

Technical Report 2603-VK8M

Validity Analyses for easyCBM[®] in Grades K-8: Mathematics

Gerald Tindal, PhD – University of Oregon

Sara McCaslin, PhD – 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.

Reference: Tindal, G. (2026). *Validity Analyses for easyCBM[®] in Grades K-8: Mathematics (Technical Report # 2603-VK8M)*. 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-VK8M_ValidityMath

Introduction and Overview¹	4-5
Summary of Technical Reports	
1002: The Alignment of easyCBM® Math Measures to Curriculum Standards	5
1228: The Alignment of the easyCBM® Grades K–2 Math Measures to the Common Core Standards	7
1229: The Alignment of the easyCBM® Grades 3–5 Math Measures to the Common Core Standards	8
1230: The Alignment of the easyCBM® Grades 6–8 Math Measures to the Common Core Standards	10
1208: The Alignment of the easyCBM® Middle School Mathematics CCSS Measures to the Common Core State Standards	11
2101: The Alignment Between easyCBM® Mathematics Assessments and State and National Standards	12
1006: Technical Adequacy of the easyCBM® Primary-Level Mathematics Measures (Grades K– 2), 2009–2010 Version	13
1007: Technical Adequacy of the easyCBM® Mathematics Measures: Grades 3–8, 2009–2010 version	14
1008: Diagnostic Efficiency of easyCBM® Math: Washington State	15
1009 Summary: Diagnostic efficiency of easyCBM® math: Oregon	16
1010: easyCBM® Mathematics Criterion Related Validity Evidence: Washington State Test	18
1011: easyCBM® Mathematics Criterion Related Validity Evidence: Oregon State	19
1402 Summary: Criterion Validity Evidence for the easyCBM® CCSS Math Measures Grades 6- 8	20
1104: A Cross-Validation of easyCBM® Mathematics Cut Scores in Oregon: 2009–2010	21
1105: A Cross-Validation of easyCBM® Mathematics Cut Scores in Washington State: 2009– 2010 Test	22
1501: An Exploration of Differential Item Functioning with the easyCBM® Middle School Mathematics Tests: Grades 6–8	23
2401: Criterion Validity and Classification Accuracy of easyCBM®: Grades 3-8	25
Appendix A: Technical Report Table Titles	28
Appendix B: Guide to Spreadsheet Technical Report Value Displays	29
Technical Report References	30-31

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

Abstract

In this series of studies on the validity of the easyCBM[®] Mathematics measures, the first set of reports summarizes a variety of analytical procedures: psychometric characteristics (difficulty and discrimination as well as fit), alignment with standards, reliability values, and criterion related associations with other measures using correlations along with regression and predictive modeling. The studies were conducted in several locations across the U.S., sampling from different populations, as part of the initial development of the measures, prior to and concurrent with implementation. **Note:** All tables and figures in this summary are examples of those presented in full within the individual Technical Reports: They are not exhaustive, but only illustrative.

Validity Evidence for easyCBM[®] Mathematics Measures

The Standards for Educational and Psychological Testing (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014)² define validity as the degree to which evidence and theory support the interpretations of test scores for intended uses. Validity is not a property of a test itself, but of the inferences drawn from test scores. The Standards conceptualize validity as a unified construct supported by multiple complementary sources of evidence. Five primary sources are identified: evidence based on test content, response processes, internal structure, relations to other variables, and consequences of testing.

Evidence based on test **content** examines the extent to which assessment items represent the intended construct domain. This includes blueprint analyses, alignment to academic standards, depth-of-knowledge reviews, and expert judgment studies. In mathematics assessment, content evidence is often gathered through structured alignment studies evaluating item-to-standard correspondence and representativeness.

Evidence based on **response processes** evaluates whether examinees engage in the intended cognitive processes when responding to items. Methods such as cognitive interviews, scoring audits, and analysis of student work are used to determine whether responses reflect targeted mathematical reasoning.

Evidence based on **internal structure** evaluates dimensionality, reliability, and item functioning. Classical test theory indices (e.g., internal consistency), item response theory modeling, Rasch analysis, and differential item functioning studies provide statistical evidence regarding construct coherence and fairness across subgroups.

Evidence based on **relations to other variables** includes correlations, regression models, classification accuracy indices, and predictive validity analyses. Convergent validity is demonstrated when scores relate strongly to other measures of mathematics achievement; discriminant validity is supported when relations with unrelated constructs are weaker; predictive validity supports forecasting future outcomes.

Evidence based on **consequences** of testing addresses intended and unintended effects of score use, including decision accuracy, instructional implications, and fairness considerations. Together, these five sources provide a comprehensive framework for evaluating the validity of mathematics assessment interpretations and supporting responsible score use.

Summary of Technical Reports on Validity with easyCBM[®]

The series of technical reports summarized in Technical Report 2603-VK8M collectively provide comprehensive validity evidence for the easyCBM[®] mathematics measures across Grades K–8. Across studies, consistent methodological frameworks were applied, including classical test theory indices, Rasch modeling, alignment reviews, and criterion-related analyses linked to statewide accountability assessments. **Note:** Alignment-focused reports are summarized in Technical Report 2603-A38RM_AlignmentReadMath, documenting strong correspondence between easyCBM[®] measures and the Common Core State Standards across grade bands. Structured expert review protocols confirmed substantial coverage of focal standards, with minor refinements identified for continuous improvement, providing evidence based on content.

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

Technical Reports 1006 and 1007 established foundational psychometric adequacy for primary (K–2) and upper-grade (3–8) mathematics measures. Internal consistency reliability coefficients were consistently within acceptable to strong ranges for screening and instructional decision-making. Item-level analyses demonstrated appropriate difficulty distributions and positive discrimination indices. Rasch analyses supported acceptable item fit statistics and stable parameter estimates, providing evidence of internal structure.

Criterion-related validity studies (1010, 1011, 1402) demonstrated moderate to strong correlations between easyCBM[®] benchmark scores and statewide mathematics assessments in Washington and Oregon. Regression and predictive modeling analyses indicated that benchmark scores contributed meaningful explanatory power in predicting proficiency outcomes. Receiver operating characteristic analyses yielded balanced sensitivity and specificity estimates, supporting risk classification decisions.

Diagnostic efficiency and cross-validation reports (1008, 1009, 1104, 1105) extended this evidence by examining cut score performance across states. Classification accuracy statistics demonstrated acceptable predictive values, supporting the consequences of score interpretation in screening contexts.

Technical Report 1501 examined differential item functioning across demographic subgroups using item response theory procedures. Results indicated minimal subgroup bias, supporting fairness and equitable measurement and Technical Report 2401 reported on classification accuracy.

Taken together, these integrated findings provide evidence across all five sources identified in the Standards (2014): content alignment, response process consistency, internal structure via reliability and Rasch modeling, relations to external variables through criterion and predictive studies, and consequential validity through classification accuracy and fairness analyses. The convergence of findings across multiple states, grade levels, and analytical methods supports the technical adequacy and validity of easyCBM[®] mathematics scores for screening, progress monitoring, and benchmark decision-making.

Summary of Technical Report 1002: The Alignment of easyCBM[®] Math Measures to Curriculum Standards (Nese, Lai, Anderson, Park, et al., 2010)

Methods

Participants included 13 certified teachers (one employed as a district curriculum specialist) recruited from Hillsboro and Springfield School Districts in Oregon. Raters were assigned to one of three grade bands (K–1, 3–5, or 6–8), and all had prior experience with easyCBM[®] math measures. The measures consisted of easyCBM[®] benchmark and progress monitoring math forms across grades K, 1, and 3–8 (grade 2 was excluded as items did not target NCTM Focal Points for that grade). Each form contained 16 items. Data were collected from November 2009 to January 2010 via online conferencing training sessions of 1.5–2 hours, followed by independent remote rating over 2–4 weeks. Raters judged item alignment to the NCTM Curriculum Focal Points on a 0–3 scale (0 = no link; 3 = clear, direct link) and assigned Depth of Knowledge (DOK) levels (1–3) to items and standards for grades 3–8. Analyses included descriptive statistics (frequency and percentage of ratings, dichotomized linked/not-linked), and inter-rater reliability was estimated using intraclass correlation coefficients (ICC) derived from a two-level hierarchical cross-classified model (HLM) that partitioned variance attributable to items, raters, and residual error.

Results

Alignment to Standards. Across grades, focal points, and test forms, teacher ratings of easyCBM[®] mathematics items showed generally strong alignment to the NCTM Curriculum Focal Points. Excluding one focal point in grade 8, the percentage of items rated as linked to standards ranged from 65% to 100%. At the kindergarten level, benchmark form alignment was 88% for Numbers/Operations, 98% for Geometry, and 67% for Measurement—with Measurement being the weakest focal point. Progress monitoring forms showed a range of 25%–100%, with Measurement again the lowest (65%). Grade 1 benchmark alignment ranged from 83% to 94% across focal points, and progress monitoring forms ranged from 84% to 95%. Grades 3 through 7 generally showed strong alignment, with most focal points exceeding 75%–88% for both benchmark and progress monitoring forms. Grade 8 showed the weakest overall alignment: benchmark forms ranged from 42% (Geometry/Measurement) to 77% (Data

Analysis/Numbers & Operations/Algebra), and progress monitoring forms from 58% to 73%. The Geometry/Masurement focal point at grade 8 was the single outlier below 65% across the entire study.

Across grades and focal points, teacher DOK ratings of easyCBM® items were predominantly at levels 1 (Recognition and Reproduction) and 2 (Skill and Concept). Consensus among raters on DOK item ratings within any given form ranged widely (0%–100%), though more typically between 13% and 50%. Consensus on the highest DOK level (Strategic Thinking, level 3) was rarely achieved. Geometry and Algebra standards were consistently rated at higher DOK levels than Number and Operations standards, suggesting these domains are conceptually more demanding. DOK ratings across raters showed relatively low inter-rater agreement overall, indicating that DOK classification is inherently subjective, limiting the strength of conclusions that can be drawn from these data. Inter-rater reliability for standard alignment ratings was high, with ICCs ranging from .80 to 1.0. Reliability for item-level and standard-level DOK ratings was moderately high, with ICCs ranging from .50 to .80. These results indicate that raters were dependably consistent in judging alignment to content standards, but more variable in their DOK classifications.

