passage Reading Fluency progress monitoring
<b>Statistical Model</b>	Rasch item-response modeling used to estimate word difficulty
<b>Form Equivalence</b>	Mean score comparisons and correlations across forms
<b>Primary Outcome</b>	Alternate forms demonstrated comparable difficulty suitable for repeated CBM progress monitoring

## Reference

Alonzo, J. & Tindal, G. (2007). *The development of word and passage reading fluency measures in a progress monitoring assessment system (Technical Report # 0040)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 0043:** The Development of Fifth-grade Passage Reading Fluency Measures for Use in a Progress Monitoring Assessment System (Alonzo & Tindal, 2008a).

### Methods

This report describes the development of 20 alternate forms of a fifth-grade Passage Reading Fluency (PRF) measure, extending an existing K-4 progress monitoring assessment system. All 20 passages (~250 words each) were written by graduate students enrolled in College of Education courses in spring and summer 2007, targeting a 5.5 grade reading level.

Each passage underwent a four-stage review: (1) the lead author edited for grammatical correctness and grade-level appropriateness; (2) two graduate students divided passages into three paragraphs of even length and verified Flesch-Kincaid readability fell between 5.4 and 5.6 per paragraph; (3) an experienced fifth-grade teacher reviewed each passage for topic, wording, and style appropriateness; and (4) the lead author conducted a final review against test specifications.

Passage difficulty was analyzed using classical statistics. All 20 forms were administered to the same group of students over one week (5 forms per day), with a consistent single assessor per student, to control for tester variability. Repeated-measures correlations and mean comparisons were used to assess cross-form equivalence. Students read each passage aloud for 60 seconds; words read correctly per minute (WRCM) constituted the score.

### Results

Pilot testing was conducted in February 2008 with 68-74 fifth-grade students per passage. Mean WRCM scores ranged from 134.96 (Gr5PRF\_1, crossword puzzles) to 170.86 (Gr5PRF\_20, career choices), indicating meaningful variation in passage difficulty that required correction. Inter-form correlations were uniformly strong across all 20 passages ( $r = .76-.96$ , all  $p < .001$ ), demonstrating that the forms measured the same underlying skill consistently. Using the nine most similarly-difficulty passages as anchors, the remaining 11 passages were revised. The two easiest passages were simplified (e.g., Gr5PRF\_1: 'vertically and horizontally' changed to 'up and down and across the paper'; Gr5PRF\_2: 'sluggishly' changed to 'slowly'). Passages scoring higher than the anchor group were made more challenging (e.g., Gr5PRF\_20 had sentences lengthened and names replaced with longer alternatives). Minor character name changes were used to adjust difficulty in several additional passages (e.g., Gr5PRF\_8, Gr5PRF\_13, Gr5PRF\_14).

These revisions produced 20 comparable alternate forms. Final student copies used size 14 Verdana font; assessor copies included administration directions and a scored version with cumulative word counts. All forms were loaded to the EasyCBM website for web-based access.

**Table 10. Example of Descriptive Statistics from Technical Report 0043**

Table 2  
*Descriptive Statistics for Grade 5 Passage Reading Measures*

Passage	Topic	<i>n</i>	<i>M</i>	<i>SD</i>
Gr5PRF_1	Crossword puzzles	74	134.96	32.239
Gr5PRF_2	Playing baseball	74	156.00	38.253
Gr5PRF_3	New baby brother	74	164.15	42.690
Gr5PRF_4	Moving to a new country	74	164.53	36.961
Gr5PRF_5	Zoo animals	74	162.54	39.541
Gr5PRF_6	Seating charts	74	161.82	39.808
Gr5PRF_7	Astronaut	74	165.95	34.937
Gr5PRF_8	Jen's crush	74	154.07	40.710
Gr5PRF_9	School pranks	74	168.99	38.482
Gr5PRF_10	Piano recital	74	165.89	43.019
Gr5PRF_11	Spider snack	70	160.26	41.254
Gr5PRF_12	Puppy of her own	70	165.74	42.009
Gr5PRF_13	Time machine	70	153.63	41.202
Gr5PRF_14	Real life hero	70	149.86	36.328
Gr5PRF_15	Spelling quiz	70	160.31	33.414
Gr5PRF_16	Tooth aches	68	164.57	38.255
Gr5PRF_17	Talent show	69	163.43	37.618
Gr5PRF_18	Book club	69	162.86	38.926
Gr5PRF_19	New bike	68	160.31	37.912
Gr5PRF_20	Career choices	69	170.86	38.611

**Table 11. Sample Correlations from Technical Report 0043**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1																			
2	.89**	1																		
3	.89**	.94**	1																	
4	.86**	.92**	.93**	1																
5	.88**	.93**	.94**	.93**	1															
6	.92**	.94**	.94**	.93**	.92**	1														
7	.89**	.92**	.92**	.89**	.91**	.93**	1													
8	.90**	.91**	.92**	.91**	.90**	.94**	.91**	1												
9	.89**	.94**	.93**	.92**	.93**	.95**	.92**	.96**	1											
10	.86**	.94**	.92**	.92**	.91**	.94**	.93**	.93**	.94**	1										
11	.87**	.93**	.94**	.93**	.93**	.94**	.92**	.93**	.92**	.94**	1									
12	.84**	.90**	.92**	.89**	.90**	.94**	.92**	.92**	.92**	.93**	.94**	1								
13	.82**	.89**	.91**	.88**	.89**	.93**	.85**	.86**	.86**	.88**	.85**	.87**	1							
14	.80**	.83**	.85**	.81**	.84**	.87**	.84**	.86**	.85**	.87**	.89**	.88**	.81**	1						
15	.80**	.85**	.85**	.83**	.84**	.85**	.83**	.84**	.84**	.86**	.85**	.81**	.76**	.79**	1					
16	.83**	.89**	.87**	.85**	.87**	.85**	.91**	.95**	.95**	.91**	.93**	.91**	.89**	.88**	.84**	1				
17	.87**	.93**	.92**	.91**	.89**	.95**	.87**	.87**	.88**	.92**	.85**	.85**	.81**	.79**	.78**	.88**	1			
18	.83**	.85**	.86**	.83**	.82**	.87**	.87**	.87**	.87**	.88**	.93**	.85**	.85**	.81**	.79**	.78**	.88**	1		
19	.85**	.91**	.92**	.89**	.92**	.92**	.89**	.90**	.93**	.93**	.95**	.94**	.90**	.87**	.87**	.90**	.93**	.83**	1	
20	.80**	.89**	.88**	.86**	.89**	.89**	.87**	.88**	.92**	.90**	.89**	.89**	.87**	.83**	.81**	.81**	.88**	.85**	.90**	1

**Table 12. Example of Changes Made: Grade 5 Passage Reading Measures in Technical Report 0043**

Passage	Changes made to the passage
Gr5PRF_1	Made less challenging by changing Mr. Jacobson to Mr. Jacobs, ‘assigned to groups’ to ‘put in groups’, replacing ‘vertically and horizontally’ with ‘up and down and across the paper’, and changing several ‘the students’ to ‘they’.
Gr5PRF_2	Made slightly less challenging by changing ‘bannister’ to ‘railing’, ‘dreamt’ to ‘had dreamed’, and ‘sluggishly’ to ‘slowly’
Gr5PRF_3	No changes made
Gr5PRF_4	No changes made
Gr5PRF_5	No changes made
Gr5PRF_6	No changes made
Gr5PRF_7	No changes made
Gr5PRF_8	Made slightly easier by changing ‘Jasmine’ to ‘Jen’
Gr5PRF_9	Made slightly more difficult by changing Jose to Mario, lengthening a few sentences.

**Table 13. Key Findings Summary from Technical Report 0043**

Category	Summary
<b>Grade Level</b>	Grade 5 only
<b>Assessment Focus</b>	Passage Reading Fluency progress monitoring
<b>Measure Type</b>	Passage Reading Fluency (PRF) — oral reading of connected narrative text
<b>Alternate Forms</b>	20 forms, each approximately 250 words, piloted February 2008
<b>Readability Target</b>	Flesch-Kincaid level 5.4–5.6 per paragraph (3 paragraphs per passage)

<b>Statistical Model</b>	Classical statistics: repeated-measures correlations and mean comparisons
<b>Pilot Sample</b>	68–74 students; each completed 5 passages per day over one week
<b>Form Equivalence</b>	Strong inter-form correlations ( $r = .76-.96$ , all $p < .001$ ); targeted revisions applied
<b>Grand Mean Range</b>	134.96 (lowest: crossword puzzles) to 170.86 (highest: career choices) WRC/min
<b>Primary Outcome</b>	20 comparable alternate forms for Grade 5 progress monitoring, loaded to EasyCBM <sup>®</sup>

## Reference

Alonzo, J., & Tindal, G. (2008). *The development of fifth-grade passage reading fluency measures for use in a progress monitoring assessment system (Technical Report 0043)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

**Summary of Technical Report 0046:** The Development of Middle School Passage Reading Fluency Measures for Use in a Progress Monitoring Assessment System (Alonzo, Park, et al., 2008).

This technical report describes the **development of middle school Passage Reading Fluency** measures for students in grades 6 through 8 as part of an expanded progress monitoring assessment system. The measures were designed to extend oral reading fluency assessment beyond the elementary grades and to provide educators with reliable tools for monitoring reading development during middle school.

## Methods

This report describes the development of Passage Reading Fluency (PRF) progress monitoring measures for middle school students in Grades 6 through 8. The goal of the project was to create multiple alternate forms that could be administered repeatedly within a curriculum-based measurement (CBM) system while maintaining comparable difficulty across passages.

Researchers initially developed a large pool of reading passages aligned with middle school reading expectations. Passages were written by graduate students and reviewed by the lead author as well as panels of graduate students and classroom teachers to ensure appropriate readability, vocabulary level, and content quality. From this pool, twenty alternate PRF passages were developed for use in a progress monitoring system.

To evaluate form equivalence, the researchers conducted pilot testing with middle school students and analyzed passage difficulty using classical statistical procedures. The analyses focused on mean score comparisons and correlations among passage scores to determine whether alternate forms measured the same reading fluency construct. Passage difficulty estimates were used to identify a subset of passages with similar difficulty levels that could serve as anchors for the set of forms.

## Results

Analyses of the pilot data indicated that the alternate passages showed acceptable levels of consistency for use in progress monitoring. Correlation analyses among passage scores demonstrated that student performance across passages was strongly related, suggesting that the passages measured the same underlying reading fluency construct.

Difficulty analyses identified nine passages that were highly similar in difficulty and could serve as a core reference set for the measure. The remaining passages were revised and adjusted so that their difficulty more closely matched the anchor passages. Revisions included modifications to vocabulary, sentence structure, and overall passage length to improve comparability.

After these revisions, the set of twenty passages demonstrated improved consistency in mean performance levels and correlations across forms. The results indicated that the passages functioned as comparable alternate forms suitable for repeated administration in a middle school reading progress monitoring system. Overall, the development process produced a technically adequate set of PRF measures that can be used to track reading fluency growth for students in Grades 6–8.