This study marked the first application of Webb’s alignment model to a curriculum-based measurement (CBM) system aligned to modified state content standards. The authors noted that the Webb model was originally designed for large-scale summative assessments, and applying it to formative measures with extensive item banks (over 11,000 items) posed logistical and methodological challenges. Notably, the standard panel consensus process was not fully implemented, as raters worked independently rather than collectively calibrating ratings. Despite these limitations, the results provide meaningful evidence of content validity for the easyCBM® math assessment system, demonstrating that the measures are generally well-aligned to nationally recognized mathematics standards across grade levels.

Table 1. Illustrative Results from Technical Report 1002

Table 76

Grade 8 Benchmark Measures: Individual Raters' Ratings on Standard Depth of Knowledge

Focal point	Term	Ratings	Raters		
			J	K	L
Algebra	Fall	Recognition and Reproduction (1)	62.5	6.3	--
		Skill and Concept (2)	37.5	56.3	--
		Strategic Thinking (3)	0	37.5	--
		Extended Thinking (4)	0	0	--
	Winter	Recognition and Reproduction (1)	81.3	18.8	--
		Skill and Concept (2)	18.8	62.5	--
		Strategic Thinking (3)	0	18.8	--
		Extended Thinking (4)	0	0	--
	Spring	Recognition and Reproduction (1)	62.5	12.5	--
		Skill and Concept (2)	37.5	50.0	--
		Strategic Thinking (3)	0	37.5	--
		Extended Thinking (4)	0	0	--
Geometry	Fall	Recognition and Reproduction (1)	87.5	--	18.8
		Skill and Concept (2)	12.5	--	56.3
		Strategic Thinking (3)	0	--	25.0
		Extended Thinking (4)	0	--	0
	Winter	Recognition and Reproduction (1)	68.8	--	18.8
		Skill and Concept (2)	31.3	--	62.5
		Strategic Thinking (3)	0	--	18.8
		Extended Thinking (4)	0	--	0
	Spring	Recognition and Reproduction (1)	87.5	--	25.0
		Skill and Concept (2)	12.5	--	62.5
		Strategic Thinking (3)	0	--	12.5
		Extended Thinking (4)	0	--	0

Reference

Nese, J. F. T., Lai, C.-F., Anderson, D., Park, B. J., Tindal, G., & Alonzo, J. (2010). *The alignment of easyCBM® math measures to curriculum standards (Technical Report No. 1002)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1228: The Alignment of the easyCBM® Grades K–2 Math Measures to the Common Core Standards (Irvin et al., 2012b).

Methods

This study examined the degree to which the easyCBM® mathematics benchmark assessments for kindergarten through grade 2 align with the Common Core State Standards (CCSS). Because formative assessments are often used to guide instructional decisions in a Response to Intervention (RTI) framework, the authors emphasized that these measures must closely reflect the academic standards used to guide classroom instruction. The purpose of the study was therefore to determine whether the existing easyCBM® math items adequately represented the CCSS domains and standards that students are expected to master at each grade level.

The alignment analysis focused on the seasonal benchmark assessments included in the easyCBM® system. These assessments contain items designed to measure students' understanding of core mathematics concepts across the school year. Although the assessments were originally written to align with the National Council of Teachers of Mathematics (NCTM) Curriculum Focal Points, the introduction of the CCSS required a systematic evaluation of how well those items corresponded with the new standards.

A panel of mathematics experts participated in the alignment study. These reviewers had expertise in mathematics instruction, standards-based curriculum, and the CCSS. Participants were recruited through an online screening process designed to confirm their qualifications and experience with standards-based mathematics education. After completing a training webinar, reviewers independently examined the easyCBM® benchmark assessment items and identified the CCSS domain and standard that each item most closely represented.

Each item was evaluated for alignment at two levels. First, reviewers determined whether an item aligned with a specific CCSS domain such as Operations and Algebraic Thinking, Number and Operations in Base Ten, or Measurement and Data. Second, reviewers identified the CCSS standard addressed by the item. Items were classified as aligning with an on-grade standard, a prior-grade standard, or not aligning with any CCSS standard. The results of these ratings were aggregated across reviewers to produce alignment summaries for kindergarten, first grade, and second grade benchmark assessments.

Results

The analysis found that the easyCBM® K–2 mathematics benchmark assessments demonstrate strong overall alignment with the Common Core State Standards. Across grades, most items aligned with either on-grade or prior-grade CCSS standards. Specifically, approximately 94% of kindergarten items, 99% of first grade items, and 96% of second grade items were judged to align with CCSS expectations. These findings suggest that the easyCBM® math assessments generally measure content that is consistent with the skills outlined in the standards.

At the domain level, alignment across grade levels was generally strong. The benchmark assessments included items representing multiple CCSS domains, providing coverage of major mathematics content areas. However, when alignment was examined at the individual standard level, several gaps became apparent. Some CCSS standards were represented by few or no assessment items, while others were represented multiple times.

For example, in grade 2 the Measurement and Data domain was broadly represented, but several standards within that domain were not addressed by any benchmark assessment items, including standards 2.MD.2, 2.MD.4, 2.MD.5, 2.MD.9, and 2.MD.10. Additional gaps were observed in the Numbers and Operations in Base Ten domain (e.g., standards 2.NBT.3, 2.NBT.6, 2.NBT.8, and 2.NBT.9) and in Operations and Algebraic Thinking (e.g., standards 2.OA.3 and 2.OA.4). Conversely, some standards such as 2.MD.1 and 2.NBT.5 appeared frequently and were therefore somewhat overrepresented within the existing item pool.

These results indicate that while the easyCBM® math measures provide broad coverage of CCSS domains, the distribution of items across specific standards is uneven. The findings were used to guide subsequent assessment development. In particular, the authors recommended writing new items to address underrepresented CCSS standards in order to strengthen alignment and ensure more balanced coverage of mathematics content across the K–2 assessments.

Table 2. Illustrative Table of Key Findings from Technical Report 1228

Table 4
Item level alignment results for the easyCBM® first grade fall benchmark in mathematics.

Item	PS1	PS1		PS1	PS2	PS2		PS2	SS1		SS1	SS2		SS2	SS3		SS3	Total	Req Skills
		N	Ave			N	Ave		N	Ave		N	Ave		N	Ave			
1	K.CC.2	2	1	2				1.NBT.1	1	2	1.OA.1	1	2	1.OA.5	1	2	5	0	
2	1.NBT.1	4	3	2				K.CC.1	1	2							5	0	
3	NS	3		**				1.MD.1	1	2	1.MD.4	1	1				5	1	
4	1.NBT.2	3	2	2				1.NBT.4	2	2							5	0	
5	NS	4		**				1.NBT.2	1	1							5	1	
6	1.NBT.3	2		1	NS	2	**	K.CC.2	1	*							5	2	
7	NS	3		**				1.MD.4	2	1							5	1	
8	NS	4		**				1.NBT.1	1	*							5	1	
9	1.NBT.1	5	4	2													5	0	
10	1.NBT.2	5	4	1.5													5	0	
11	1.NBT.1	4	3	2				1.OA.4	1	1							5	0	
12	NS	4		**				1.NBT.1	1	*							5	1	
13	1.NBT.2	4	3	1.67				NS	1	**							5	1	
14	1.NBT.2	4	3	2				1.NBT.4	1	2							5	0	
15	1.NBT.3	4		2				1.OA.8	1	*							5	0	
16	1.NBT.1	5	4	2													5	0	
17	K.G.4	5	4	1.5													5	0	
18	K.G.6	5	4	2													5	0	
19	K.G.4	5	4	1.5				K.G.2	1	*							6	0	
20	K.G.4	4	3	2				1.G.4	1	2							5	0	
21	1.G.2	3	2	1.5				K.G.5	1	2	K.G.6	1	2				5	0	
22	K.G.4	4	3	1.67				1.G.1	1	1							5	0	
23	K.G.4	3	2	1.5				1.G.1	1	1	NS	1	**				5	1	
24	K.G.4	4	3	1.67				1.G.1	1	1							5	0	
25	K.G.6	3		2				1.G.2	2	2							5	0	
26	K.G.2	3	2	2				1.G.2	1	2	NS	1	**				5	0	
27	NS	3		**				1.G.3	2	1.5							5	1	

Reference

Irvin, P. S., Bitnara J. P., Alonzo, J., & Tindal, G. (2012). *The Alignment of the easyCBM® grades K–2 math measures to the Common Core Standards (Technical Report 1228)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1229: The Alignment of the easyCBM® Grades 3–5 Math Measures to the Common Core Standards (Park et al., 2012).

Methods

This study evaluated how well the easyCBM® seasonal mathematics benchmark assessments for grades 3 through 5 align with the Common Core State Standards (CCSS). The easyCBM® system includes formative benchmark assessments administered three times per year to monitor student progress. The mathematics items used in these assessments were originally developed to align with the National Council of Teachers of Mathematics (NCTM) Curriculum Focal Points. With the adoption of the CCSS, the researchers conducted a systematic alignment study to determine how well the existing items correspond to the newer standards.

The study was conducted in two phases. In Phase 1, one educator per grade level (K–8) conducted an initial review of easyCBM® benchmark items and their alignment with CCSS standards. Reviewers were selected based on their experience teaching mathematics and their familiarity with the Common Core. The nine Phase 1 reviewers included general education teachers, special education teachers, teachers who taught both settings, and district-level mathematics specialists. Participants averaged approximately 12 years of mathematics teaching experience, with experience ranging from 3 to 31 years. The reviewers represented multiple states including Washington, Ohio, South Carolina, New Jersey, Indiana, Kansas, and Arizona.

Phase 2 expanded the review process by adding four additional educators per grade level. These reviewers were again selected for their expertise in mathematics instruction and familiarity with CCSS. Each reviewer examined grade-level benchmark assessment items and identified the CCSS domain and specific standard that best aligned with each item.

The alignment analysis included all seasonal benchmark mathematics items for grades 3–5. Each grade-level benchmark assessment contains three testing periods (fall, winter, and spring), with 45 items per assessment. This resulted in 135 items per grade being evaluated for alignment. Reviewers determined whether each item aligned with an on-grade CCSS standard, a prior-grade standard, or did not align with any standard. These classifications were aggregated to examine alignment patterns across grade levels, domains, and individual standards.

Results

Overall, the analysis found strong alignment between easyCBM® benchmark mathematics items and the CCSS across grades 3–5. Approximately 98% of third-grade items, 100% of fourth-grade items, and 97% of fifth-grade items were aligned with either grade-level or prior-grade CCSS standards. These results indicate that the benchmark assessments largely measure mathematical content consistent with the Common Core standards.

Although overall alignment was high, the analysis revealed uneven representation across domains and individual standards. At the domain level, the assessments generally covered the major CCSS content areas. However, certain domains and standards were either underrepresented or overrepresented within the item pool.

For example, in grade 5 the Number and Operations in Base Ten domain was highly represented, with more than 70 items across the three benchmark assessments aligning with standards from that domain either as primary or secondary standards. In contrast, several standards within the Number and Operations—Fractions domain were underrepresented, including standards 5.NF.3 through 5.NF.7. Additional gaps were identified in several domains, including Operations and Algebraic Thinking (5.OA.1–5.OA.3), Number and Operations in Base Ten (5.NBT.1–5.NBT.3 and 5.NBT.5), Measurement and Data (5.MD.1–5.MD.2), and Geometry (5.G.1–5.G.4).

The results indicate that while easyCBM® benchmark assessments broadly align with CCSS expectations, the distribution of items across standards is uneven. The findings were used to guide subsequent assessment development, particularly the creation of new items targeting underrepresented standards to strengthen alignment between easyCBM® mathematics measures and the Common Core standards.

Table 3. Illustrative Table of Key Findings from Technical Report 1229

Table 1
Item level alignment results for the easyCBM® third grade fall benchmark in mathematics.

Item	PS1	PS1 N	PS1 Ave N	PS1 Ave	PS2	PS2 N	PS2 Ave N	PS2 Ave	SS1	SS1 N	SS1 Ave N	SS1 Ave	SS2	SS2 N	SS2 Ave N	SS2 Ave	SS3	SS3 N	SS3 Ave N	SS3 Ave	Total n	Req Skills
1	3.NF.1	2		2					2.G.3	1		1	3.G.2	1		1	NS	1	**			
2	3.NF.1	4	3	2					3.G.2	1		2										
3	3.NF.1	4	3	2					3.G.2	1		2										
4	3.NF.1	4	3	2					3.G.2	1		2										
5	3.NF.1	4	3	2					3.G.2	1		2										
6	3.NF.1	4	3	2					3.G.2	1		2										
7	3.NF.1	4	3	2					3.G.2	1		1										
8	3.NF.1	4	3	1.67					3.G.2	1		2										
9	3.NF.1	4	3	1.67					3.G.2	1		1										
10	3.NF.3	5	4	2																		
11	3.NF.3	4	3	2					3.NF.2	1		2										
12	3.NF.3	5	4	1.5																		
13	3.NF.3	5	4	2																		
14	3.NF.3	4	3	2					3.NF.1	1		1										
15	NS	3		**					2.G.1	1		1	3.G.2	1		1						
16	NS	3		**					2.G.1	1		1	3.G.2	1								
17	NS	3		**					2.G.1	1		1	3.G.2	1								
18	3.G.2	2		1.5	NS	2	1	**	2.G.1	1		2										
19	NS	2	1	**					2.OA.1	1		2	2.MD.1	1		1	3.MD.6	1				
20	2.G.1	2		1.5	NS	2	1	**	4.G.1	1		1										
21	2.G.1	2		2	2.G.1	2	1	1	4.G.2	1		2										
22	NS	3		**					2.G.1	2		1.5										
23	2.G.1	2		2					3.G.1	1		2	3.G.2	1		1	NS	1		**		
24	NS	4		**					2.G.1	1		1										
25	***								2.G.2	1		2	2.OA.1	1		2	2.OA.2	1		2	3.G.2	1
26	NS	4		**					3.G.1	1		2										
27	3.G.1	2		2					4.G.1	1		2	2.G.1	1		*	NS	1		**		

Reference

Park, B. J., Irvin, P. S., Alonzo, J., & Tindal, G. (2012). *The Alignment of the easyCBM® grades 3–5 math measures to the Common Core Standards (Technical Report 1229)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1230: The Alignment of the easyCBM® Grades 6–8 Math Measures to the Common Core Standards (Irvin et al., 2012a).

Methods

This technical report examined the alignment between easyCBM® seasonal benchmark mathematics assessments and the Common Core State Standards (CCSS) for grades 6 through 8. The easyCBM® system provides formative benchmark and progress-monitoring assessments designed to support instructional decision-making within Response to Intervention (RTI) frameworks. Because instructional decisions depend on the degree to which assessments reflect current academic standards, the study evaluated how closely existing easyCBM® mathematics items corresponded to CCSS expectations.

The study used a structured alignment review conducted in two phases. In Phase 1, a single educator for each grade level (K–8) conducted an initial review of easyCBM® mathematics benchmark items and their correspondence to CCSS standards. Reviewers were selected based on their experience teaching mathematics and familiarity with the Common Core. Phase 2 expanded the process by adding four additional educators per grade level to review the items and confirm alignment judgments. Participants included general education teachers, special education teachers, and district mathematics specialists with substantial classroom experience. Reviewers evaluated each benchmark item and identified the CCSS domain and specific standard to which it aligned. The analysis included all seasonal benchmark assessments in fall, winter, and spring, with ~135 items per grade level evaluated for alignment.

Results

Results indicated strong overall alignment between easyCBM® mathematics benchmark items and CCSS standards for grades 6–8. Approximately 99% of sixth-grade items, 93% of seventh-grade items, and 96% of eighth-grade items aligned with either on-grade or prior-grade CCSS standards.

While the overall alignment rates were high, the analysis revealed uneven representation of individual standards across domains. Some CCSS standards were heavily represented by multiple items, whereas others were minimally represented or absent. These findings provided guidance for future item development within the easyCBM® system, particularly for creating new assessment items that address underrepresented standards and improve alignment with the Common Core.

Table 4. Illustrative Table of Key Findings from Technical Report 1230

Table 1
Item level alignment results for the easyCBM® sixth grade fall benchmark in mathematics.

Item	PS1	PS1 N	PS1 Ave	PS2	PS2 N	PS2 Ave	SS1	SS1 N	SS1 Ave	SS2	SS2 N	SS2 Ave	SS3	SS3 N	SS3 Ave	SS4	SS4 N	SS4 Ave	SS5	SS5 N	SS5 Ave	Total n	Req Skills
1	NS	4	**				5NBT.3	1	1													5	3
2	5NF.4	2	1	2	5NF.6	2	1	1	NS	1	**											5	1
3	5NBT.7	2	2				6NS.2	1	2	6NS.3	1	*	NS	1	**							5	1
4	NS	3	**				5NBT.3	1	1	5NF.3	1	1										5	2
5	5NF.6	3	2				5NF.4	2	1	2												5	0
6	5NF.4	4	3	1.67			5NF.6	1	2													5	0
7	5NF.4	2	1	2	5NF.6	2	1	5NF.5	1	2												5	0
8	5NF.4	3	2	1.5	5NF.7	2	1.5	5NF.7	2	1.5												7	0
9	5NBT.2	3	1.67				6EE.1	2	1	2												5	0
10	NS	3	**				5NBT.3	1	2	5NF.3	1	1										5	2
11	NS	3	**				5NBT.7	1	1	6EE.3	1	2										5	2
12	NS	4	**				5NF.5	1	1													5	3
13	NS	3	**				5NBT.3	1	1	5NF.3	1	2										5	2
14	NS	3	**				5NBT.3	2	1													5	2
15	5NF.2	3	2	2			NS	2	**													5	1
16	6EE.2	3	2	2			6EE.5	1	2	6EE.6	1	1										5	0
17	6EE.2	4	3	1.67			6EE.5	1	2													5	0
18	6EE.2	2	1.5				6EE.5	1	2	6EE.6	1	1	NS	1	**							5	0
19	6EE.2	2	2				6EE.6	1	*	6EE.7	1	2	NS	1	**							5	0
20	NS	4	**				6XP.2	1	2													5	2
21	6EE.7	2	1	2	6EE.2	2	2	6EE.5	1	1												5	0
22	6EE.6	2	1	2	6EE.2	2	1.5	6EE.7	1	2												5	0
23	6EE.5	2	1	2			6EE.2	1	2	6EE.3	1	2	NS	1	**							5	0
24	6EE.1	5	4	2			6EE.2	1	*	6EE.3	1	*										7	0
25	6EE.1	5	4	2																		5	0
26	6EE.1	2	1	2	6EE.4	2	2	6EE.2	1	2	6EE.3	1	*									6	0
27	6EE.2	3	2				6EE.3	1	2	NS	1	**										5	0

Reference

Irvin, P. S., Bitnara, J. P., Alonzo, J., & Tindal, G. (2012). *The alignment of the easyCBM® grades 6–8 math measures to the Common Core Standards (Technical Report No. 1230)*. Behavioral Research and Teaching, Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1208: The Alignment of the easyCBM® Middle School Mathematics CCSS Measures to the Common Core State Standards (Anderson et al., 2012).