**Table 14. Illustrative Descriptive Statistics from Technical Report 0046**

Table 2  
Descriptive Statistics for Grade 6 Passage Reading Measures

Passage	Topic	n	M	SD
Gr6PRF_1	Finding Money	55	185.70	35.18
Gr6PRF_2	Being Bilingual	55	166.20	38.20
Gr6PRF_3	Flying	55	171.96	34.19
Gr6PRF_4	Shopping Mall	55	186.20	39.59
Gr6PRF_5	New School	55	183.43	37.14
Gr6PRF_6	Cousin Envy	55	164.85	31.40
Gr6PRF_7	Pet Rabbit	55	184.26	32.95
Gr6PRF_8	Lost Keys	55	185.09	40.14
Gr6PRF_9	Singing	55	177.91	33.13
Gr6PRF_10	New Video Game	55	195.72	32.60
Gr6PRF_11	Embarrassment	55	186.17	38.02
Gr6PRF_12	Grandma	55	176.28	33.22
Gr6PRF_13	Sign Language	55	189.20	35.97
Gr6PRF_14	Lost Bike	55	199.74	38.60
Gr6PRF_15	Budget Woes	54	193.92	36.09
Gr6PRF_16	Asking for a Date	54	183.75	37.98
Gr6PRF_17	Going to College	54	203.00	44.55
Gr6PRF_18	Group Project	54	177.62	36.50
Gr6PRF_19	Roller Coasters	54	184.00	37.65
Gr6PRF_20	Saturday Morning	53	201.58	45.77

**Table 15. Typical Correlation from Technical Report 0046**

Table 3  
Correlations between Each of the Sixth-Grade Passage Reading Fluency Forms

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	1																				
2	.90**	1																			
3	.88**	.90**	1																		
4	.93**	.92**	.92**	1																	
5	.93**	.91**	.91**	.94**	1																
6	.77**	.76**	.76**	.82**	.80**	1															
7	.83**	.82**	.78**	.85**	.87**	.79**	1														
8	.90**	.89**	.88**	.89**	.90**	.80**	.85**	1													
9	.88**	.86**	.85**	.86**	.85**	.76**	.81**	.88**	1												
10	.91**	.88**	.88**	.93**	.90**	.80**	.83**	.89**	.86**	1											
11	.90**	.89**	.90**	.91**	.90**	.81**	.87**	.90**	.90**	.92**	1										
12	.91**	.88**	.91**	.92**	.90**	.81**	.87**	.89**	.91**	.91**	.93**	1									
13	.88**	.83**	.87**	.87**	.88**	.75**	.85**	.88**	.90**	.86**	.92**	.92**	1								
14	.89**	.84**	.86**	.89**	.89**	.78**	.83**	.90**	.88**	.91**	.89**	.88**	.90**	1							
15	.88**	.87**	.86**	.89**	.88**	.82**	.86**	.89**	.94**	.90**	.93**	.92**	.93**	.93**	1						
16	.87**	.84**	.83**	.88**	.86**	.80**	.84**	.88**	.90**	.90**	.88**	.92**	.89**	.88**	.92**	1					
17	.89**	.86**	.83**	.90**	.89**	.84**	.84**	.88**	.88**	.92**	.89**	.93**	.88**	.89**	.91**	.92**	1				
18	.85**	.80**	.84**	.87**	.86**	.80**	.85**	.86**	.86**	.87**	.89**	.92**	.90**	.87**	.91**	.90**	.89**	1			
19	.89**	.87**	.87**	.89**	.91**	.84**	.87**	.90**	.88**	.89**	.89**	.91**	.89**	.89**	.91**	.92**	.91**	.92**	1		
20	.89**	.84**	.86**	.87**	.88**	.81**	.85**	.89**	.88**	.91**	.90**	.90**	.91**	.93**	.92**	.88**	.92**	.88**	.90**	1	

**Table 16. Key Findings Summary from Technical Report 0046**

Category	Summary
Grade Levels	Grades 6–8
Assessment Focus	Passage Reading Fluency
Statistical Model	Classical Test Theory
Forms Developed	17–20 forms per grade
Primary Outcome	Reliable fluency measures for middle school progress monitoring

## Reference

Alonzo, J., Park, B. J., & Tindal, G. (2008). *The development of middle school passage reading fluency measures for use in a progress monitoring assessment system (Technical Report 0046)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

### Summary of Technical Report 0041: Examining the Technical Adequacy of Reading Comprehension Measures in a Progress Monitoring Assessment System (Alonzo et al., 2007).

This technical report evaluates the reliability and performance of reading progress monitoring measures, focusing on methodological design and psychometric outcomes. The study aimed to determine whether the measures provide reliable and valid information for educational decision-making.

## Methods

The sample consisted of students across multiple grade levels, including both general education students and those receiving targeted instructional support. Assessments were administered in standardized settings by trained examiners. Data collection protocols emphasized consistent administration procedures, scoring accuracy, and repeated measurement to support reliability analyses.

Analytical methods included classical test theory statistics such as internal consistency estimates, as well as item response theory modeling to examine item-level performance. Item difficulty, discrimination, and fit indices were evaluated. Additional analyses examined score distributions and stability to assess suitability for progress monitoring.

## Results

Findings indicated that the reading measures demonstrated moderate to strong reliability across grade levels. Item difficulty values were generally centered within acceptable ranges, indicating appropriate calibration for the intended populations. Most items exhibited acceptable fit statistics, with few extreme or poorly functioning items identified. Overall, the results support the use of these measures for monitoring student progress and informing instructional decisions.

**Table 17. Example of Item Statistics from Technical Report 0041**

Table 4  
Item Statistics, Entry Order, Gr3MCI

Item Number	Raw Score	Count	Measure	Model Standard Error	Mean Square Outfit
1	78	88	-1.29	.35	.89
2	54	87	0.48	.24	.96
3	75	88	-0.96	.31	.80
4	45	87	0.96	.23	1.17
5	73	88	-0.77	.30	1.10
6	62	88	0.03	.25	.94
7	67	89	-0.24	.26	.91
8	75	88	-0.95	.31	.79
9	79	89	-1.30	.35	.98
10	27	89	1.98	.25	1.17
11	69	89	-0.39	.27	.71
12	54	89	0.53	.23	1.10
13	69	88	-0.39	.27	.78
14	53	89	0.54	.23	1.22
15	75	88	-0.87	.31	.55
16	59	87	0.22	.24	.86
17	59	89	0.17	.25	1.14
18	33	89	1.64	.23	1.16
19	45	89	1.01	.23	1.05
20	69	89	-0.39	.27	.82

**Table 18. Typical Distractor Analysis from Technical Report 0041**

Table 5  
*Distractor Analysis, Gr3MCI, Continued*

Entry #	Data Code	Score Value	Count	%	Average Measure	S.E. Mean
11	B	0	5	6	0.08	.28
	C	0	15	17	0.30	.17
	A	1	69	78	1.33	.12
	Missing	**				
12	B	0	4	4	0.57	.48
	C	0	31	35	0.84	.11
	A	1	54	61	1.27	.16
	Missing	**				
13	C	0	10	11	0.29	.27
	A	0	10	11	0.33	.18
	B	1	69	78	1.31	.12
	Missing	**				
14	B	0	15	17	0.61	.17
	C	0	20	23	1.07	.20
	A	1	53	60	1.20	.15
	Missing	**	1	1	2.66	
15	A	0	8	9	-0.20	.20
	B	0	6	7	0.13	.26
	C	1	75	84	1.30	.11
	Missing	**				
16	A	0	25	28	0.44	.16
	C	0	4	5	0.45	.44
	B	1	59	67	1.41	.12
	Missing	**	1	1	0.75	
17	B	0	8	9	0.72	.27
	A	0	20	23	0.85	.15
	C	1	59	68	1.22	.15
	Missing	**	2	2	0.95	.21
18	C	0	22	25	0.80	.16
	A	0	34	38	0.92	.13
	B	1	33	37	1.45	.22
	Missing	**				
19	C	0	18	20	0.69	.16
	B	0	26	29	0.81	.17
	A	1	45	51	1.41	.16
	Missing	**				
20	A	0	13	15	0.30	.20
	B	0	7	8	0.48	.33
	C	1	69	78	1.30	.12
	Missing	**				

**Summary of Technical Report 0807:** Examining the Technical Adequacy of Fifth Grade Reading Comprehension Measures in a Progress Monitoring Assessment System (Alonzo & Tindal, 2008b).

This technical report presents the development and evaluation of reading progress monitoring measures, with emphasis on methodological design and empirical performance. The study examined whether the measures yield reliable, instructionally meaningful data for diverse student populations.

### Methods

Participants included students sampled from 5<sup>th</sup> grade, representing both general education populations and students receiving additional instructional support. Assessments were administered under standardized conditions by trained examiners. Data collection procedures emphasized consistency in administration, scoring, and timing to support repeated-measures analyses and ensure score comparability. Analyses incorporated both classical test theory and item response theory approaches. Internal consistency estimates were calculated to evaluate reliability. Item-level analyses examined difficulty, discrimination, and fit statistics to assess measurement quality. Additional analyses explored score distributions and stability across administrations to evaluate sensitivity for progress monitoring purposes.

**Results**

Results indicated that the reading measures demonstrated moderate to strong reliability across grade levels. Item difficulty values were generally centered within acceptable ranges, supporting differentiation among students with varying skill levels. Most items exhibited acceptable fit statistics, with few items identified as poorly functioning. Overall, the findings support the technical adequacy of the measures for monitoring student progress and informing instructional decision-making.

**Table 19. Example of Item Statistics from Technical Report 0807**

Table 3  
Item Statistics, Entry Order, Gr5MCI

Item Number	Raw Score	Count	Measure	Model Standard Error	Mean Square Outfit
1	155	176	-2.27	0.26	2.64
2	105	176	0.01	0.19	0.61
3	106	176	-0.03	0.20	0.52
4	103	176	0.09	0.19	0.58
5	145	176	-1.70	0.23	2.96
6	143	176	-1.59	0.22	1.20
7	84	176	0.78	0.19	0.69
8	111	176	-0.22	0.20	0.46
9	129	176	-0.95	0.21	1.09
10	101	175	0.15	0.19	0.71
11	108	175	-0.11	0.20	0.56
12	43	174	2.27	0.20	1.73
13	75	174	1.10	0.19	0.98
14	113	174	-0.32	0.20	0.64
15	65	174	1.46	0.19	1.47
16	99	174	0.22	0.19	0.98
17	108	174	-0.13	0.20	0.50
18	114	174	-0.36	0.20	0.73
19	93	174	0.44	0.19	0.95
20	73	174	1.17	0.19	1.72

**Table 20. Illustrative Distractor Analysis from Technical Report 0807**

Table 4  
Distractor Analysis, Gr5MCI

Entry #	Data Code	Score Value	Count	%	Average Measure	S.E. Mean
1	A	0	12	7	-0.67	0.30
	C	0	9	5	0.75	0.41
	B	1	155	88	0.81	0.15
	Missing	**				
2	A	0	51	29	-1.25	0.13
	C	0	20	11	0.03	0.23
	B	1	105	60	1.79	0.14
	Missing	**				
3	C	0	55	31	-1.12	0.13
	B	0	15	9	-0.35	0.26
	A	1	106	60	1.81	0.14
	Missing	**				
4	A	0	59	34	-1.04	0.14
	C	0	14	8	-0.09	0.27
	B	1	103	59	1.82	0.14
	Missing	**				
5	B	0	4	2	-0.44	0.51
	A	0	27	15	0.50	0.23
	C	1	145	82	0.78	0.16
	Missing	**				
6		0	12	7	-0.54	0.14
		0	21	12	-0.41	0.27
		1	143	81	0.98	0.16
	Missing	**				
7	B	0	60	34	-0.97	0.14
	A	0	32	18	0.37	0.23
	C	1	84	48	2.04	0.15
	Missing	**				
8	C	0	47	27	-1.28	0.11
	B	0	18	10	-0.55	0.23
	A	1	111	63	1.76	0.14
	Missing	**				
9	B	0	22	13	-1.21	0.25
	C	0	25	14	-0.51	0.24
	A	1	129	73	1.27	0.15
	Missing	**				
10	B	0	45	26	-1.11	0.15
	A	0	29	16	-0.17	0.22
	C	1	101	57	1.8	0.15
	Missing	**	1	1	-1.51	0

**Table 21. Summary of Key Findings from Technical Report 0807**

Category	Summary
<b>Sample</b>	The sample consisted of fifth-grade students drawn from participating schools, with sample sizes for individual analyses generally ranging from approximately 48 to 50 students per form. Students represented typical general education classrooms used to evaluate progress monitoring measures.
<b>Assessment Forms</b>	Multiple fifth grade reading comprehension forms were developed and analyzed. Forms were designed to be equivalent in structure and difficulty to support repeated administration for screening and progress monitoring.
<b>Analysis Method</b>	Statistical analyses included classical test theory methods and Rasch modeling. Analyses focused on form equivalence, within-subjects comparisons across forms, and item-level performance.
<b>Items Analyzed</b>	Reading comprehension items embedded within multiple assessment forms were analyzed. Items were designed to reflect grade-level comprehension skills and varied in difficulty to capture a range of student ability.
<b>Problematic Items</b>	A small number of items exhibited weaker performance, including reduced variability or fit concerns. These items were identified for potential revision or further investigation in subsequent studies.
<b>Item Fit</b>	Rasch and descriptive analyses indicated that most items demonstrated acceptable fit and functioned as intended. Item difficulty was generally centered within recommended ranges, supporting differentiation among students.
<b>Overall Conclusion</b>	Results supported the technical adequacy of the fifth grade reading comprehension measures for use in a progress monitoring assessment system. While reliability was acceptable for screening and instructional use, the report notes that larger samples and additional analyses could further strengthen measurement precision.

### Reference

Alonzo, J., & Tindal, G. (2008). *Examining the technical adequacy of fifth grade reading comprehension measures in a progress monitoring assessment system (Technical Report 0807)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

### Summary of Technical Report 0808: Examining the Technical Adequacy of Second Grade Reading Comprehension Measures in a Progress Monitoring Assessment System (Alonzo, Liu, et al., 2008).

#### Methods

This report describes the development and pilot testing of second grade reading comprehension measures intended for use in a progress monitoring assessment system. The goal was to create multiple equivalent forms that could be administered repeatedly to track student growth.

Reading passages were written by elementary and secondary teachers and graduate students. Comprehension questions were then developed by an assessment design expert. Items were organized according to cognitive attributes (e.g., types of comprehension processes) and estimated difficulty levels to ensure coverage of different comprehension skills.