Methods

Overall alignment was high. Of the 1,345 items reviewed, 1,180 (87.73%) received an adjusted MFRM rating at or above 2.0, indicating alignment to their corresponding CCSS standard. Of the 165 items rated as not aligned, 160 (97.00%) were judged by consensus to address a requisite skill to the standard. Combined, 99.6% of all items aligned either directly to a standard or to a requisite skill. Rater severity varied substantially across the 15 raters, with the most severe rater scoring items nearly a full category lower on average than the most lenient. Despite this range, all raters fit the MFRM expectations well, with mean square outfit statistics ranging from 0.76 to 1.16, indicating consistent rating behavior. The exploratory analysis found minimal variation in ratings attributable to item domain (5% of variance) or grade level (0.25% of variance), suggesting raters applied the alignment criteria consistently across content areas and grades.

Results

Overall alignment was high. Of the 1,345 items reviewed, 1,180 (87.73%) received an adjusted MFRM rating at or above 2.0, indicating alignment to their corresponding CCSS standard. Of the 165 items rated as not aligned, 160 (97.00%) were judged by consensus to address a requisite skill to the standard. Combined, 99.6% of all items aligned either directly to a standard or to a requisite skill. Rater severity varied substantially across the 15 raters, with the most severe rater scoring items nearly a full category lower on average than the most lenient. Despite this range, all raters fit the MFRM expectations well, with mean square outfit statistics ranging from 0.76 to 1.16, indicating consistent rating behavior. The exploratory analysis found minimal variation in ratings attributable to item domain (5% of variance) or grade level (0.25% of variance), suggesting raters applied the alignment criteria consistently across content areas and grades.

Table 5. Illustrative Table of Key Findings from Technical Report 1208

Item	ScoreTot	Obs. Avg.	Adj. Avg.	Endorsability	S.E.	Fit Statistics			
						InfitMS	InfitZ	OutfitMS	OutfitZ
6EE1003	9	3	2.91	-3.8	1.85	1	0	1	0
6EE1004	9	3	2.9	-3.64	1.85	1	0	1	0
6EE1005	9	3	2.86	-3.28	1.85	1	0	1	0
6EE1007	8	2.67	2.51	-1.93	1.06	.49	-.28	.43	-.3
6EE1011	9	3	2.9	-3.64	1.85	1	0	1	0
6EE1012	9	3	2.83	-3.12	1.85	1	0	1	0
6EE1013	9	3	2.91	-3.8	1.85	1	0	1	0
6EE1014	9	3	2.85	-3.21	1.85	1	0	1	0
6EE1017	9	3	2.83	-3.12	1.85	1	0	1	0
6EE1020	9	3	2.86	-3.28	1.85	1	0	1	0
6EE1023	9	3	2.91	-3.8	1.85	1	0	1	0
6EE1024	9	3	2.9	-3.64	1.85	1	0	1	0
6EE2002	7	2.33	2.11	-1.18	.82	.17	-1.68	.2	-1.41
6EE2003	8	2.67	2.51	-1.93	1.06	.49	-.28	.43	-.3
6EE2007	8	2.67	2.47	-1.85	1.05	.54	-.2	.48	-.24
6EE2008	9	3	2.9	-3.64	1.85	1	0	1	0
6EE2009	9	3	2.85	-3.21	1.85	1	0	1	0
6EE2012	9	3	2.91	-3.8	1.85	1	0	1	0
6EE2015	8	2.67	2.47	-1.85	1.05	.54	-.2	.48	-.24
6EE2016	8	2.67	2.54	-2.01	1.06	.68	-.01	.61	-.04
6EE2017	9	3	2.91	-3.8	1.85	1	0	1	0
6EE2019	9	3	2.9	-3.64	1.85	1	0	1	0
6EE2022	8	2.67	2.54	-2.01	1.06	.68	-.01	.61	-.04
6EE2024	9	3	2.85	-3.21	1.85	1	0	1	0
6EE3001	9	3	2.9	-3.64	1.85	1	0	1	0
6EE3003	7	2.33	2.02	-1.02	.81	1.73	1.09	1.65	.99
6EE3006	9	3	2.91	-3.8	1.85	1	0	1	0
6EE3008	9	3	2.85	-3.21	1.85	1	0	1	0
6EE3011	9	3	2.86	-3.28	1.85	1	0	1	0
6EE3012	9	3	2.83	-3.12	1.85	1	0	1	0
6EE3014	9	3	2.91	-3.8	1.85	1	0	1	0
6EE3016	9	3	2.9	-3.64	1.85	1	0	1	0
6EE3017	9	3	2.85	-3.21	1.85	1	0	1	0

Note: Endorsability reported on logit scale, with higher values indicating a "harder" to endorse item.

Reference

Anderson, D., Irvin, P. S., Alonzo, J., & Tindal, G. (2012). *The Alignment of the easyCBM® middle school mathematics CCSS measures to the Common Core State Standards (Technical Report 1208)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 2101: The Alignment Between easyCBM[®] Mathematics Assessments and State and National Standards (Saez et al., 2021).

Methods

All 50 U.S. states' K–8 Mathematics content standards served as the subject of analysis, using the Common Core State Standards (CCSS) as the primary referent. Data were drawn from individual state department of education websites, corestandards.org, easyCBM[®] math item development files, easyCBM[®] test items (lite and district versions), and the easyCBM[®] user manual. Both easyCBM[®] Basic Math (formerly NCTM Math) and easyCBM[®] Proficient Math (formerly CCSS Math) were included. The Proficient Math measures were developed in 2012, after the release of the CCSS, and were explicitly designed to reflect CCSS expectations across grades K–8. Mathematics standards were gathered from November 2020 through March 2021. States were classified into three groups: CCSS Adopted (20 states), CCSS Revised (28 states), and State Unique (2 states: Texas and Virginia). Standards text was input into a multi-tabbed Excel file covering all 11 CCSS mathematics domains.

Using archived math development alignment files, an alignment dataset was constructed to document item coverage across all CCSS clusters and topics. A minimum threshold required coverage on at least 3 of 3 benchmark forms. For standards with only 1–2 benchmark forms covered, follow-up analysis of progress monitoring (PM) coverage was conducted. Alignment was rated on a four-level scale: Strong (all 3 benchmark forms), Moderate (2 benchmark forms + $\geq 50\%$ of PM forms), Limited (2 benchmark forms + $< 50\%$ of PM forms), or Insufficient (< 2 benchmark forms). A second reviewer verified alignment classifications, and disagreements were resolved through consensus discussion.

Results

The easyCBM[®] mathematics assessments demonstrated strong and broad alignment with state mathematics standards, particularly for states that had fully or partially adopted the CCSS. Because the Proficient Math measures were explicitly developed with CCSS alignment in mind, coverage across the 11 mathematics domains was systematically documented and generally robust.

For the 20 states that fully adopted the CCSS (e.g., Colorado, Oregon, Washington, Maryland), alignment results were uniform and applied equally across all states in this group. The Proficient Math measures showed Strong to Moderate alignment across all 11 CCSS domains, with the strongest alignment found in domains most heavily represented in the item development process, such as Operations and Algebraic Thinking, Numbers & Operations in Base 10, and Measurement & Data. The 28 CCSS Revised states showed largely consistent alignment with the CCSS-based findings, given the substantial overlap between their standards and the CCSS. State-specific ADDITIONAL standards—those extending beyond the CCSS framework—were rated separately. Most ADDITIONALS reflected off-grade associations with CCSS clusters (e.g., first-grade Counting and Cardinality content appearing in kindergarten standards) and were generally rated Limited, as easyCBM[®] items partially but not fully captured these extended competencies. Texas and Virginia, whose standards substantially deviate from the CCSS, required individual alignment analyses. Given the limited overlap between their frameworks and the CCSS, alignment ratings for non-CCSS standards were more frequently Limited to Insufficient. Educators in these states should consult the state-specific Google Sheets datasets to identify which easyCBM[®] items and domains best correspond to their standards.

Across all states, the 11 CCSS mathematics domains were reviewed: Geometry; Measurement & Data; Counting & Cardinality; Operations and Algebraic Thinking; Numbers & Operations in Base 10; Numbers & Operations—Fractions; The Number System; Ratios & Proportional Relations; Expressions & Equations; Statistics & Probability; and Functions. Domains assessed across multiple grade levels (e.g., Geometry, Measurement & Data) showed more robust item coverage, while upper-grade domains (e.g., Functions, Statistics & Probability) showed more variability in alignment strength given the narrower grade band in which they appear.

These findings indicate that easyCBM[®] mathematics assessments—particularly the Proficient Math measures—provide well-aligned measurement tools for educators in CCSS-aligned states. The explicit CCSS-based development of these measures supports their use as progress monitoring and benchmark tools in states with full or modified CCSS adoption. Users in State Unique contexts should consult the detailed Google Sheets results to determine which domains and clusters offer adequate alignment support.

Reference

Saez, L., Whitney, M., Swanson, D., & Alonzo, J. (2021). *The alignment between easyCBM® mathematics and literacy assessments and state and national standards (Technical Report # 2101)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1006: Technical Adequacy of the easyCBM® Primary-Level Mathematics Measures (Grades K–2), 2009–2010 Version (Anderson, Lai, et al., 2010).

Methods

This report examined the technical adequacy of the easyCBM® mathematics benchmark assessments across Kindergarten, Grade 1, and Grade 2 using data collected throughout the 2009–2010 school year. Criterion and construct validity analyses used a nationally stratified random sample drawn from 76 schools across 26 states.

For criterion validity, a national stratified random sample across 76 schools in 26 states was used. The TerraNova 3 mathematics assessment, administered in May, served as the criterion. Regression and correlation analyses were run in four models at each grade: a full model (all three seasonal scores as predictors) and three seasonal models. Fall and winter models were interpreted as predictive validity evidence, the spring model as concurrent validity evidence. Construct validity was assessed through Rasch item fit analyses and confirmatory factor analysis (CFA), testing a unidimensional model against a three-factor alternative.