A total of 20 alternate test forms were created. The forms were piloted using a common person / common item design. Groups of students each completed five forms, with overlapping forms across groups to allow comparisons across administrations. This design allowed the researchers to evaluate item performance and establish relationships among the forms. Content and bias reviews were conducted to ensure that passages

and questions were appropriate for second-grade students and free from unnecessary linguistic or cultural bias. Item-level analyses were then conducted using Rasch modeling, which estimated item difficulty based on student performance and allowed the researchers to examine how well individual items functioned within the measurement model.

## Results

0808 provided difficulty estimates for each comprehension item, allowing researchers to evaluate how items performed relative to student ability levels. The analysis supported the creation of a calibrated pool of items across the 20 forms.

Overall, the results indicated that the items covered a range of difficulty levels appropriate for second-grade readers. The Rasch modeling also allowed the researchers to examine the comparability of the different test forms. Items that appeared to function less effectively were retained for additional examination in future studies with larger samples. The pilot results demonstrated that the forms could be used to support a progress monitoring system for second grade reading comprehension, with item-level information helping guide future refinement of the measures.

**Table 22. Sample Item Statistics from Technical Report 0808**

Table 3

*Item Statistics, Entry Order, Gr2MC1*

Item Number	Raw Score	Count	Measure	Model Standard Error	Mean Square Outfit
1	100	118	-1.49	0.27	0.76
2	78	118	-0.28	0.21	0.75
3	71	118	0.03	0.21	1.09
4	53	118	0.77	0.20	1.16
5	68	118	0.15	0.21	1.00
6	59	118	0.52	0.20	1.19
7	79	118	-0.32	0.21	0.75
8	71	118	0.03	0.21	0.88
9	61	118	0.44	0.20	1.27
10	72	117	-0.04	0.21	1.02
11	88	117	-0.80	0.23	0.75
12	47	117	1.00	0.21	1.10

**Table 23. Illustrative Distractor Analysis from Technical Report 0808**

Table 4  
*Distractor Analysis, Gr2MC1*

Entry #	Data Code	Score Value	Count	%	Average Measure	S.E. Mean
1	B	0	5	4	-0.60	.52
	C	0	13	11	-0.40	.21
	A	1	100	85	0.87	.13
	Missing	**				
2	B	0	14	12	-0.37	.25
	A	0	26	22	-0.21	.14
	C	1	78	66	1.15	.15
	Missing	**				
3	C	0	20	17	-0.09	.19
	B	0	27	23	0.31	.24
	A	1	71	60	1.02	.16
	Missing	**				
4	A	0	30	25	-0.02	.18
	C	0	35	30	0.56	.14
	B	1	53	45	1.13	.21
	Missing	**				
5	A	0	25	21	-0.25	.13
	B	0	25	21	0.37	.30
	C	1	68	58	1.11	.15
	Missing	**				
6	C	0	18	15	-0.44	.19
	A	0	41	35	0.66	.19
	B	1	59	50	1.01	.17
	Missing	**				
7	C	0	10	8	-0.49	.32
	B	0	29	25	-0.23	.14
	A	1	79	67	1.15	.14
	Missing	**				
8	A	0	21	18	-0.22	.17
	B	0	26	22	0.04	.24
	C	1	71	60	1.16	.15
	Missing	**				
9	A	0	26	22	0.27	.15
	C	0	31	26	0.45	.26
	B	1	61	52	0.95	.18
	Missing	**				
10	B	0	18	15	-0.06	.22
	C	0	27	23	0.18	.16
	A	1	72	62	1.04	.17
	Missing	**	1	1	0.25	

**Table 24. Summary of Key Findings from Technical Report 0808**

Category	Summary
<b>Purpose</b>	Develop and pilot second-grade reading comprehension progress monitoring measures.
<b>Item Development</b>	Teachers created passages; comprehension questions written and grouped by cognitive skill and difficulty.
<b>Design</b>	20 alternate forms piloted using a common person / common item design.
<b>Analysis Method</b>	Rasch modeling used to estimate item difficulty and examine item functioning.
<b>Key Finding</b>	Items covered appropriate difficulty levels and supported development of comparable progress monitoring forms.

**Reference**

Alonzo, J., Liu, K., & Tindal, G. (2008). *Examining the technical adequacy of second-grade reading comprehension measures in a progress monitoring assessment system (Technical Report No. 0808)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Table 25. Summary of Key Findings from Technical Report 0041, 0807, and 0808**

The methods and results for TR 0041 are virtually identical for all the Categories listed below in two other Technical Reports: Alonzo, Liu, et al. (2008) and Alonzo and Tindal (2008b).

Category	Summary
<b>Sample</b>	The sample included a range of reading ability levels suitable for evaluating item performance and measurement adequacy.
<b>Assessment Forms</b>	Twenty forms were designed to be equivalent in structure and difficulty to support repeated administration within a progress monitoring framework.
<b>Analysis Method</b>	Item Response Theory (IRT) was used to report fit statistics, difficulty parameters, and discrimination indices as indices of technical adequacy.
<b>Items Analyzed</b>	Items spanned a range of difficulty levels intended to capture variation in student reading ability.
<b>Item Fit</b>	Most items exhibited acceptable fit statistics and functioned as intended within the IRT model.
<b>Overall Conclusion</b>	Results supported the technical adequacy of the reading comprehension measures for progress monitoring purposes. Most items functioned well, with only minor revisions needed to improve measurement precision and instructional utility.

**Reference**

Alonzo, J., Liu, K., & Tindal, G. (2007). *Examining the technical adequacy of reading comprehension measures in a progress monitoring assessment system (Technical Report 0041)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Alonzo, Liu, et al. (2008). *Examining the technical adequacy of second-grade reading comprehension measures in a progress monitoring assessment system (Technical Report # 0808)*. Eugene, OR, Behavioral Research and Teaching, University of Oregon.

Alonzo and Tindal (2008b). *Examining the technical adequacy of fifth-grade reading comprehension measures in a progress monitoring assessment system (Technical Report # 0807)*. Eugene, OR, Behavioral Research and Teaching, University of Oregon.

---

**Summary of Technical Report 1102:** The Development and Technical Adequacy of Seventh-Grade Reading Comprehension Measures in a Progress Monitoring Assessment (Park et al., 2011).

### **Methods**

This report describes the development and pilot testing of seventh grade reading comprehension measures designed for use in the easyCBM® progress monitoring and screening system. The goal was to create a set of technically sound comprehension assessments appropriate for middle school students and suitable for repeated administration.

Each assessment form consisted of an original narrative passage approximately 1,600–1,900 words long followed by 20 multiple-choice comprehension questions. The questions were designed to measure different levels of comprehension, including literal, inferential, and evaluative understanding. Each item included three response options: a correct answer, a near-correct distractor, and a clearly incorrect distractor. This structure was intended to capture gradations in student understanding while maintaining efficient scoring. Passages and items were developed specifically for the assessment system and reviewed to ensure appropriate reading level, clarity, and alignment with comprehension constructs. The measures were piloted with seventh-grade students using the easyCBM® online platform. Data from the pilot administrations were used to evaluate item performance and the comparability of the assessment forms.

To evaluate item functioning and test structure, the researchers applied Rasch measurement modeling. Rasch analysis allowed the authors to estimate item difficulty and examine whether individual items functioned consistently within the measurement model. This approach also allowed the researchers to determine whether items covered an appropriate range of difficulty for the target population and whether the forms could support progress monitoring.

### **Results**

The Rasch analyses produced difficulty estimates for each comprehension item and allowed the researchers to evaluate the functioning of the assessment forms. Overall, the items demonstrated an appropriate range of difficulty for seventh-grade students, suggesting that the measures could distinguish among different levels of reading comprehension ability.

The analyses also indicated that the test forms functioned comparably and that most items fit the Rasch model adequately. These results supported the technical adequacy of the assessment forms for use in a progress monitoring system. The calibrated item information provides a foundation for constructing multiple assessment forms capable of tracking student performance over time. Overall, the findings indicate that the seventh-grade comprehension measures developed for easyCBM provide a structured and psychometrically sound approach to assessing reading comprehension and monitoring student growth in middle school reading skills.

**Table 26. Example Item Statistics from Technical Report 1102***Item Statistics, Entry Order, Gr7MC1.*

Item Number	Raw Score	Count	Measure	Model Standard Error	Mean Square Outfit
1	61	64	-2.55	0.60	0.32
2	42	64	0.07	0.28	1.02
3	53	64	-1.00	0.35	0.80
4	10	64	2.76	0.36	1.25
5	58	64	-1.76	0.44	0.64
6	41	64	0.15	0.28	0.91
7	16	64	2.10	0.31	1.75
8	54	64	-1.12	0.36	0.78
9	43	64	-0.02	0.29	0.72
10	30	64	0.97	0.27	1.06
11	62	64	-2.99	0.73	0.28
12	51	64	-0.76	0.33	0.96
13	33	64	0.75	0.27	1.18
14	26	64	1.27	0.28	0.88
15	33	64	0.75	0.27	1.17
16	14	64	2.30	0.32	1.66
17	38	64	0.38	0.28	1.19
18	48	64	-0.46	0.31	0.93
19	50	64	-0.66	0.32	0.83
20	45	64	-0.19	0.29	1.03

**Table 27. Key Findings Summary from Technical Report 1102**

Category	Summary
<b>Purpose</b>	Develop and evaluate seventh grade reading comprehension measures for the easyCBM® progress monitoring system.
<b>Assessment Design</b>	Passages of 1,600–1,900 words with 20 multiple-choice comprehension items assessing literal, inferential, and evaluative understanding.
<b>Data Collection</b>	Pilot testing conducted with seventh-grade students using the easyCBM® online platform.
<b>Analysis Method</b>	Rasch modeling used to estimate item difficulty and evaluate item functioning.
<b>Key Results</b>	Items showed an appropriate range of difficulty and acceptable model fit, supporting use for comprehension assessment and monitoring.

**Reference**

Park, B. J., Alonzo, J., & Tindal, G. (2011). *The development and technical adequacy of seventh-grade reading comprehension measures in a progress monitoring assessment (Technical Report # 1102)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1209:** The Development of CBM Vocabulary Measures: Grade 2 (Alonzo, Anderson, et al., 2012a).

### **Methods**

This technical report describes the development and piloting of curriculum-based vocabulary measures designed for second-grade students within a response-to-intervention (RTI) framework. Approximately 1,200 students from schools across the United States participated, with data collected during spring benchmark administrations via an online assessment platform. Items were presented individually on computer screens with responses recorded automatically.

Item development was guided by principles of universal design for assessment to ensure accessibility for diverse learners, including students with disabilities and English language learners. Items were written to target specific grade-level vocabulary standards and underwent extensive review prior to piloting. Anchor items were embedded across forms to support calibration. Statistical analyses employed a one-parameter logistic (1PL) Rasch model using Winsteps software. Item difficulty estimates, fit statistics, and standard errors were examined for all items. Overfitting items were retained when distractor analyses confirmed appropriate functioning; no underfitting items were identified.

### **Results**

Rasch analysis was used to evaluate the psychometric properties of all items on the Gr2Vocab1 form. The mean item measure for this form was -1.85 logits, indicating that, on average, items were relatively easy for the sampled population of second-grade students. Item difficulty (measure) values ranged from -3.91 (Item 1) to -0.99 (Item 8), reflecting a moderate spread of difficulty across the form. All items were on the easier end of the logit scale, consistent with expectations for a Grade 2 vocabulary assessment administered in the spring. Item response counts ranged from 77 (Item 10) to 95 (Item 11), indicating adequate sample sizes for estimation across all items. Standard errors of measurement were generally small, ranging from 0.39 to 1.84, with the largest error associated with Item 1 (measure = -3.91), which was the most extreme in difficulty and had the lowest response count ( $n = 78$ ). Items with more central difficulty estimates tended to have smaller standard errors, consistent with Rasch modeling expectations.

Mean Square Outfit (MSQ Outfit) values were examined to evaluate model fit. Acceptable fit is typically defined as values between 0.50 and 1.50. Most items fell within this range. Item 2 demonstrated notable underfit with an MSQ Outfit of 0.07, and Items 9, 10, and 11 also showed low outfit values (0.16, 0.16, and 0.19 respectively), suggesting these items may have been overly predictable for students of varying ability levels. Despite these low outfit values, distractor analyses confirmed that all items were functioning appropriately and no items were flagged for removal. Item 3 had the highest MSQ Outfit value of 1.45, approaching but remaining within the acceptable threshold.

Discrimination (Disc) values ranged from 0.94 to 1.23, with all items exceeding the minimum acceptable threshold, indicating that each item reliably differentiated between students of higher and lower vocabulary ability. Items targeted a range of grade-level standards, including EL.03, EL.04, EL.05, and EL.06 reading standards related to vocabulary, contextual meaning, figurative language, synonyms, antonyms, homographs, and multiple-meaning words. The breadth of standards covered supports the content validity of the form. Overall, results supported the retention of all 12 items on the Gr2Vocab1 form without revision. These item-level findings were consistent across the other piloted forms, ultimately supporting the creation of ten progress monitoring forms and three benchmark forms with comparable difficulty levels and strong technical adequacy for use in instructional decision-making.

**Table 28. Example of Summary Statistics from Technical Report 1209**

Table 1a  
*Item Information and Statistics, Gr2Vocab1 (Mean Measure = -1.85)*

Item	Item Name	Standard	Measure	Count	Error	MSQ Outfit	Disc
1	Vocab_ite m_55410	EL.05.RE.11 Determine meanings of words using contextual and structural clues.	-3.91	78	1.84	1.00	1.00
2	Vocab_ite m_65641	EL.06.RE.12 Interpret figurative language, including similes, metaphors, and words with multiple meanings.	-3.00	93	1.04	0.07	1.15
3	Vocab_ite m_35734	EL.03.RE.15 Use sentence and word context to find the meaning of unknown words.	-1.39	94	0.46	1.45	1.07
4	Vocab_ite m_45641	EL.04.RE.11 Determine meanings of words using contextual and structural clues.	-1.37	89	0.50	0.21	1.23
5	Vocab_ite m_45736	EL.04.RE.13 Apply knowledge of synonyms, antonyms, homographs, and idioms to determine the meaning of words and phrases.	-1.23	87	0.49	1.18	0.99
6	Vocab_ite m_45644	EL.04.RE.12 Distinguish and interpret words with multiple meanings (i.e., quarter) by using context clues.	-1.19	89	0.42	1.21	1.06
7	Vocab_ite m_35274	EL.03.RE.14 Determine the meanings of words using knowledge of antonyms, synonyms, homophones, and homographs.	-1.08	94	0.39	0.91	0.94
8	Vocab_ite m_35151	EL.03.RE.15 Use sentence and word context to find the meaning of unknown words.	-0.99	93	0.45	1.05	1.08
9	Vocab_ite m_45645	EL.04.RE.12 Distinguish and interpret words with multiple meanings (i.e., quarter) by using context clues.	-2.41	89	0.74	0.16	1.13
10	Vocab_ite m_55650	EL.05.RE.13 Determine the meanings of figurative expressions, such as those in similes and metaphors.	-2.22	77	0.73	0.16	1.16
11	Vocab_ite m_35589	EL.03.RE.14 Determine the meanings of words using knowledge of antonyms, synonyms, homophones, and homographs.	-1.94	95	0.55	0.19	1.20
12	Vocab_ite m_35109	EL.03.RE.15 Use sentence and word context to find the meaning of unknown words.	-1.50	94	0.50	0.26	1.16

**Table 29. Key Findings Summary from Technical Report 1209**

Category	Summary
<b>Grade Level</b>	Grade 2
<b>Assessment Focus</b>	Vocabulary
<b>Statistical Model</b>	1PL Rasch
<b>Forms Developed</b>	10 progress monitoring and 3 benchmark forms
<b>Primary Outcome</b>	Reliable vocabulary CBMs for RTI use

**Reference**

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012). *The development of CBM vocabulary measures: Grade 2 (Technical Report 1209)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

**Summary of Technical Report 1210:** The Development of CBM Vocabulary Measures: Grade 3 (Alonzo, Anderson, et al., 2012b).

### Methods

Researchers at the University of Oregon developed vocabulary assessments for grades 2–8 as part of the easyCBM® online progress monitoring system. Item development followed Universal Design for Assessment (UDA) principles, prioritizing accessibility for English language learners and students with disabilities. Eight item writers with backgrounds in education and special education created approximately 800 items per grade level, guided by Oregon State Vocabulary Standards (EL.03.RE.14 and EL.03.RE.15). Items were multiple-choice with three answer options, designed to minimize cognitive load and linguistic complexity. After a thorough review by six researchers from June–August 2009, approximately 5,600 items were prepared for piloting. In spring 2011, roughly 1,200 third-grade students from schools across the United States completed pilot assessments delivered online using a common-item design. Each student was administered anchor items plus randomly selected grade-level items. Results were analyzed using a one-parameter logistic (1PL) Rasch model via Winsteps software, with Mean Square Outfit values (acceptable range: 0.50–1.50) and distractor analysis used to evaluate item quality.

### Results

Although no items were originally written specifically for grade 3, extremely easy items identified during piloting were repurposed for this grade. Of the items analyzed, 107 were over-fit (Mean Square Outfit: 0.08–0.49); 90 of these were retained based on satisfactory distractor functioning. An additional 26 under-fit items (Mean Square Outfit: 1.52–3.13) were all removed due to poor distractor performance. In total, 43 items were excluded because higher-ability students did not consistently select the correct answer. The retained items were organized into 13 third-grade vocabulary measures: 10 progress monitoring forms and 3 benchmark screener forms, each containing 20 items. All forms were calibrated to comparable difficulty, with mean measures ranging narrowly from –0.08 to –0.03, confirming strong form equivalence suitable for reliable progress monitoring.

**Table 30. Example Item Information and Statistics from Technical Report 1210**

Table 1a  
*Item Information and Statistics, Gr3Vocab1 (Mean Measure = -0.07)*

Item	Item Name	Standard	Measure	Count	Error	MSQ Outfit	Disc
1	Vocab_It m_35781	EL.03.RE.15 Use sentence and word context to find the meaning of unknown words.	-1.55	94	0.48	0.43	1.11
2	Vocab_It m_35734	EL.03.RE.15 Use sentence and word context to find the meaning of unknown words.	-1.39	94	0.46	1.45	1.07
3	Vocab_It m_35692	EL.03.RE.15 Use sentence and word context to find the meaning of unknown words.	-0.76	95	0.37	0.55	1.17
4	Vocab_It m_35014	EL.03.RE.15 Use sentence and word context to find the meaning of unknown words.	-0.73	94	0.36	0.70	1.12
5	Vocab_It m_35334	EL.03.RE.15 Use sentence and word context to find the meaning of unknown words.	-0.03	95	0.29	0.57	1.33
6	Vocab_It m_35415	EL.03.RE.14 Determine the meanings of words using knowledge of antonyms, synonyms, homophones, and homographs.	0.01	94	0.32	0.90	0.98
7	Vocab_It m_35776	EL.03.RE.15 Use sentence and word context to find the meaning of unknown words.	0.80	94	0.27	0.91	1.10
8	Vocab_It m_35868	EL.03.RE.15 Use sentence and word context to find the meaning of unknown words.	0.93	95	0.26	0.74	1.34
9	Vocab_It m_35508	EL.03.RE.15 Use sentence and word context to find the meaning of unknown words.	-1.21	94	0.44	0.25	1.24
10	Vocab_It m_35859	EL.03.RE.14 Determine the meanings of words using knowledge of antonyms, synonyms, homophones, and homographs.	-0.98	94	0.40	0.44	1.18

**Table 31. Typical Distractor Analysis Results from Technical Report 1210**

Table 2  
*Distractor Analysis Results*

Item	Answer choice	Score	# of students	Average measure	Standard error	Mean square outfit
v35001	A	1	3414	-4.18	0.02	1.97
v35001	B	0	666	-4.60	0.05	1.39
v35001	C	0	147	-5.06	0.12	1.07
v35002	A	1	88	-3.78	0.03	0.89
v35002	B	0	3	-4.06	0.10	0.91
v35002	C	0	3	-4.17	0.14	0.82
v35003	A	1	87	-3.89	0.04	1.02
v35003	B	0	2	-4.64	0.54	0.58
v35003	C	0	5	-4.51	0.23	0.63
v35004	A	1	80	-3.83	0.04	0.96
v35004	B	0	9	-4.31	0.17	0.77
v35004	C	0	4	-4.59	0.22	0.57
v35005	A	1	62	-3.83	0.04	0.95
v35005	B	0	7	-4.40	0.34	0.83
v35005	C	0	24	-4.15	0.09	0.88
v35006	A	1	92	-3.85	0.04	0.99
v35006	B	0	1	-8.69	0.00	0.00
v35006	C	0	1	-4.89	0.00	0.39
v35007	A	1	90	-3.89	0.04	1.02
v35007	B	0	3	-4.73	0.45	0.57
v35007	C	0	2	-4.57	0.61	0.64
v35008	A	1	80	-3.82	0.04	0.95
v35008	B	0	7	-4.87	0.31	0.50
v35008	C	0	6	-4.65	0.15	0.53
v35009	A	1	69	-3.87	0.05	1.04
v35009	B	0	19	-4.33	0.16	0.82
v35009	C	0	6	-4.67	0.17	0.53
v35010	A	1	90	-3.95	0.05	1.17
v35010	B	0	4	-5.37	0.74	0.43
v35010	C	0	1	-5.17	0.00	0.30
v35011	A	1	62	-3.84	0.04	0.96
v35011	B	0	8	-4.82	0.23	0.50
v35011	C	0	24	-4.05	0.10	0.99
v35012	A	1	86	-3.82	0.04	1.02
v35012	B	0	4	-5.93	0.62	0.23
v35012	C	0	4	-4.75	0.43	0.56
v35013	A	1	81	-3.80	0.03	0.89
v35013	B	0	6	-4.20	0.17	0.83
v35013	C	0	7	-4.37	0.19	0.74
v35014	A	1	81	-3.85	0.03	0.95
v35014	B	0	6	-5.89	0.75	0.51
v35014	C	0	7	-4.69	0.32	0.62

**Table 32. Key Findings Summary from Technical Report 1210**

Category	Summary
<b>Grade Level</b>	Grade 3
<b>Assessment Focus</b>	Vocabulary curriculum-based measurement
<b>Statistical Model</b>	1PL Rasch
<b>Forms Developed</b>	10 progress monitoring and 3 benchmark forms
<b>Primary Outcome</b>	Technically adequate vocabulary measures for RTI use

**Reference**

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012). *The development of CBM vocabulary measures: Grade 3 (Technical Report 1210)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

**Summary of Technical Report 1211:** The Development of CBM Vocabulary Measures: Grade 4 (Alonzo, Anderson, et al., 2012c).

## Methods

This study developed curriculum-based measurement (CBM) vocabulary assessments for students in grades 2–8, delivered via the easyCBM<sup>®</sup> online platform. Item development was grounded in the Oregon State Standards for Vocabulary, with eight item writers recruited from Oregon and California — most with Special Education backgrounds — producing approximately 800 items per grade level between October 2008 and May 2009. Writers were instructed to reduce cognitive complexity, use simple and direct language, and cover a range of student ability. Items used a multiple-choice format with three answer options, displayed one at a time on screen with the target word bolded in context. A team of six researchers reviewed all items for clarity, alignment, and formatting from June through August 2009, yielding approximately 5,600 items for piloting. The pilot was conducted in spring 2011 with approximately 1,200 students per grade across U.S. schools, using a common-item design with five anchor items to place all items on a shared scale. Data were analyzed using a one-parameter logistic (1PL) Rasch model via Winsteps software, evaluating each item's difficulty (Measure), Standard Error, and Mean Square Outfit statistic.

## Results

Although no items were written specifically for Grade 4, extremely easy items from adjacent grades were identified through piloting and organized into a Grade 4 item bank. Of the items analyzed, 122 were over-fit (Mean Square Outfit: 0.11–0.49; average 0.34) but retained because distractor analysis confirmed appropriate functioning: higher-ability students consistently selected the correct answer. Twenty-six under-fit items (Mean Square Outfit: 1.52–8.78; average 2.00) were removed because distractor analysis revealed reversed performance patterns. One additional item was removed because neither distractor was chosen by any student, resulting in 27 total items excluded. The retained items were used to construct 10 progress monitoring forms and 3 benchmark screener forms (13 forms total), each containing 20 items. Mean item difficulty (Measure) ranged narrowly from –0.09 to –0.05 across all forms, confirming comparable difficulty. Test information functions and characteristic curves confirmed that all forms performed consistently across the ability spectrum.