Results

Criterion-related validity was evaluated using regression and correlation analyses with the TerraNova 3 mathematics assessment as the external criterion. Predictive validity was assessed using fall and winter easyCBM® scores, while concurrent validity was examined using spring scores. Full regression models were significant at all three grade levels, explaining 53% (K), 59% (Gr. 1), and 66% (Gr. 2) of TerraNova variance. Individual seasonal models were also all significant. Concurrent validity (spring model) accounted for approximately 52–53% of TerraNova variance across grades. Predictive validity (fall and winter models) explained between 27% and 54% of TerraNova variance. Rasch and CFA results supported the unidimensional structure of the assessments, with items generally fitting the model well.

Construct validity was investigated through Rasch item-fit analyses and confirmatory factor analysis. Item outfit statistics largely fell within acceptable ranges, supporting unidimensionality. CFA results further indicated that a unidimensional model fit the data as well as or better than competing three-factor models aligned with NCTM focal points. Overall, the findings provide strong evidence supporting the technical adequacy of the easyCBM® K–2 mathematics measures for educational decision-making

Table 6. Example Summary of Key Findings from Technical Report 1006

Table 82
Correlations

		TerraNova3 Math Scale Score	Fall09T ot	Wint10T ot	Spmg10T ot
Pearson Correlation	TerraNova3 Math Scale Score	1.000	.594	.509	.653
	Fall09Tot	.594	1.000	.529	.478
	Wint10Tot	.509	.529	1.000	.661
	Spmg10Tot	.653	.478	.661	1.000
Sig. (1-tailed)	TerraNova3 Math Scale Score	.	.000	.000	.000
	Fall09Tot	.000	.	.000	.000
	Wint10Tot	.000	.000	.	.000
	Spmg10Tot	.000	.000	.000	.
N	TerraNova3 Math Scale Score	153	153	153	153
	Fall09Tot	153	153	153	153
	Wint10Tot	153	153	153	153
	Spmg10Tot	153	153	153	153

Reference

Anderson, D., Cheng-Fei, L., Nese, J. F. T., Bitnara, J. P., Sáez, L., Jamgochian, E., Alonzo, J., & Tindal, G. (2010). *Technical Adequacy of the easyCBM® Primary-Level Mathematics Measures (Grades K–2), 2009–2010 Version (Technical Report 1006)* Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1007: Technical Adequacy of the easyCBM® Mathematics Measures: Grades 3–8, 2009–2010 version. (Nese, Lai, Anderson, Jamgochian, et al., 2010).

Methods

This report examined the technical adequacy of the easyCBM® mathematics progress monitoring assessments for students in grades 3 through 8. The study evaluated the reliability, validity, and practical utility of the 2009–2010 version of the easyCBM® mathematics measures, which are designed to support instructional decision-making within Response to Intervention (RTI) frameworks. The system includes seasonal benchmark assessments administered in fall, winter, and spring as well as alternate forms of progress monitoring measures used throughout the academic year. The study used data collected from four school districts located in two states: three districts in Oregon and one in Washington. All students present on the testing days were included in the analyses, providing large samples across grades 3–8. Researchers conducted several statistical analyses to evaluate different aspects of the measures’ technical properties. These analyses included estimation of minimum acceptable within-year growth rates, determination of minimum acceptable end-of-year benchmark performance levels, and examination of internal consistency and split-half reliability. Additional analyses evaluated the reliability of slope estimates derived from repeated progress-monitoring assessments. Validity evidence was examined through construct, concurrent, and predictive validity analyses, including correlations and regression analyses comparing easyCBM® scores with year-end state mathematics test results.

Results

Results indicated that the easyCBM® mathematics measures demonstrate strong reliability and validity across grades and samples. Internal consistency and split-half reliability estimates supported the stability of the measures, while slope reliability analyses suggested that growth estimates derived from repeated administrations were dependable indicators of student progress. Validity analyses showed moderate to strong correlations between easyCBM® scores and state mathematics assessment outcomes across grades and demographic groups. Regression analyses further demonstrated that fall and winter easyCBM® scores were strong predictors of year-end state mathematics performance, while spring scores showed strong concurrent validity with the same outcomes. Overall, the findings support the technical adequacy of the easyCBM® mathematics measures for monitoring student progress and predicting performance on state mathematics assessments.

Table 7. Illustrative Table of Key Findings from Technical Report 1007

Oregon Predictive Validity for White Students in Grade 4, Regressing Winter easyCBM® Math Benchmark on Year-End State Math Test

Descriptive Statistics			
	Mean	Std. Deviation	N
OAKSMathTot	220.13	9.918	1265
wint_tot	32.6798	6.04462	1265

Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics			Sig. F Change	
					R Square Change	F Change	df1		df2
1	.733 ^a	.538	.537	6.745	.538	1469.772	1	1263	.000

a. Predictors: (Constant), wint_tot

Coefficients ^a											
Model		Unstandardized Coefficients		Standardized Coefficients		95.0% Confidence Interval for B			Correlations		
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part
1	(Constant)	180.804	1.043		173.336	.000	178.757	182.850			
	wint_tot	1.203	.031	.733	38.338	.000	1.142	1.265	.733	.733	.733

a. Dependent Variable: OAKSMathTot

Reference

Nese, J. F. T., Cheng-Fei, L., Anderson, D., Jamgochian E. M., Kamata, A., Sáez, L., Bitnara J. P., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM[®] mathematics measures: grades 3–8, 2009–2010 version (Technical Report 1007)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1008: Diagnostic Efficiency of easyCBM[®] Math: Washington State (Anderson, Alonzo, et al., 2010b).

Methods

This technical report examined the diagnostic efficiency of the easyCBM[®] mathematics benchmark assessments for predicting student performance on the Washington State mathematics assessment. The purpose of the study was to determine optimal cut scores on the easyCBM[®] measures that could accurately classify students as likely to meet or not meet proficiency standards on the state test. The analysis focused on students in grades 3 through 8 and evaluated how effectively seasonal easyCBM[®] benchmark scores predicted outcomes on the statewide assessment.

Data for the study were collected from three school districts in Washington State that had implemented district-wide Response to Intervention (RTI) models. All students present during the scheduled assessment periods were included in the analyses, and the sample included students across grades 3–8 as well as various demographic subgroups, including English language learners and students receiving special education services.

Two primary measures were used. The predictor variable was the easyCBM[®] mathematics benchmark assessment, a computer-administered test consisting of 45 multiple-choice items per form. The system includes 13 equivalent forms per grade level that were calibrated using a one-parameter Rasch model to ensure comparable difficulty across administrations. Three of these forms are used for seasonal benchmark screening (fall, winter, and spring), while the remaining forms are used for progress monitoring throughout the year. The outcome measure was the Washington Measures of Student Progress (MSP) mathematics assessment, the statewide accountability test used to determine student proficiency.

Receiver Operating Characteristic (ROC) curve analyses were conducted to evaluate the diagnostic accuracy of easyCBM[®] scores in predicting MSP proficiency outcomes. These analyses were used to estimate classification accuracy statistics and determine optimal cut scores for each grade level and benchmark period. Results were reported separately by grade and season and were also examined across demographic subgroups.

Results

The results indicated that the easyCBM[®] mathematics benchmark assessments demonstrated strong diagnostic efficiency for predicting student performance on the Washington state mathematics test. ROC analyses showed that easyCBM[®] scores were effective at distinguishing between students who ultimately met proficiency standards and those who did not. Optimal cut scores were identified for each grade and seasonal benchmark, allowing educators to classify students into risk categories with acceptable levels of sensitivity and specificity.

Overall, the findings suggest that easyCBM[®] mathematics benchmarks can serve as effective screening tools within RTI systems, enabling educators to identify students at risk of failing the state assessment early in the school year and to monitor their progress toward proficiency over time.

Table 8. Illustrative Table of Key Findings from Technical Report 1008

Table 4
Resulting Statistics for Each Chosen Cut Score: Full Sample

Grd	Season	Meets Score	<i>n</i>	Failure Base Rate	False Positive Rate	False Negative Rate	Sensitivity	Specificity	Positive Predictive Power	Negative Predictive Power	Overall Correct Classification	AUC
3	Fall	31	594	.39	.25	.21	.79	.75	.67	.85	.77	.84
	Winter	36	721	.33	.29	.15	.85	.71	.59	.90	.75	.87
	Spring	39	923	.36	.35	.13	.88	.65	.59	.90	.73	.88
4	Fall	33	671	.37	.16	.17	.83	.84	.75	.89	.84	.90
	Winter	36	857	.37	.20	.16	.84	.80	.71	.89	.81	.90
	Spring	39	911	.38	.25	.12	.88	.75	.69	.91	.80	.93
5	Fall	33	640	.35	.19	.16	.84	.81	.71	.91	.82	.91
	Winter	37	776	.37	.16	.13	.87	.84	.76	.91	.85	.93
	Spring	42	1042	.39	.27	.11	.89	.73	.68	.91	.79	.92
6	Fall	32	829	.36	.27	.13	.87	.73	.65	.91	.78	.90
	Winter	33	830	.36	.18	.14	.86	.82	.72	.91	.83	.92
	Spring	38	1668	.39	.22	.11	.89	.78	.73	.92	.83	.94
7	Fall	29	753	.36	.18	.20	.80	.82	.72	.88	.82	.90
	Winter	29	772	.35	.17	.20	.80	.83	.71	.89	.82	.91
	Spring	34	1590	.41	.22	.11	.89	.78	.74	.91	.82	.93
8	Fall	32	513	.29	.20	.14	.86	.80	.63	.94	.82	.92
	Winter	35	636	.35	.25	.10	.90	.75	.66	.93	.81	.92
	Spring	35	1462	.40	.23	.20	.80	.77	.70	.86	.78	.91

Note. AUC = Area Under the ROC Curve

Reference

Anderson, D., Alonzo, J., & Tindal, G. (2009). *Diagnostic efficiency of easyCBM® math: Washington state (Technical Report 1008)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1009 Summary: Diagnostic efficiency of easyCBM® math: Oregon (Anderson, Alonzo, et al., 2010a).

Methods

The study examined the diagnostic efficiency of the easyCBM® mathematics benchmark assessments for predicting student performance on the Oregon Assessment of Knowledge and Skills (OAKS) state test. Participants included students in grades 3 through 8 from three Oregon school districts. Two districts implemented a district-wide Response to Intervention (RTI) framework and administered easyCBM® benchmark assessments to all students, including English language learners and students with disabilities. A third district administered the benchmark assessments to a representative subset of classes designed to mirror district demographics.

Two primary measures were used. The predictor measure was the easyCBM® mathematics benchmark assessment, a computer-administered test consisting of 45 multiple-choice items with multiple alternate forms equated using a Rasch (1PL) model. The criterion measure was the mathematics portion of OAKS, a computer adaptive statewide assessment reported on a Rasch scale and classified into performance levels. For analysis, OAKS performance was dichotomized as meeting or not meeting expectations. Receiver Operating Characteristic (ROC) curve analyses were conducted to evaluate diagnostic efficiency. Sensitivity, specificity, predictive power, and classification accuracy were calculated for all possible cut scores across fall, winter, and spring administrations. Optimal cut scores were determined using decision rules based on Silbergliitt and Hintze (2005), with emphasis on maximizing sensitivity while maintaining acceptable specificity.

Results

Results demonstrated that the easyCBM[®] mathematics benchmarks provided strong predictive accuracy for determining whether students would meet expectations on the Oregon state test. Across grades 3–8 and across seasonal benchmark administrations (fall, winter, and spring), the Area Under the ROC Curve (AUC) ranged from approximately .86 to .92, indicating high diagnostic accuracy. These values suggest that the easyCBM[®] measures were effective in distinguishing students who would meet or exceed state standards from those who would not. Optimal cut scores were established for each grade and season. These cut points balanced sensitivity and specificity while prioritizing the identification of students at risk of failing the state test. Sensitivity values were generally high, often exceeding .80 and in some cases approaching .90 or higher, meaning the measures were effective at correctly identifying students likely to meet state standards. Specificity values were also acceptable, typically above .70, indicating that most students identified as at risk were correctly classified.

The study also reported classification accuracy statistics including false positive and false negative rates, positive predictive power, negative predictive power, and overall correct classification. Overall classification accuracy ranged approximately from the low .70s to the mid .80s depending on grade and season. Negative predictive power values were particularly high, often above .95, indicating that students identified as not at risk were very likely to meet state expectations. Seasonal differences in cut scores were also observed. In lower grades, cut scores increased more noticeably across the school year. For example, in grade 3 the optimal cut score increased substantially from fall to spring, reflecting expected student learning growth. In contrast, upper grades showed smaller seasonal shifts in cut scores. This pattern may reflect slower growth rates in later grades or differences in scaling between easyCBM[®] and the state test.

Subgroup analyses were conducted by ethnicity and English language learner status. Although sample sizes varied, the general pattern of diagnostic efficiency remained consistent across groups. The findings therefore support the use of easyCBM[®] mathematics benchmarks as an early screening and progress monitoring tool within RTI frameworks, particularly for identifying students at risk of failing the state accountability assessment.

Table 9. Summary of Main Findings from Technical Report 1009

Finding	Evidence	Interpretation
Diagnostic accuracy	AUC values ranged from .86 to .92	easyCBM [®] strongly predicts performance on the Oregon state test
Sensitivity	Typically, .80–.93 across grades and seasons	Measures effectively identify students likely to meet standards
Specificity	Generally, above .70	Students flagged as at risk are usually correctly identified
Seasonal cut score change	Lower grades showed larger fall–spring increases	Reflects greater academic growth in early grades
Predictive power	Negative predictive power often above .95	Students identified as safe are very likely to meet expectations

Reference

Anderson, D., Alonzo, J., & Tindal, G. (2010). *Diagnostic efficiency of easyCBM[®] math: Oregon (Technical Report No. 1009)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1010: easyCBM[®] Mathematics Criterion Related Validity Evidence: Washington State Test (Anderson, Alonzo, et al., 2010d).

Methods

This report examined the criterion-related validity of the easyCBM[®] mathematics benchmark assessments in Grades 3–8 using Washington State's Measures of Student Progress (MSP) as the external criterion. The study was conducted in one medium-sized Washington school district. Sample sizes ranged from approximately 417 (Grade 8) to 673 (Grade 4) students per grade, with demographic data detailed by grade including percentages of English Language Learners, students receiving special education services, sex, and ethnicity.

The easyCBM[®] mathematics tests consisted of 45 multiple-choice items aligned to NCTM Focal Point Standards, administered in fall, winter, and spring. The MSP, newly implemented for 2009–2010, included multiple-choice and short-answer items and was administered at the end of the school year. Because the MSP was administered only once at year-end, fall and winter easyCBM[®] scores were analyzed as predictive validity evidence, while the spring administration served as concurrent validity evidence.

Four regression models were run at each grade level: a full model including all three seasonal easyCBM[®] scores simultaneously, and three individual seasonal models. Scatterplots were also produced for each grade and season, with vertical lines marking the 20th and 50th easyCBM[®] percentiles and a horizontal line marking the MSP proficiency cut score, allowing visual examination of classification accuracy.

Results

Results indicated a strong relationship between easyCBM[®] and the MSP across all grades. The full regression model accounted for between 59% and 75% of MSP variance depending on grade. Individual seasonal models explained between 48% and 67% of MSP variance. Pearson correlations between easyCBM[®] seasonal scores and the MSP were consistently high, generally ranging from approximately .68 to .83 across grades and seasons, with the strongest relationships observed in Grades 6–8. Examination of the scatterplots showed that even in fall, very few students scoring below the 20th easyCBM[®] percentile reached the MSP proficiency level, while most students above the 50th percentile did reach proficiency. The authors note that easyCBM[®] demonstrated stronger predictive accuracy for identifying students who would not reach proficiency than for those who would.

Table 10. Illustrative Table of Key Findings from Technical Report 1010

		Correlations			
		Washington State Assessment Scale Score	Fall09TotMath	Wint10TotMath	Spr10TotMath
Pearson Correlation	Washington State Assessment Scale Score	1.000	.682	.705	.689
	Fall09TotMath	.682	1.000	.735	.682
	Wint10TotMath	.705	.735	1.000	.710
	Spr10TotMath	.689	.682	.710	1.000
Sig. (1-tailed)	Washington State Assessment Scale Score	.	.000	.000	.000
	Fall09TotMath	.000	.	.000	.000
	Wint10TotMath	.000	.000	.	.000
	Spr10TotMath	.000	.000	.000	.
N	Washington State Assessment Scale Score	463	463	463	463
	Fall09TotMath	463	463	463	463
	Wint10TotMath	463	463	463	463
	Spr10TotMath	463	463	463	463

Reference

Anderson, D., Alonzo, J., & Tindal, G. (2009). *easyCBM® Mathematics Criterion Related Validity Evidence: Washington State Test (Technical Report 1010)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1011: easyCBM® Mathematics Criterion Related Validity Evidence: Oregon State Test (Anderson, Alonzo, et al., 2010c).

Methods

This technical report examined the criterion-related validity of the easyCBM® mathematics benchmark assessments by evaluating their relationship with the Oregon statewide mathematics assessment. The study investigated whether easyCBM® benchmark scores could predict student performance on the state test under the revised proficiency standards introduced for the 2010–2011 school year. The analyses focused on students in grades 3 through 8 and explored the strength of the relationship between seasonal benchmark scores and outcomes on the state assessment.

Data were collected from school districts in Oregon that administered the easyCBM® mathematics benchmark assessments as part of a Response to Intervention (RTI) framework. Students completed seasonal benchmark assessments (fall, winter, and spring) through the online easyCBM® system. These computer-based assessments include multiple equivalent forms designed for progress monitoring and screening within RTI systems. The sample included students across grades 3–8. Demographic information was reported by district and grade level. For example, in one district the grade-level samples included 1,311 students in Grade 3, 1,299 in Grade 4, 1,357 in Grade 5, 1,329 in Grade 6, and 1,262 in Grade 7. Student demographic variables included English language learner status, economic disadvantage (free or reduced lunch), special education participation, gender, and ethnicity.

To examine criterion-related validity, the study used both correlation and regression analyses to evaluate the relationship between easyCBM® mathematics scores and performance on the Oregon statewide mathematics assessment. The analyses focused on determining how well benchmark scores predicted whether students would meet the state proficiency standard.

Results

Results indicated meaningful relationships between easyCBM® benchmark scores and the Oregon statewide mathematics test. Correlation analyses demonstrated moderate to strong associations between the seasonal easyCBM® scores and the state test outcomes. Regression analyses further showed that easyCBM® scores provided significant predictive information regarding student proficiency status on the state assessment.

Overall, the findings provided evidence that the easyCBM® mathematics benchmarks function as valid indicators of student performance relative to Oregon's statewide mathematics standards. The results support the use of easyCBM® benchmark assessments as screening tools within RTI systems for identifying students who may be at risk of not meeting state proficiency expectations.

Table 11. Illustrative Table of Key Findings from Technical Report 1011

Model	Coefficients ^a					
	Unstandardized Coefficients		Standardized Coefficients			
	B	Std. Error	Beta	t	Sig.	
1	(Constant)	180.471	.580		311.140	.000
	fall_tot	1.065	.019	.694	55.366	.000

a. Dependent Variable: OAKSMathTot

Reference

Anderson, D., Alonzo, J., & Tindal, G. (2009). *easyCBM® Mathematics criterion related validity evidence: Oregon state test. Technical Report No. 1011*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1402 Summary: Criterion Validity Evidence for the easyCBM® CCSS Math Measures Grades 6-8 (Anderson et al., 2014).

Methods

This study investigated the criterion validity of the easyCBM® CCSS Math benchmark assessments for Grades 6–8 by examining their relationship with the Stanford Achievement Test, Tenth Edition (SAT-10).