**Table 33. Sample Item Information and Statistics from Technical Report 1211**

Table 1a  
Item Information and Statistics, Gr4Vocab1 (Mean Measure = -0.08)

Item	Item Name	Standard	Measure	Count	Error	MSQ Outfit	Disc
1	Vocab_Itm_45181	EL.04.RE.11 Determine meanings of words using contextual and structural clues.	-2.05	89	0.64	0.12	1.21
2	Vocab_Itm_45010	EL.04.RE.12 Distinguish and interpret words with multiple meanings (i.e., quarter) by using context clues.	-1.70	89	0.54	0.26	1.16
3	Vocab_Itm_45692	EL.04.RE.12 Distinguish and interpret words with multiple meanings (i.e., quarter) by using context clues.	-0.91	89	0.39	0.82	1.07
4	Vocab_Itm_55187	EL.05.RE.11 Determine meanings of words using contextual and structural clues.	-0.86	78	0.47	0.52	1.10
5	Vocab_Itm_45694	EL.04.RE.12 Distinguish and interpret words with multiple meanings (i.e., quarter) by using context clues.	0.56	89	0.32	0.86	1.10
6	Vocab_Itm_45594	EL.04.RE.11 Determine meanings of words using contextual and structural clues.	0.65	89	0.28	0.79	1.11
7	Vocab_Itm_45047	EL.04.RE.11 Determine meanings of words using contextual and structural clues.	1.50	89	0.25	1.21	0.53
8	Vocab_Itm_45335	EL.04.RE.13 Apply knowledge of synonyms, antonyms, homographs, and idioms to determine the meaning of words and phrases.	1.66	89	0.25	0.96	1.06
9	Vocab_Itm_55144	EL.05.RE.12 Understand and explain frequently used synonyms, antonyms, and homographs.	-1.53	78	0.53	0.45	1.07
10	Vocab_Itm_45466	EL.04.RE.13 Apply knowledge of synonyms, antonyms, homographs, and idioms to determine the meaning of words and phrases.	-1.36	89	0.49	0.51	1.04

**Table 34. Key Findings Summary from Technical Report 1211**

Category	Summary
<b>Grade Level</b>	Grade 4
<b>Assessment Focus</b>	Vocabulary curriculum-based measurement
<b>Statistical Model</b>	1PL Rasch
<b>Forms Developed</b>	10 progress monitoring and 3 benchmark forms
<b>Primary Outcome</b>	Technically adequate vocabulary measures for RTI use

### Reference

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012). *The development of CBM vocabulary measures: Grade 4 (Technical Report 1211)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

---

**Summary of Technical Report 1212:** The Development of CBM Vocabulary Measures: Grade 5 (Alonzo, Anderson, et al., 2012d).

### Methods

Researchers at the University of Oregon developed vocabulary Curriculum-Based Measures (CBMs) for grades 2–8 to support Response to Intervention (RTI) screening and progress monitoring. Item development was guided by Universal Design for Assessment (UDA) principles, prioritizing accessibility for English language learners and students receiving special education services. Eight item writers from Oregon and California created approximately 800 items per grade level beginning in October 2008, with completion in May 2009. Items were multiple-choice with three answer options, aligned to Oregon State Vocabulary Standards, and designed to minimize cognitive and linguistic complexity. A six-member research team reviewed all items for clarity, alignment, and formatting from June through August 2009, resulting in approximately 5,600 items for piloting. In spring 2011, roughly 1,200 students per grade across the United States completed pilot assessments administered online via the easyCBM platform. A common-item design using anchor items allowed all items to be calibrated to a single scale. Data were analyzed using a one-parameter logistic (1PL) Rasch model with Winsteps software, examining Mean Square Outfit, Standard Error of Measure, item difficulty, and distractor functioning.

### Results

Although no items were originally written specifically for Grade 5, a subset of very easy items from adjacent grades was identified through piloting and organized into Grade 5 measures. Of the Grade 5 items analyzed, 135 were over-fit (Mean Square Outfit: 0.07–0.49; average 0.31), with 134 retained after distractor analysis confirmed appropriate functioning. An additional 31 items were under-fit (Mean Square Outfit: 1.51–4.95; average 2.17); all were removed because higher-ability students were not reliably selecting the correct answer. In total, 32 items were excluded from the Grade 5 item bank. The retained items were organized into 13 alternate forms: 10 progress monitoring forms and 3 benchmark screener forms, each containing 20 items. Mean difficulty measures across progress monitoring forms ranged from –0.24 to –0.19 (average –0.22), and benchmark forms ranged from –0.22 to –0.21 (average –0.22), indicating comparable and consistently accessible difficulty levels suitable for both general education students and those receiving intensive intervention.

**Table 35. Example of Item Information and Statistics from Technical Report 1212**

Table 1a  
*Item Information and Statistics, Gr5Vocab1 (Mean Measure = -0.23)*

Item	Item Name	Standard	Measure	Count	Error	MSQ Outfit	Disc
1	Vocab_It m_55232	EL.05.RE.12 Understand and explain frequently used synonyms, antonyms, and homographs.	-2.59	78	0.75	0.12	1.18
2	Vocab_It m_65229	EL.06.RE.11 Determine the meaning of unknown words or words with unusual meanings in informational and narrative text by using word, sentence, and paragraph clues.	-2.27	93	0.63	0.70	1.01
3	Vocab_It m_55867	EL.05.RE.12 Understand and explain frequently used synonyms, antonyms, and homographs.	-1.26	78	0.44	1.05	0.90
4	Vocab_It m_55017	EL.05.RE.13 Determine the meanings of figurative expressions, such as those in similes and metaphors.	-1.23	79	0.49	1.11	1.00
5	Vocab_It m_55594	EL.05.RE.12 Understand and explain frequently used synonyms, antonyms, and homographs.	0.54	78	0.30	1.07	0.82
6	Vocab_It m_55452	EL.05.RE.11 Determine meanings of words using contextual and structural clues.	0.58	78	0.29	0.85	1.06
7	Vocab_It m_55732	EL.05.RE.12 Understand and explain frequently used synonyms, antonyms, and homographs.	1.49	78	0.27	0.70	1.48
8	Vocab_It m_55694	EL.05.RE.13 Determine the meanings of figurative expressions, such as those in similes and metaphors.	1.66	78	0.26	0.94	1.19
9	Vocab_It m_65106	EL.06.RE.12 Interpret figurative language, including similes, metaphors, and words with multiple meanings.	-1.79	92	0.51	0.53	1.02
10	Vocab_It m_55463	EL.05.RE.12 Understand and explain frequently used synonyms, antonyms, and homographs.	-1.53	78	0.54	0.49	1.08

**Table 36. Key Findings Summary from Technical Report 1212**

Category	Summary
<b>Grade Level</b>	Grade 5
<b>Assessment Focus</b>	Vocabulary curriculum-based measurement
<b>Statistical Model</b>	1PL Rasch
<b>Forms Developed</b>	10 progress monitoring and 3 benchmark forms
<b>Primary Outcome</b>	Technically adequate vocabulary measures for RTI use

### Reference

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012). *The development of CBM vocabulary measures: Grade 5 (Technical Report 1212)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

**Summary of Technical Report 1213:** The Development of CBM Vocabulary Measures: Grade 6 (Alonzo, Anderson, et al., 2012).

## Methods

Participants included approximately 1,200 students drawn from schools across the United States. Data collection occurred during spring benchmark administrations conducted via computer-based delivery. Items were presented individually, and student responses were captured electronically, ensuring standardized administration conditions and consistent scoring procedures across sites.

Item development was guided by principles of universal design for assessment to reduce construct-irrelevant barriers related to language, formatting, and working memory demands. Items were aligned to grade-level vocabulary standards and reviewed extensively by trained researchers prior to piloting. Anchor items were embedded within forms to support calibration and comparability across administrations.

Statistical analyses were conducted using a one-parameter logistic Rasch model implemented in Winsteps. Item difficulty estimates, fit statistics, and standard errors were examined to evaluate item functioning. Distractor analyses were used to identify items with inadequate response patterns. Items exhibiting underfit or non-functioning distractors were removed from the item bank, while overfitting items were retained when distractor functioning was acceptable.

## Results

Results supported the creation of ten progress monitoring forms and three benchmark forms with comparable difficulty levels. Mean form measures indicated strong alignment across forms, supporting interpretability for instructional decision-making. Overall findings indicate that the Grade 6 vocabulary measures demonstrate strong technical adequacy and are appropriate for use in benchmarking and progress monitoring.

**Table 37. Example of Item Information and Statistics from Technical Report 1213**

Table 1a  
*Item Information and Statistics, Gr6Vocab1 (Mean Measure = -0.23)*

Item	Item Name	Standard	Measure	Count	Error	MSQ Outfit	Disc
1	Vocab_Itm_75188	EL.07.RE.11 Demonstrate understanding of idioms and comparisons, such as analogies, metaphors, and similes, in prose (informational and literary text) and poetry.	-2.49	69	1.02	0.14	1.13
2	Vocab_Itm_75870	EL.07.RE.12 Clarify word meanings through the use of definition, inference, example, restatement, or contrast.	-2.21	68	0.76	0.45	1.04
3	Vocab_Itm_75871	EL.07.RE.12 Clarify word meanings through the use of definition, inference, example, restatement, or contrast.	-1.05	69	0.47	0.58	1.10
4	Vocab_Itm_65334	EL.06.RE.12 Interpret figurative language, including similes, metaphors, and words with multiple meanings.	-1.01	94	0.44	0.23	1.28
5	Vocab_Itm_65407	EL.06.RE.11 Determine the meaning of unknown words or words with unusual meanings in informational and narrative text by using word, sentence, and paragraph clues.	0.33	92	0.30	0.72	1.19
6	Vocab_Itm_65605	EL.06.RE.12 Interpret figurative language, including similes, metaphors, and words with multiple meanings.	0.38	92	0.34	0.43	1.42
7	Vocab_Itm_65147	EL.06.RE.12 Interpret figurative language, including similes, metaphors, and words with multiple meanings.	1.40	92	0.27	0.81	1.27
8	Vocab_Itm_65634	EL.06.RE.11 Determine the meaning of unknown words or words with unusual meanings in informational and narrative text by using word, sentence, and paragraph clues.	1.59	95	0.26	0.80	1.31
9	Vocab_Itm_65106	EL.06.RE.12 Interpret figurative language, including similes, metaphors, and words with multiple meanings.	-1.79	92	0.51	0.53	1.02
10	Vocab_Itm_75680	EL.07.RE.10 Determine meanings of words using contextual and structural clues.	-1.55	68	0.57	0.44	1.03

**Table 38. Key Findings Summary from Technical Report 1213**

Category	Summary
<b>Grade Level</b>	Grade 6
<b>Assessment Focus</b>	Vocabulary curriculum-based measurement
<b>Statistical Model</b>	1PL Rasch
<b>Forms Developed</b>	10 progress monitoring and 3 benchmark forms
<b>Primary Outcome</b>	Technically adequate vocabulary measures for RTI use

### Reference

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012). *The development of CBM vocabulary measures: Grade 6 (Technical Report 1213)*. Eugene OR: University of Oregon, Behavioral Research and Teaching.

---

**Summary of Technical Report 1214:** The Development of CBM Vocabulary Measures: Grade 7 (Alonzo, Anderson, et al., 2012e).

This technical report documents the development and pilot evaluation of curriculum-based measurement vocabulary assessments designed for use with students in Grade 7. The measures were developed as part of the easyCBM® assessment system to support both benchmarking and progress monitoring within a response-to-intervention framework. Emphasis was placed on producing brief, repeatable measures that were sensitive to growth while remaining accessible to students with a wide range of ability levels.

### Methods

Participants included approximately 1,200 students drawn from schools across the United States. Data collection occurred during spring benchmark administrations conducted via computer-based delivery. Items were presented individually, and student responses were captured electronically, ensuring standardized administration conditions and consistent scoring procedures across sites.

Item development was guided by principles of universal design for assessment to reduce construct-irrelevant barriers related to language, formatting, and working memory demands. Items were aligned to grade-level vocabulary standards and reviewed extensively by trained researchers prior to piloting. Anchor items were embedded within forms to support calibration and comparability across administrations.

Statistical analyses were conducted using a one-parameter logistic Rasch model implemented in Winsteps. Item difficulty estimates, fit statistics, and standard errors were examined to evaluate item functioning. Distractor analyses were used to identify items with inadequate response patterns. Items exhibiting underfit or non-functioning distractors were removed from the item bank, while overfitting items were retained when distractor functioning was acceptable.

### Results

Results supported the creation of ten progress monitoring forms and three benchmark forms with comparable difficulty levels. Mean form measures indicated strong alignment across forms, supporting interpretability for instructional decision-making. Overall findings indicate that the Grade 7 vocabulary measures demonstrate strong technical adequacy and are appropriate for use in benchmarking and progress monitoring.

**Table 39. Example of Item Information and Statistics from Technical Report 1214**

Table 1a  
*Item Information and Statistics, Gr7Vocab1 (Mean Measure = -0.16)*

Item	Item Name	Standard	Measure	Count	Error	MSQ Outfit	Disc
1	Vocab_It m_75097	EL.07.RE.10 Determine meanings of words using contextual and structural clues.	-2.03	68	0.63	0.24	1.17
2	Vocab_It m_75766	EL.07.RE.10 Determine meanings of words using contextual and structural clues.	-1.89	69	0.75	0.18	1.16
3	Vocab_It m_75557	EL.07.RE.12 Clarify word meanings through the use of definition, inference, example, restatement, or contrast.	-1.02	68	0.50	1.37	0.87
4	Vocab_It m_85139	EL.08.RE.10 Determine meanings of words using contextual and structural clues.	-0.96	58	0.61	0.38	1.13
5	Vocab_It m_75815	EL.07.RE.10 Determine meanings of words using contextual and structural clues.	0.35	69	0.33	0.79	1.09
6	Vocab_It m_75013	EL.07.RE.11 Demonstrate understanding of idioms and comparisons, such as analogies, metaphors, and similes, in prose (informational and literary text) and poetry.	0.49	69	0.34	1.20	0.91
7	Vocab_It m_75510	EL.07.RE.12 Clarify word meanings through the use of definition, inference, example, restatement, or contrast.	1.54	69	0.30	1.27	0.68
8	Vocab_It m_75691	EL.07.RE.12 Clarify word meanings through the use of definition, inference, example, restatement, or contrast.	1.68	68	0.29	1.19	0.39
9	Vocab_It m_75151	EL.07.RE.12 Clarify word meanings through the use of definition, inference, example, restatement, or contrast.	-1.50	69	0.56	0.45	1.10
10	Vocab_It m_85098	EL.08.RE.11 Analyze idioms and comparisons, such as analogies, metaphors, and similes, to infer the literal and figurative meanings of phrases.	-1.39	58	0.76	0.17	1.23

**Table 40. Key Findings Summary from Technical Report 1214**

Category	Summary
<b>Grade Level</b>	Grade 7
<b>Assessment Focus</b>	Vocabulary curriculum-based measurement
<b>Statistical Model</b>	1PL Rasch
<b>Forms Developed</b>	10 progress monitoring and 3 benchmark forms
<b>Primary Outcome</b>	Technically adequate vocabulary measures for RTI use

**Reference**

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012). *The development of CBM vocabulary measures: Grade 7 (Technical Report 1214)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

**Summary of Technical Report 1215:** The Development of CBM Vocabulary Measures: Grade 8 (Alonzo, Anderson, et al., 2012f).

This technical report documents the development and pilot evaluation of curriculum-based measurement vocabulary assessments designed for use with students in Grade 8. The measures were developed as part of the easyCBM® assessment system to support both benchmarking and progress monitoring within a response-to-intervention framework. Emphasis was placed on producing brief, repeatable measures that were sensitive to growth while remaining accessible to students with a wide range of ability levels.

## Methods

Participants included approximately 1,200 students drawn from schools across the United States. Data collection occurred during spring benchmark administrations conducted via computer-based delivery. Items were presented individually, and student responses were captured electronically, ensuring standardized administration conditions and consistent scoring procedures across sites.

Item development was guided by principles of universal design for assessment to reduce construct-irrelevant barriers related to language, formatting, and working memory demands. Items were aligned to grade-level vocabulary standards and reviewed extensively by trained researchers prior to piloting. Anchor items were embedded within forms to support calibration and comparability across administrations. Statistical analyses were conducted using a one-parameter logistic Rasch model implemented in Winsteps. Item difficulty estimates, fit statistics, and standard errors were examined to evaluate item functioning. Distractor analyses were used to identify items with inadequate response patterns. Items exhibiting underfit or non-functioning distractors were removed from the item bank, while overfitting items were retained when distractor functioning was acceptable.

## Results

Results supported the creation of 10 progress monitoring forms and 3 benchmark forms with comparable difficulty levels. Mean form measures indicated strong alignment across forms, supporting interpretability for instructional decision-making. Overall findings indicate that the Grade 8 vocabulary measures demonstrate strong technical adequacy and are appropriate for benchmarking and progress monitoring.

**Table 41. Example of Item Information and Statistics from Technical Report 1215**

Table 1a  
Item Information and Statistics, Gr8Vocab1 (Mean Measure = -0.15)

Item	Item Name	Standard	Measure	Count	Error	MSQ Outfit	Disc
1	Vocab_ite m_85372	EL.08.RE.11 Analyze idioms and comparisons, such as analogies, metaphors, and similes, to infer the literal and figurative meanings of phrases.	-2.49	58	1.03	0.16	1.13
2	Vocab_ite m_85640	EL.08.RE.11 Analyze idioms and comparisons, such as analogies, metaphors, and similes, to infer the literal and figurative meanings of phrases.	-2.03	58	1.02	0.32	1.06
3	Vocab_ite m_85049	EL.08.RE.10 Determine meanings of words using contextual and structural clues.	-0.94	58	0.52	1.30	0.89
4	Vocab_ite m_85369	EL.08.RE.11 Analyze idioms and comparisons, such as analogies, metaphors, and similes, to infer the literal and figurative meanings of phrases.	-0.88	58	0.58	0.22	1.30
5	Vocab_ite m_85638	EL.08.RE.11 Analyze idioms and comparisons, such as analogies, metaphors, and similes, to infer the literal and figurative meanings of phrases.	0.04	58	0.41	1.43	1.01
6	Vocab_ite m_85453	EL.08.RE.10 Determine meanings of words using contextual and structural clues.	0.16	58	0.42	0.67	1.15
7	Vocab_ite m_85417	EL.08.RE.11 Analyze idioms and comparisons, such as analogies, metaphors, and similes, to infer the literal and figurative meanings of phrases.	0.59	58	0.43	1.07	0.86
8	Vocab_ite m_85238	EL.08.RE.11 Analyze idioms and comparisons, such as analogies, metaphors, and similes, to infer the literal and figurative meanings of phrases.	0.73	58	0.37	1.17	0.84
9	Vocab_ite m_85150	EL.08.RE.12 Verify the meaning of a word in its context, even when its meaning is not directly stated, through the use of definition, restatement, example, comparison, or contrast.	-1.61	57	0.76	0.19	1.18
10	Vocab_ite m_85871	EL.08.RE.12 Verify the meaning of a word in its context, even when its meaning is not directly stated, through the use of definition, restatement, example, comparison, or contrast.	-1.22	58	0.59	0.23	1.27

**Table 42. Key Findings Summary from Technical Report 1215**

Category	Summary
<b>Grade Level</b>	Grade 8
<b>Assessment Focus</b>	Vocabulary curriculum-based measurement
<b>Statistical Model</b>	1PL Rasch
<b>Forms Developed</b>	10 progress monitoring and 3 benchmark forms
<b>Primary Outcome</b>	Technically adequate vocabulary measures for RTI use

## Reference

Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012). *The development of CBM vocabulary measures: Grade 8 (Technical Report 1215)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

## Summary of Technical Report 1221: The Development of the easyCBM<sup>®</sup> CCSS Reading Assessments: Grade 3 (Alonzo, Park, et al., 2012a).

This technical report documents the development and piloting of the easyCBM<sup>®</sup> Common Core State Standards Reading Assessment for Grade 3. The assessment was designed to support screening and progress monitoring for students at risk for reading difficulties and low-performing students.

## Methods

Item development followed principles of Universal Design for Assessment and alignment with CCSS reading standards. Educators with experience teaching struggling readers authored literary, informational, and task-based passages, each paired with multiple-choice comprehension questions. All items underwent expert review for bias, accessibility, and construct relevance. Pilot testing occurred online across multiple U.S. states using convenience samples. Item Response Theory testlet models were used to analyze item difficulty, discrimination, and fit.

## Results

Results showed that Read to Perform a Task items were generally less difficult, while literary and informational texts exhibited wider variability. Overall findings support the technical adequacy of the measures for progress monitoring, though subtest-level interpretation is not recommended.

**Table 43. Typical Results of IRT Analysis from Technical Report 1221**

Table 2  
Results of IRT Analysis, Mean Measure of Progress Monitoring Measure 3\_2 = 0.30

Test Form	Item #	Reading Type	Standard	Item Name	Measure	Count	Error	Mean Square Outfit	Discrimination
2	1	Literary Text: Demonstrate General Understanding	Identify the speaker or narrator in a selection. Distinguish the order of events or a specific event from a sequence of events. Determine significant events from the story.	35756 to 35760	-0.71	61	0.35	1.25	0.95
				Determine what characters are like by what they say or do and by how the author or illustrator portrays them. Predict probable future outcomes or actions. Determine and discuss the underlying theme or author's message in literary text. Recognize cause-and-effect relationships in literary text.					
2	2	Literary Text: Develop an Interpretation	Recall major points in the text and make predictions about forthcoming information. Distinguish cause-and-effect and fact and opinion.	35851 to 35855	1.68	65	0.35	0.78	0.99
				Recall major points in the text and make predictions about forthcoming information. Distinguish cause-and-effect and fact and opinion.					
2	3	Informational Text: Develop an Interpretation	Recall major points in the text and make predictions about forthcoming information. Distinguish cause-and-effect and fact and opinion.	35256 to 35260	0.07	65	0.40	0.43	1.50
				Recall major points in the text and make predictions about forthcoming information. Distinguish cause-and-effect and fact and opinion.					
2	4	Informational Text: Develop an Interpretation	Recall major points in the text and make predictions about forthcoming information. Distinguish cause-and-effect and fact and opinion.	35706 to 35710	1.18	64	0.26	0.71	1.23
				Follow simple multiple-step written instructions (e.g., how to assemble a product or play a board game).					
2	5	Read to Perform a Task	Follow simple multiple-step written instructions (e.g., how to assemble a product or play a board game).	35516 to 35520	-0.74	66	0.24	1.87	0.10

**Table 44. Key Findings Summary from Technical Report 1221**

Category	Summary
<b>Sample</b>	Grade 3 students from multiple U.S. states; convenience sample
<b>Assessment Types</b>	Literary Text, Informational Text, Read to Perform a Task
<b>Analysis Method</b>	Item Response Theory (testlet-based modeling)
<b>Key Finding</b>	Read to Perform a Task items were least difficult
<b>Conclusion</b>	Measures demonstrate strong technical adequacy for monitoring growth

### Reference

Alonzo, J., Park, B. J., & Tindal, G. (2012). *The development of the easyCBM® CCSS reading assessments: Grade 3 (Technical Report 1221)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report No 1222:** The Development of the easyCBM® CCSS Reading Assessments: Grade 4 (Alonzo, Park, et al., 2012b).

This report describes the development and validation of the Grade 4 easyCBM® CCSS Reading Assessment, created to measure reading comprehension for screening and progress monitoring within an RTI framework. The assessment targets students who struggle with reading comprehension.

### Methods

Assessment items were developed by trained educators and aligned to CCSS standards, incorporating literary text, informational text, and task-based reading. Universal Design principles guided item construction to ensure accessibility for diverse learners, including students with disabilities. Pilot data were collected online and analyzed using Item Response Theory.

### Results

Results indicated substantial variability in difficulty across passages, with no consistent pattern indicating a single most challenging subtest. Findings support the use of the assessment as a unidimensional measure of reading comprehension, suitable for monitoring student growth over time.

**Table 45. Typical Results of IRT Analysis for Technical Report 1222**

Table 1  
Results of IRT Analysis, Mean Measure of Progress Monitoring Measure 4\_1 = 0.52

Test Form	Item #	Reading Type	Standard	Item Name	Measure	Count	Error	Mean Square Outfit	Discrimination
1	1	Literary Text: Demonstrate General Understanding	Identify and/or summarize sequence of events, main ideas, and supporting details in literary selections. Identify the main problem or conflict of the plot, and explain how it is resolved.	45716 to 45720	-2.66	58	0.49	9.90	0.71
1	2	Literary Text: Demonstrate General Understanding	Identify and/or summarize sequence of events, main ideas, and supporting details in literary selections. Identify the main problem or conflict of the plot, and explain how it is resolved.	45846 to 45850	1.28	61	0.41	0.41	1.56
1	3	Informational Text: Demonstrate General Understanding Informational Text: Develop an Interpretation	Identify and/or summarize sequence of events, main ideas, facts, supporting details, and opinions in informational and practical selections. Make and confirm predictions about text by using prior knowledge and ideas presented in the text itself, including illustrations, titles, topic sentences, and important words. Draw inferences or conclusions about an author's meaning supported by facts and events from the text.	45071 to 45075	-0.52	60	0.29	0.93	0.58

**Table 46. Key Findings Summary from Technical Report 1222**

Category	Summary
<b>Sample</b>	Grade 4 students across multiple U.S. states
<b>Assessment Types</b>	Literary Text, Informational Text, Read to Perform a Task
<b>Analysis Method</b>	Item Response Theory with testlet structure
<b>Key Finding</b>	Wide variability in item difficulty across subtests
<b>Conclusion</b>	Assessment supports valid, reliable progress monitoring

### Reference

Alonzo, J., Park, B. J., & Tindal, G. (2012). *The development of the easyCBM<sup>®</sup> CCSS reading assessments: Grade 4 (Technical Report 1222)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1223:** The Development of the easyCBM<sup>®</sup> CCSS Reading Assessments: Grade 5 (Alonzo, Park, et al., 2012c).

This technical report documents the development and pilot evaluation of the easyCBM<sup>®</sup> Common Core State Standards–aligned reading comprehension assessments for Grade 5. The assessments were designed to support both universal screening and progress monitoring, with particular emphasis on students who demonstrate persistent difficulty in reading. The measures target comprehension across literary text, informational text, and read-to-perform-a-task formats in alignment with grade-level CCSS expectations.