Participants were randomly selected from one middle school in the Pacific Northwest. Researchers selected approximately 65 students per grade to ensure adequate statistical power. The final samples included 67 students in Grade 6, 63 in Grade 7, and 64 in Grade 8. Demographic information included gender, special education status, English language learner status, eligibility for free or reduced lunch, and ethnicity. All students in the school completed the winter easyCBM® CCSS Math benchmark assessment. Within one week of this benchmark administration, the randomly selected sample completed the SAT-10 mathematics test. Both assessments were administered online. The easyCBM® CCSS Math measures were developed to align with the Common Core State Standards and are used within Response to Intervention (RTI) systems for screening and progress monitoring. Each test form included multiple-choice items designed to measure mathematical skills aligned with the standards.

The SAT-10 forms administered were Intermediate 3 for Grade 6, Advanced 1 for Grade 7, and Advanced 2 for Grade 8. Each SAT-10 mathematics test contained 80 multiple-choice items, including 48 problem-solving items and 32 procedural mathematics items. Items were scored dichotomously and scaled using a Rasch model. To evaluate the relationship between the assessments, researchers conducted bivariate correlation analyses and simple linear regression analyses. Exploratory regression models were also estimated to examine whether demographic variables contributed additional explanatory power.

Results

The results indicated strong relationships between the easyCBM® CCSS Math benchmark and the SAT-10 mathematics scores across all grade levels. Correlation analyses showed coefficients of .82 for Grade 6, .77 for Grade 7, and .75 for Grade 8, indicating strong positive associations between the two measures.

Simple linear regression analyses demonstrated that easyCBM® scores were significant predictors of SAT-10 performance for all grades ($p < .001$). The easyCBM® measure explained 67% of the variance in SAT-10 scores for Grade 6, 59% for Grade 7, and 56% for Grade 8. Regression coefficients indicated that a one-point increase in easyCBM® score corresponded to approximately a 3.5 to 4.5 point increase in SAT-10 scale score, depending on grade level. Exploratory regression models that included demographic variables produced similar results. Most demographic variables were not significant predictors of SAT-10 performance, although special education status was significant for Grade 7. Overall, the findings indicated that easyCBM® CCSS Math scores were strongly related to SAT-10 scores, providing evidence that the assessments measure similar underlying mathematics constructs.

Table 12. Illustrative Table of Key Findings from Technical Report 1402

Table 3

Multiple Regression Results: Grade 6

Parameter	Estimate		95% CI		S _r ²
	Standardized	Raw	Lower	Upper	
Intercept	-	564.53	537.81	591.25	-
FRL	-0.07	-5.34	-17.34	6.66	0.004
Female	0.14	10.46	-0.89	21.80	0.017
ELL	0.06	8.05	-13.42	29.53	0.003
SPEd	0.02	2.67	-16.58	21.93	0.001
easyCBM®	0.82	3.98	3.18	4.79	0.494

Note. The overall model was significant, $F = 27.60$ (5, 61), $R^2 = 0.69$.

Reference

Anderson, D., Rowley, B., Alonzo, J., & Tindal, G. (2014). *Criterion validity evidence for the easyCBM[®] CCSS math measures: Grades 6-8 (Technical Report # 1402)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1104: A Cross-Validation of easyCBM[®] Mathematics Cut Scores in Oregon: 2009–2010 (Anderson et al., 2011a).

Methods

This study evaluated the diagnostic efficiency of the easyCBM[®] mathematics benchmark assessments for predicting performance on the Oregon statewide mathematics assessment. The analyses focused on determining cut scores that could accurately classify students as likely to meet or not meet the state proficiency standard. Participants were students in Grades 3 through 8 from school districts in Oregon that had implemented the easyCBM[®] assessment system within a Response to Intervention (RTI) framework. Students completed the easyCBM[®] mathematics benchmark assessments during the fall, winter, and spring screening periods. Each benchmark form consisted of 45 multiple-choice items designed to measure grade-level mathematics skills. The assessments were computer-administered and drawn from a larger pool of calibrated items developed using Rasch measurement procedures, which ensured comparable difficulty across forms.

The outcome measure was the Oregon statewide mathematics assessment, which served as the criterion measure for determining student proficiency. To evaluate the predictive accuracy of the easyCBM[®] benchmarks, the researchers used Receiver Operating Characteristic (ROC) curve analyses. These analyses allowed the researchers to estimate classification accuracy statistics and determine optimal benchmark cut scores. Diagnostic efficiency was evaluated through measures such as sensitivity, specificity, and overall classification accuracy across grades and benchmark seasons.

Results

The analyses showed that the easyCBM[®] mathematics benchmarks demonstrated strong diagnostic efficiency for predicting performance on the Oregon state mathematics assessment. ROC analyses indicated that the benchmark scores were effective at distinguishing between students who would meet proficiency standards and those who would not. Across grades and testing periods, Area Under the Curve (AUC) values were consistently high, indicating strong classification performance.

Optimal cut scores were identified for each grade and seasonal benchmark period. These cut scores allowed students to be classified into risk categories with acceptable levels of sensitivity (correctly identifying students at risk) and specificity (correctly identifying students not at risk). The results also showed that predictive accuracy generally improved across the school year, with the spring benchmark demonstrating the strongest relationship with the state assessment.

Overall, the findings indicate that easyCBM[®] mathematics benchmark scores provide meaningful information about students' likelihood of meeting state mathematics proficiency standards, supporting their use as screening tools within RTI systems.

Table 13. Illustrative Table of Key Findings from Technical Report 1104

Independent Samples Test (continued)										
		Levene's Test for Equality of Variances				t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
EconDsvntg	Equal variances assumed	.279	.597	-.266	2722	.791	-.005	.019	-.043	.032
	Equal variances not assumed			-.266	2718.968	.791	-.005	.019	-.043	.032
OAKS	Equal variances assumed	.157	.692	.300	3702	.764	.096	.322	-.535	.728
Math Tot	Equal variances not assumed			.300	3700.198	.764	.096	.322	-.535	.728
Fall	Equal variances assumed	1.077	.299	1.508	3923	.132	.308	.204	-.093	.709
easyCBM	Equal variances not assumed			1.508	3922.475	.132	.308	.204	-.093	.709
Wint	Equal variances assumed	.496	.481	-.067	2717	.947	-.017	.251	-.509	.475
easyCBM	Equal variances not assumed			-.067	2713.431	.947	-.017	.251	-.509	.476
Spring	Equal variances assumed	2.834	.092	.654	3710	.513	.126	.193	-.252	.504
easyCBM	Equal variances not assumed			.654	3709.839	.513	.126	.193	-.252	.504
PLC	Equal variances assumed	.865	.352	-.465	3739	.642	-.006	.013	-.031	.019
	Equal variances not assumed			-.465	3738.656	.642	-.006	.013	-.031	.019

Reference

Anderson, D., Alonzo, J., & Tindal, G. (2009). *A cross-validation of easyCBM® mathematics cut scores in Oregon: 2009–2010. Technical Report No. 1104*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1105: A Cross-Validation of easyCBM® Mathematics Cut Scores in Washington State: 2009–2010 Test (Anderson et al., 2011b).

Methods

The study drew on data from three Washington State school districts that had implemented district-wide Response to Intervention (RTI) programs, enrolling students in grades 3 through 8, including English language learners and students with learning disabilities. Two assessments provided the data: the easyCBM® mathematics benchmark (fall, winter, and spring administrations), and the Measures of Student Progress (MSP), Washington's statewide accountability test. easyCBM® forms were aligned to NCTM Focal Point Standards and scaled using a 1PL Rasch model; MSP scores classified students as below basic, basic, proficient, or advanced, then collapsed dichotomously into "meeting" (proficient/advanced) versus "not meeting" (below basic/basic) for analysis. The full sample was randomly split into two roughly equal groups using a Bernoulli random-value function in SPSS 18.0 ($p = 0.50$). Independent-samples t-tests verified demographic and achievement comparability across groups on ten subgroup variables. Receiver Operating Characteristic (ROC) curve analyses were then conducted at each grade for each group, with area under the curve (AUC) statistics compared via 95% confidence intervals.

Results

The random split produced two groups with closely matched demographic profiles. T-tests revealed few statistically significant differences in student subgroup composition or achievement between groups, supporting the assumption that the two samples were equivalent and that any differences in outcomes could be attributed to sampling or measurement error rather than systematic bias.

Across all grades and measurement occasions (fall, winter, and spring), the optimal cut scores derived independently for each group were strikingly similar, typically differing by no more than one to two points. In Grade 3, for example, fall benchmark meeting scores were 32 and 31 for Groups 1 and 2 respectively, converging to an identical score of 39 by spring. Grade 4 showed similarly small divergence: fall scores were 34 and 33, rising to an identical spring score of 39. Grades 5 through 8 followed comparable patterns, with most within-season differences between groups being one point or less. Grade 7 was a slight exception, with a two-point difference observed on the winter benchmark (29 vs. 31), though this was still within a narrow range. The general trajectory across all grades was an increase in the optimal meeting score from fall to spring, reflecting expected growth in mathematics achievement over the school year.

The most significant finding was that in no case did the AUC statistics differ significantly between the two randomly selected groups at any grade or measurement occasion. The AUC values themselves were consistently high across all grades, ranging from approximately 0.82 to 0.94 across grades and seasons, indicating strong overall discriminative accuracy. The overlapping 95% confidence intervals for AUC comparisons at every grade confirm that the two groups' ROC curves were statistically equivalent, lending strong validity evidence to the identified cut scores. ROC curve figures displayed visually similar curve shapes for Groups 1 and 2 at every grade level.

The study also examined whether strictly following the Silbergliitt and Hintze (2005) decision rules versus the modified rules would have affected stability. In most cases, the modified rules (which prioritized maximizing sensitivity while keeping specificity above 0.70) produced stable or more stable results. The Grade 6 fall benchmark was a notable case: for Group 1, no cut score achieved both sensitivity and specificity above 0.80, and the modified rules selected a meeting score of 32; for Group 2, a score of 30 met the higher threshold. Had the unmodified rules been applied uniformly, the two groups would have been only one point apart (31 vs. 30), but the authors argued the modified rules were appropriate given the RTI framework's priority in minimizing false negatives.