### Methods

Participants included students from multiple states who participated through convenience sampling via the online easyCBM<sup>®</sup> assessment platform. Item piloting occurred during fall administrations, with students completing subsets of items to support large-scale calibration while minimizing individual testing burden. Administration procedures were standardized and computer-based, allowing automated capture of student responses. Item development followed principles of Universal Design for Assessment to reduce construct-irrelevant variance. Educators with experience teaching struggling readers authored passages and associated multiple-choice items, which were reviewed for bias, accessibility, and alignment prior to piloting. Each passage and its associated items were treated as testlets to account for local item dependence. Data analyses were conducted using Item Response Theory testlet models. Item difficulty, fit statistics, and discrimination indices were examined to evaluate psychometric functioning. Results indicated that most items demonstrated acceptable fit and contributed to well-targeted forms. Ten progress monitoring forms and three benchmark forms were assembled with comparable difficulty levels to support growth.

### Results

Overall findings indicate that the Grade 5 easyCBM<sup>®</sup> CCSS reading assessments demonstrate strong technical adequacy. The measures provide reliable and interpretable information for identifying students at risk and monitoring growth over time within data-based decision-making frameworks.

**Table 47. Key Findings Summary from Technical Report 1223**

Category	Summary
<b>Grade Level</b>	Grade 5
<b>Assessment Focus</b>	CCSS-aligned reading comprehension
<b>Sample</b>	Multi-state convenience sample
<b>Statistical Model</b>	IRT testlet modeling
<b>Forms Developed</b>	10 progress monitoring, 3 benchmarks
<b>Intended Use</b>	Screening and progress monitoring

## Reference

Alonzo, J., Park, B. J., & Tindal, G. (2012). *The development of the easyCBM® CCSS reading assessments: Grade 5 (Technical Report 1223)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1224:** The Development of the easyCBM® CCSS Reading Assessments: Grade 6 (Alonzo, Park, et al., 2012d).

## Methods

Researchers developed **vocabulary** Curriculum-Based Measures (CBMs) for use with students in grades 2–8 as part of the easyCBM® online assessment system. Item development was guided by Oregon State Vocabulary Standards and principles of Universal Design for Assessment (UDA), with a focus on reducing cognitive and linguistic complexity to ensure accessibility for students with disabilities, English language learners, and those receiving special education services. Eight qualified item writers created approximately 800 items per grade level between October 2008 and May 2009, producing three-choice multiple-choice items contextualized in brief sentences with bolded target words. A team of six researchers reviewed all items for clarity, alignment, and formatting from June through August 2009, yielding approximately 5,600 items for piloting. Items were administered online to approximately 1,200 students per grade from schools across the United States during spring 2011, using a common-item design with anchor items to enable calibration across forms. Data were analyzed using a one-parameter logistic (1PL) Rasch model in Winsteps, examining Mean Square Outfit, Standard Error of Measure, item difficulty, and distractor functioning.

## Results

Although no items were originally written for Grade 6, a set of items identified as extremely easy during piloting were organized into sixth-grade vocabulary measures. Of the Grade 6 items analyzed, 146 were over-fit (Mean Square Outfit: 0.07–0.49; average 0.33), of which 145 were retained based on appropriate distractor functioning. A total of 41 items were under-fit (Mean Square Outfit: 1.51–9.64; average 2.32) and removed due to poor distractor performance, along with one additional item removed on content review grounds, yielding 43 total removed items. Retained items were organized into 13 forms: ten progress monitoring forms and three benchmark screener forms, each comprising 20 items. Mean item difficulty measures across progress monitoring forms ranged from –0.23 to –0.19 (average –0.21), and benchmark forms ranged from –0.22 to –0.21 (average –0.21), indicating comparable difficulty across all alternate forms.

**Table 48. Key Findings Summary from Technical Report 1224**

Category	Summary
<b>Grade Level</b>	Grade 6
<b>Assessment Focus</b>	CCSS-aligned reading comprehension
<b>Sample</b>	Multi-state convenience sample
<b>Statistical Model</b>	IRT testlet modeling
<b>Forms Developed</b>	10 progress monitoring, 3 benchmarks
<b>Intended Use</b>	Screening and progress monitoring

## Reference

Alonzo, J., Park, B. J., & Tindal, G. (2012). *The development of the easyCBM® CCSS reading assessments: Grade 6 (Technical Report 1224)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1225:** The Development of the easyCBM® CCSS Reading Assessments: Grade 7 (Alonzo, Park, et al., 2012).

### Methods

This study developed reading comprehension assessments for 7th-grade students aligned to the Common Core State Standards (CCSS), with particular focus on persistently low-performing students (approximately 2% of the student population) who do not respond to interventions. The assessment team applied Universal Design for Assessment principles, using the NCEO guidelines and the Test Accessibility and Modification Inventory (TAMI) to maximize fairness and accessibility. Item writers — educators experienced with struggling readers — independently created passages and graphics across three sub-test types: Short Literary Text, Informational Text, and Read to Perform a Task. Each passage was paired with five multiple-choice questions targeting specific CCSS reading standards, with one correct answer and two plausible distractors. A research team of 3–5 members reviewed all 900 items per grade for bias, age-appropriateness, and Universal Design compliance. Items were piloted online in November–December 2011 with convenience samples of students from eight states, randomly assigned to sub-test sets of 25 items to control for class effects and ensure adequate scaling data.

### Results

Piloted data were analyzed using Item Response Theory (IRT), treating each passage and its five questions as a testlet. Results yielded 10 progress monitoring forms and 3 benchmark forms (Fall, Winter, Spring), each comprising two Literary Text sub-tests, two Informational Text sub-tests, and one Read to Perform a Task sub-test. IRT measures (item difficulty) for the Grade 7 progress monitoring forms ranged widely: Literary Text from -1.63 to 2.95, Informational Text from -1.66 to 5.89, and Read to Perform a Task from -0.91 to 3.73. Benchmark form means were approximately 1.14 (Fall), 1.20 (Winter), and 1.17 (Spring). No consistent pattern emerged indicating which sub-test type was most difficult. Discrimination values were generally acceptable, though a small number of items showed poor fit and were excluded (Table 14). Importantly, the authors caution that sub-test scores should not be used to draw conclusions about student mastery of specific text types, as construct validity studies confirm all three sub-tests measure a single underlying reading construct.

**Table 49. Key Findings Summary from Technical Report 1225**

Category	Summary
<b>Grade Level</b>	Grade 7
<b>Assessment Focus</b>	CCSS-aligned reading comprehension
<b>Sample</b>	Multi-state convenience sample
<b>Statistical Model</b>	IRT testlet modeling
<b>Forms Developed</b>	10 progress monitoring, 3 benchmarks
<b>Intended Use</b>	Screening and progress monitoring

### Reference

Alonzo, J., Park, B. J., & Tindal, G. (2012). *The development of the easyCBM® CCSS reading assessments: Grade 7 (Technical Report 1225)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1226:** The Development of the easyCBM® CCSS Reading Assessments: Grade 8 (Alonzo, Park, et al., 2012e).

### Methods

This study developed and piloted Grade 8 reading comprehension assessments aligned to the Common Core State Standards (CCSS), designed primarily for persistently low-performing students — a subgroup representing roughly 2% of the student population who do not respond consistently to interventions. Assessment items were authored by experienced educators trained in Universal Design for Assessment and CCSS content standards. Item writers independently produced literary passages, informational texts, and graphics-based prompts, each accompanied by five multiple-choice questions targeting specific grade-level reading standards. A graphic artist rendered all visual materials for computer-based delivery. All items — totaling 900 across grades 3–8 — were reviewed by 3–5 trained researchers for bias, grade-level appropriateness, and universal design compliance before piloting. Items were piloted online in November–December 2011 using convenience samples from Oregon, Washington, Montana, Florida, Texas, Illinois, California, and Wisconsin. Each student completed 25 items across five randomly assigned sub-tests to reduce class-level bias. Data were analyzed using Item Response Theory (IRT), treating each passage and its five questions as a testlet.

### Results

IRT analyses produced 10 progress monitoring forms and 3 benchmark forms (Fall, Winter, Spring), each containing two Literary Text sub-tests, two Informational Text sub-tests, and one Read to Perform a Task sub-test. Item difficulty (IRT Measure) varied widely across and within sub-test types. For progress monitoring forms, Read to Perform a Task difficulty ranged from -0.71 to 2.08; Literary Text ranged from -3.84 to 3.08; and Informational Text ranged from -4.84 to 3.63. Benchmark assessments showed similar variability. No consistent pattern emerged indicating which sub-test type was most challenging — difficulty appeared to depend more on the specific passage than on text category. The authors caution that sub-test scores should not be interpreted as separate indicators of skill by text type, as construct validity studies confirm all three sub-tests measure a single underlying reading construct. A small number of items were excluded from final forms due to extreme difficulty values or poor psychometric fit.

**Table 50. Key Findings Summary from Technical Report 1226**

Category	Summary
<b>Grade Level</b>	Grade 8
<b>Assessment Focus</b>	CCSS-aligned reading comprehension
<b>Sample</b>	Multi-state convenience sample
<b>Statistical Model</b>	IRT testlet modeling
<b>Forms Developed</b>	Benchmark and progress monitoring forms
<b>Intended Use</b>	Screening and progress monitoring

### Reference

Alonzo, J., Park, B. J., & Tindal, G. (2012). *The development of the easyCBM® CCSS reading assessments: Grade 8 (Technical Report 1226)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

**Summary of Technical Report 1110:** Progress Monitoring Instrument Development: Silent Reading Fluency, Vocabulary, and Reading Comprehension (Nese et al., 2011).

This technical report describes the development of computer-based curriculum-based measurement (CBM) instruments designed to assess silent reading fluency, vocabulary, and reading comprehension in elementary students. The study focused on creating measures that could be administered digitally, addressing limitations associated with traditional oral reading fluency assessments that require individual administration.

### Methods

A team of researchers and elementary teachers developed a series of experimental reading measures for Grades 3 through 5. Five types of instruments were created: silent reading fluency sentences, silent reading fluency maze passages, vocabulary maze passages, vocabulary sentence items, and reading comprehension passages. Each measure included multiple alternate forms to support progress monitoring. Items were constructed using grade-appropriate vocabulary and readability targets determined by the Flesch-Kincaid formula. The measures were designed for computer delivery, allowing automated capture of student responses and timing data. For fluency tasks, the system recorded reading time and converted it into words-per-minute estimates. Vocabulary and comprehension measures focused on accuracy of responses to multiple-choice items embedded in passages or sentences. All items were reviewed by a research team to ensure clarity, appropriate difficulty, and absence of bias before being placed into the computer interface.

### Results

The development process produced multiple parallel forms of computer-administered reading assessments across the three reading domains. The resulting instruments generated automated data on reading speed, accuracy, and item responses, demonstrating the feasibility of collecting progress-monitoring data through computer-based testing. The measures were designed to maintain student engagement while minimizing cognitive load from distracters so that reading fluency and vocabulary knowledge could be captured efficiently. The report concludes that computerized CBM instruments could improve the efficiency of classroom progress monitoring by allowing groups of students to complete assessments simultaneously while performance data are automatically recorded.

### Reference

Nese, J. F. T., Anderson, D., Hoelscher, K., Tindal, G., & Alonzo, J. (2011). *Progress monitoring instrument development: Silent reading fluency, vocabulary, and reading comprehension (Technical Report # 1110)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

---

**Summary of Technical Report 1301:** The Development of easyCBM® Spanish Literacy Assessments for Use in Grades K–2 (Alonzo et al., 2013).

Technical Report 1301 documents the development and evaluation of the easyCBM® **Spanish literacy assessments** for use with students in kindergarten through Grade 2. The purpose of the study was to identify early literacy measures appropriate for students receiving instruction in Spanish, recognizing important linguistic differences between Spanish and English that influence literacy development and assessment. The assessments were intended to support progress monitoring and instructional decision making within Spanish literacy programs.

### Methods

Two initial item development studies were conducted with distinct student populations. The first included native English speakers enrolled in a Spanish immersion program, while the second involved native Spanish speakers enrolled in a dual-language bilingual program. Across both studies, students received all literacy instruction in Spanish. Trained researchers individually administered multiple early literacy measures, including letter sounds, syllable sounds, phoneme segmenting, syllable segmenting, word reading fluency, and sentence reading fluency, following standardized protocols.

Analyses focused on examining relationships between early literacy measures and reading fluency outcomes. Correlational analyses and stepwise linear regression models were used to identify which measures best predicted word and sentence reading performance. Across studies and grade levels, performance on syllable-level measures, particularly syllable sounds and syllable reading, consistently emerged as the strongest predictors of reading fluency. Phoneme-level measures showed weaker and less consistent relationships with reading outcomes, especially for younger students.

A subsequent reliability study examined alternate form reliability across kindergarten, Grade 1, and Grade 2. Data were collected from native Spanish-speaking students using multiple forms of each measure. Classical test theory analyses demonstrated very strong correlations among alternate forms of syllable reading, word reading fluency, and sentence reading fluency measures. Regression analyses further indicated that word reading fluency accounted for substantial variance in sentence reading fluency, particularly at Grade 2.

## Results

Overall findings support the technical adequacy of the easyCBM® Spanish literacy assessments. Results underscore the importance of syllable-level measures in Spanish literacy development and provide empirical support for assessment systems aligned with the linguistic structure of Spanish for early readers.

**Table 51. Example of Descriptive Statistics from Report 1301**

Table 2

*Descriptive Statistics for Study 1 and Study 2, Grade 1*

Measure	Study 1			Study 2		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Letter Sounds	48	14.00	7.31	35	16.06	9.40
Syllable Sounds	48	20.44	7.89	35	24.94	9.63
Phoneme Segmenting	48	22.35	11.17	35	22.29	9.23
Syllable Segmenting	48	17.31	5.93	35	19.34	4.33
Word Reading	48	14.17	5.17	35	22.94	7.97
Sentence Reading	-NA-	-NA-	-NA-	34	24.12	11.67

**Table 52. Typical Set of Relations Among Measures from Report 1301**

Table 4

*Correlations Among the Grade 1 Measures in Study 1 (n = 48) & Study 2, Grade 1 (n = 35)*

	Letter Sounds		Syllable Sounds		Phoneme Segmenting		Syllable Segmenting		Word Reading	
	1	2	1	2	1	2	1	2	1	2
Syllable Sounds	.46**	.28								
Phoneme Segmenting	.42**	.13	.33*	.01						
Syllable Segmenting	-.11	-.07	.05	-.04	.23	-.06				
Word Reading	.31*	.09	.87**	.56**	.21	.02	.19	.22		
Sentence Reading	NA	.06	NA	.45**	NA	-.10	NA	.11	NA	.57**

\* $p < .05$

\*\* $p < .01$

**Table 53. Example of Sentence Reading Regression Summary from Technical Report 1301**

Table 7  
*Sentence Reading Regression Summary for Grade 2, Study 1*

Independent Variables	Unstandardized Coefficients		Standardized Coefficients	t	Partial and Part Correlations	
	B	Std. Error	Beta		Partial	Part
Syllable Sounds	1.34	0.25	0.57	5.29	0.61	0.52
Syllable Segmenting	0.96	0.37	0.28	2.62	0.36	0.26
Constant	-24.15	8.08		-2.99		

**Table 54. Key Findings Summary from Technical Report 1301**

Category	Summary
<b>Grades</b>	Kindergarten through Grade 2
<b>Assessment Focus</b>	Spanish early literacy
<b>Sample</b>	Spanish immersion and dual-language bilingual students
<b>Statistical Models</b>	Correlations and linear regression
<b>Primary Finding</b>	Syllable-level measures best predict reading fluency
<b>Intended Use</b>	Progress monitoring and instructional decision making

**Reference**

Alonzo, J., Gonzalez, M., & Tindal, G. (2013). *The development of easyCBM® Spanish literacy assessments for use in Grades K–2 (Technical Report 1301)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

---

**Summary of Technical Report 1901: Initial Norms for the easyCBM® Spanish Language Vocabulary Assessments (Grades 2 – 8) (Alonzo, 2019).**
**Methods**

This technical report describes the development and initial norming of the easyCBM® Spanish Language Vocabulary Assessments for Grades 2–8. The assessments were developed in summer and fall 2016 by a team of native Spanish-speaking educators working collaboratively with University of Oregon assessment developers. Item writers included teachers and paraprofessionals with grade-specific experience in Spanish-language instruction. Prior to writing items, the team participated in a structured training workshop focused on bias reduction, avoidance of regional dialect differences, and the importance of writing original Spanish items rather than translating English vocabulary content.

Items were written in small grade-band teams (Grades 2–3, 4–5, 6–8) and overseen by a Lead Item Developer with expertise in Spanish literature and bilingual education. Each vocabulary item consisted of a sentence stem designed to provide contextual meaning, followed by three answer choices: one correct answer, a near distractor, and a far distractor. After internal review, items were piloted online during the 2017 and 2018 school years. Item Response Theory (IRT) analyses were conducted in spring 2018 to calibrate items and construct 13 comparable alternate forms per grade. Three forms were designated for benchmark screening, with ten retained for progress monitoring. Initial norms were generated using data from Fall 2018 and Winter 2018 benchmark administrations. The sample included students from 19 states (see page 6). The report notes that the sample represents a convenience sample and that winter sample sizes were particularly small in several grades. Benchmark assessments were restricted to on-grade administration within district-designated testing windows.

## Results

The initial normative data demonstrate expected developmental and seasonal trends in vocabulary performance across Grades 2–8. Across grades, winter percentile ranks generally reflect higher performance at equivalent raw scores compared to fall administrations, suggesting growth over the school year.

In Grades 2 through 5, raw score distributions show steady upward shifts from fall to winter. Students achieving moderate raw scores in fall often corresponded to substantially higher percentile ranks in winter, reflecting improved performance relative to peers. This pattern is consistent with expected midyear vocabulary development and suggests sensitivity to instructional exposure over time.

In Grades 6 through 8, similar seasonal increases are observed, although variability in percentile distributions is somewhat greater. Some grades display clustering of higher percentile ranks at relatively low raw scores, which may reflect smaller winter sample sizes and limited score spread in the initial dataset. In the upper grades, several raw scores correspond to very high percentile rankings, indicating possible ceiling effects or compressed distributions in the preliminary norming sample.

Overall, the data reflect reasonable growth patterns across seasons and grades but should be interpreted cautiously. Because norms are based on a convenience sample and winter participation was uneven across grades, the author emphasizes that these norms are preliminary. Larger and more representative samples are needed before the benchmarks can be considered fully stable for high-stakes decisions. The report recommends using these norms as provisional interpretive guidance rather than definitive cut scores until further data collection strengthens the normative base.

## Reference

Alonzo, J. (2019). *Initial norms for the easyCBM® Spanish language vocabulary assessments (Technical Report No. 1901)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

---

**Appendix A: Technical Report Table Titles**

Table 1. Tasks for Kindergarten Blueprint

Table 2. Tasks for Grade 1 Blueprint

Table 3. Tasks for Grades 2–8 Blueprint

Table 4. Standards Crosswalk with CCSS, NRP, and easyCBM®

Table 5. Example of IRT Analysis for Technical Report 0039

Table 6. Summary Table for Technical Report 0039

Table 7. Typical Results of IRT Analysis from Technical Report 0040

Table 8. Example of Descriptive Statistics from Technical Report 0040

Table 9. Key Findings Summary from Technical Report 0040

Table 10. Example of Descriptive Statistics from Technical Report 0043

Table 11. Sample Correlations from Technical Report 0043

Table 12. Example of Changes Made: Grade 5 Passage Reading Measures in Technical Report 0043

Table 13. Key Findings Summary from Technical Report 0043

Table 14. Illustrative Descriptive Statistics from Technical Report 0046

Table 15. Typical Correlation from Technical Report 0046

Table 16. Key Findings Summary from Technical Report 0046

Table 17. Example of Item Statistics from Technical Report 0041

Table 18. Typical Distractor Analysis from Technical Report 0041

Table 19. Example of Item Statistics from Technical Report 0807

Table 20. Illustrative Distractor Analysis from Technical Report 0807

Table 21. Summary of Key Findings from Technical Report 0807

Table 22. Sample Item Statistics from Technical Report 0808

Table 23. Illustrative Distractor Analysis from Technical Report 0808

Table 24. Summary of Key Findings from Technical Report 0808

Table 25. Summary of Key Findings from Technical Report 0041, 0807, and 0808

Table 26. Example Item Statistics from Technical Report 1102

Table 27. Key Findings Summary from Technical Report 1102

Table 28. Example of Summary Statistics from Technical Report 1209

Table 29. Key Findings Summary from Technical Report 1209

Table 30. Example Item Information and Statistics from Technical Report 1210

Table 31. Typical Distractor Analysis Results from Technical Report 1210

Table 32. Key Findings Summary from Technical Report 1210

Table 33. Sample Item Information and Statistics from Technical Report 1211

Table 34. Key Findings Summary from Technical Report 1211

Table 35. Example of Item Information and Statistics from Technical Report 1212

Table 36. Key Findings Summary from Technical Report 1212

Table 37. Example of Item Information and Statistics from Technical Report 1213

Table 38. Key Findings Summary from Technical Report 1213

Table 39. Example of Item Information and Statistics from Technical Report 1214

Table 40. Key Findings Summary from Technical Report 1214

Table 41. Example of Item Information and Statistics from Technical Report 1215

Table 42. Key Findings Summary from Technical Report 1215

Table 43. Typical Results of IRT Analysis from Technical Report 1221

Table 44. Key Findings Summary from Technical Report 1221

Table 45. Typical Results of IRT Analysis for Technical Report 1222

Table 46. Key Findings Summary from Technical Report 1222

Table 47. Key Findings Summary from Technical Report 1223

Table 48. Key Findings Summary from Technical Report 1224

Table 49. Key Findings Summary from Technical Report 1225

Table 50. Key Findings Summary from Technical Report 1226

Table 51. Example of Descriptive Statistics from Report 1301

Table 52. Typical Set of Relations Among Measures from Report 1301

Table 53. Example of Sentence Reading Regression Summary from Technical Report 1301

Table 54. Key Findings Summary from Technical Report 1301

### Technical Report References

- Alonzo, J. (2019). *Initial norms for the easyCBM® Spanish language vocabulary assessments (Technical Report # 1901)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012a). *The development of CBM vocabulary measures: Grade 2 (Technical Report # 1209)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012b). *The development of CBM vocabulary measures: Grade 3 (Technical Report # 1210)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012c). *The development of CBM vocabulary measures: Grade 4 (Technical Report # 1211)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012d). *The development of CBM vocabulary measures: Grade 5 (Technical Report # 1212)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012). *The development of CBM vocabulary measures: Grade 6 (Technical Report # 1213)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012e). *The development of CBM vocabulary measures: Grade 7 (Technical Report # 1214)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Anderson, D., Park, B. J., & Tindal, G. (2012f). *The development of CBM vocabulary measures: Grade 8 (Technical Report # 1215)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Gonzalez, M., & Tindal, G. (2013). *The development of easyCBM Spanish literacy assessments for use in Grades K-2 (Technical Report # 1301)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Liu, K., & Tindal, G. (2007). *Examining the technical adequacy of reading comprehension measures in a progress monitoring assessment system (Technical Report # 41)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Liu, K., & Tindal, G. (2008). *Examining the technical adequacy of second-grade reading comprehension measures in a progress monitoring assessment system (Technical Report # 0808)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Park, B. J., & Tindal, G. (2008). *The development of middle school passage reading fluency measures in a progress monitoring assessment system (Technical Report # 46)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Park, B. J., & Tindal, G. (2012a). *The development of the easyCBM CCSS reading assessments: Grade 3 (Technical Report # 1221)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Park, B. J., & Tindal, G. (2012b). *The development of the easyCBM CCSS reading assessments: Grade 4 (Technical Report # 1222)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Park, B. J., & Tindal, G. (2012c). *The development of the easyCBM CCSS reading assessments: Grade 5 (Technical Report # 1223)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Park, B. J., & Tindal, G. (2012d). *The development of the easyCBM CCSS reading assessments: Grade 6 (Technical Report # 1224)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Park, B. J., & Tindal, G. (2012). *The development of the easyCBM CCSS reading assessments: Grade 7 (Technical Report # 1225)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., Park, B. J., & Tindal, G. (2012e). *The development of the easyCBM CCSS reading assessments: Grade 8 (Technical Report # 1226)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

- Alonzo, J., & Tindal, G. (2007a). *The development of early literacy measures for use in a progress monitoring assessment system: Letter names, letter sounds and phoneme segmenting (Technical Report # 39)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., & Tindal, G. (2007b). *The development of word and passage reading fluency measures in a progress monitoring assessment system (Technical Report # 40)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., & Tindal, G. (2008a). *The development of fifth-grade passage reading fluency measures for use in a progress monitoring assessment system (Technical Report # 43)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Alonzo, J., & Tindal, G. (2008b). *Examining the technical adequacy of fifth-grade reading comprehension measures in a progress monitoring assessment system (Technical Report # 0807)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Nese, J. F. T., Anderson, D., Hoelscher, K., Tindal, G., & Alonzo, J. (2011). *Progress monitoring instrument development: Silent reading fluency, vocabulary, and reading comprehension (Technical Report # 1110)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Park, B. J., Alonzo, J., & Tindal, G. (2011). *The development and technical adequacy of seventh-grade reading comprehension measures in a progress monitoring assessment (Technical Report # 1102)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.