Overall, these findings provide compelling cross-validation evidence that the easyCBM® mathematics cut scores recommended for use in Washington State are stable across student samples, and that the diagnostic efficiency of the instrument is consistent and replicable.

Table 14. Illustrative Table of Key Findings from Technical Report 1105

		Independent Samples Test									
		Levene's Test for Equality of Variances				t-test for Equality of Means					
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
									Lower	Upper	
AmerInd/ AkNative	Equal variances assumed	30.605	.000	-2.744	1930	.006	-.020	.007	-.034	-.006	
	Equal variances not assumed			-2.762	1726.160	.006	-.020	.007	-.034	-.006	
Asian/ PacIslander	Equal variances assumed	.699	.403	-.418	1930	.676	-.006	.015	-.035	.023	
	Equal variances not assumed			-.418	1929.682	.676	-.006	.015	-.035	.023	
Black	Equal variances assumed	.452	.502	-.336	1930	.737	-.003	.010	-.023	.016	
	Equal variances not assumed			-.336	1929.968	.737	-.003	.010	-.023	.016	
Hispanic	Equal variances assumed	8.528	.004	1.458	1930	.145	.020	.014	-.007	.047	
	Equal variances not assumed			1.456	1901.980	.146	.020	.014	-.007	.047	
White	Equal variances assumed	.115	.735	.169	1930	.866	.004	.022	-.040	.048	
	Equal variances not assumed			.169	1927.770	.866	.004	.022	-.040	.048	
Multethnic	Equal variances assumed	.008	.928	.045	1930	.964	.001	.014	-.027	.028	
	Equal variances not assumed			.045	1927.239	.964	.001	.014	-.027	.028	
Decline	Equal variances assumed	2.779	.096	.833	1930	.405	.005	.006	-.006	.016	
	Equal variances not assumed			.831	1867.408	.406	.005	.006	-.006	.016	
SPED	Equal variances assumed	4.101	.043	-1.011	1930	.312	-.016	.016	-.047	.015	
	Equal variances not assumed			-1.012	1929.659	.312	-.016	.016	-.047	.015	
Female	Equal variances assumed	3.195	.074	1.298	1930	.195	.030	.023	-.015	.074	
	Equal variances not assumed			1.298	1927.373	.195	.030	.023	-.015	.074	
ELL	Equal variances assumed	7.802	.005	1.394	1930	.163	.014	.010	-.006	.034	
	Equal variances not assumed			1.391	1881.700	.164	.014	.010	-.006	.035	

Reference

Anderson, D., Alonzo, J., & Tindal, G. (2009). *A cross-validation of easyCBM® mathematics cut scores in Washington state: 2009–2010 test (Technical Report 1105)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1501: An Exploration of Differential Item Functioning with the easyCBM® Middle School Mathematics Tests: Grades 6–8 (Anderson et al., 2015).

Data were drawn from a large, extant dataset collected during the 2013–2014 school year across multiple states. The sample included middle school students in Grades 6–8 participating in seasonal (fall, winter, spring) easyCBM® mathematics benchmark assessments. Each test form contained 45 multiple-choice items aligned to the Common Core State Standards. Subgroups examined included gender (female/male), English language learners (ELL/non-ELL), race (non-White/White), ethnicity (Latino/non-Latino), and special education status (received/did not receive services). Data collection occurred through operational administration of benchmark assessments, with student demographic data linked to test performance. Differential item functioning (DIF) analyses were conducted using the

Mantel-Haenszel (MH) procedure with iterative purification. The raw total test score served as the matching criterion, allowing comparisons of item performance across focal and reference groups at equivalent ability levels.

Statistical analyses included calculation of MH odds ratios and log transformations, with confidence intervals. Items were classified using ETS criteria into A (negligible DIF), B (moderate DIF), or C (large DIF). Purification removed flagged items from the matching criterion and re-estimated DIF iteratively. Over 2,000 DIF evaluations were conducted across forms and groups. Overall findings indicated that most items functioned equivalently across groups, supporting fairness and validity.

Results

Results showed minimal differential item functioning across grades, seasons, and subgroup comparisons. Approximately 97% of items were classified as “A” (negligible DIF), while about 3% were “B” (moderate DIF), and very few were “C” (large DIF).

Patterns were consistent across gender, ELL status, race/ethnicity, and special education status. Many test forms showed no DIF items, and others contained only one or two flagged items. Some variability appeared in specific Grade 8 winter comparisons, but overall stability remained high. DIF does not imply bias but indicates items for review. The results suggest easyCBM® mathematics measures are largely free from systematic bias and appropriate for diverse populations. Flagged items were recommended for monitoring and potential revision rather than removal.

Conclusions

- Overall DIF Rates were ~97% negligible DIF; ~3% moderate; few were large. These results indicate strong evidence of fairness.
- Subgroup Comparisons were consistent across gender, ELL, race, and SPED indicating valid across populations.
- Form-Level Results demonstrated that many forms had zero DIF items, thus reflecting stable test construction.
- Flagged Items were small in the number of B/C items implying little need to monitor and revise.

Table 15. Illustrative Table of Key Findings from Technical Report 1501

Table 5: g6fall_ell

item	DIF Grade	Alpha	Alpha_LB	Alpha_UB	Beta	Beta_LB	Beta_UB
IGEE5014	A	1.12	1.02	1.22	0.11	0.02	0.20
IGEE5006	A	1.07	0.97	1.18	0.07	-0.03	0.17
IGEE7007	A	1.00	0.91	1.10	0.00	-0.09	0.09
IGEE1019	A	1.19	1.09	1.30	0.18	0.09	0.27
IGEE7015	A	1.03	0.94	1.12	0.03	-0.07	0.12
IGC1005	A	0.85	0.76	0.94	-0.17	-0.27	-0.07
IGC2024	A	0.86	0.88	1.04	-0.05	-0.13	0.04
IGC3010	A	1.00	0.91	1.09	-0.01	-0.09	0.08
IGC1033	A	0.89	0.81	0.98	-0.12	-0.21	-0.02
IGC4023	A	0.96	0.87	1.05	-0.05	-0.14	0.05
IGC3038	A	1.15	1.06	1.26	0.14	0.05	0.23
IGNS2017	A	0.96	0.87	1.06	-0.04	-0.14	0.06
IGNS4010	A	0.96	0.87	1.06	-0.04	-0.13	0.06
IGNS8016	A	1.12	1.02	1.23	0.11	0.02	0.21
IGNS5011	A	1.00	0.91	1.10	0.00	-0.09	0.09
IGRP1006	A	1.03	0.93	1.13	0.03	-0.07	0.12
IGRP3006	A	0.93	0.85	1.03	-0.07	-0.17	0.03
IGRP1028	A	0.90	0.83	0.99	-0.10	-0.19	-0.01
IGRP3029	A	1.03	0.95	1.13	0.03	-0.06	0.12
IGRP2035	A	0.95	0.87	1.04	-0.05	-0.14	0.04
IGSP2025	A	1.00	0.92	1.10	0.00	-0.09	0.09
IGSP2034	A	0.97	0.89	1.06	-0.03	-0.12	0.06
IGSP1001	A	0.87	0.80	0.95	-0.14	-0.23	-0.05
IGSP4012	A	0.98	0.89	1.07	-0.03	-0.12	0.06
IGSP5008	A	1.07	0.98	1.17	0.07	-0.02	0.16
60097	A	0.82	0.72	0.93	-0.20	-0.33	-0.07
60265	A	1.07	0.97	1.19	0.07	-0.03	0.17
60300	A	1.12	1.01	1.23	0.11	0.01	0.21
60353	A	1.05	0.95	1.17	0.05	-0.05	0.15
60366	A	0.92	0.85	1.01	-0.08	-0.17	0.01
IGNS7011	A	1.05	0.94	1.17	0.05	-0.06	0.16
IGNS5015	A	0.91	0.83	1.00	-0.09	-0.19	0.00
IGNS5006	A	0.89	0.81	0.98	-0.11	-0.21	-0.02
IGRP1023	A	0.88	0.81	0.96	-0.13	-0.22	-0.04
IGEE2010	A	0.90	0.80	1.10	-0.01	-0.11	0.10
ITC3019	A	1.02	0.90	1.15	0.02	-0.10	0.14
ITRP2003	A	0.88	0.79	0.97	-0.13	-0.23	-0.03
ITRP2005	A	0.96	0.88	1.05	-0.04	-0.13	0.05
ITNS3059	A	1.03	0.94	1.12	0.03	-0.06	0.12
ITRP1047	A	1.04	0.94	1.14	0.04	-0.06	0.13
65GA34	A	1.10	0.95	1.26	0.09	-0.05	0.23
65NBT25	A	1.11	0.96	1.29	0.11	-0.04	0.26
65OA11	A	1.07	0.95	1.19	0.06	-0.05	0.17
65G412	A	0.91	0.83	0.99	-0.10	-0.19	-0.01
65MD415	A	0.91	0.83	1.00	-0.10	-0.19	-0.01

Reference

Anderson, D., Park, S., Alonzo, J., & Tindal, G. (2015). *An exploration of differential item functioning with the easyCBM[®] middle school mathematics tests: Grades 6–8 (Technical Report 1501)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 2401: Criterion Validity and Classification Accuracy of easyCBM[®]: Grades 3–8 (Tindal & Nese, 2024).

Methods

Technical Report 2401 evaluates criterion validity and classification accuracy for the easyCBM[®] Math Benchmark in Grades 3–8, using the Smarter Balanced (SB) Mathematics test as the external criterion. The study uses well over 8,000 anonymized student records from two states and four school districts. Students completed easyCBM[®] Math benchmarks in Fall 2023, Winter 2024, and Spring 2024, and SB Math was administered in spring. The report first presents grade-by-season descriptive statistics for easyCBM[®] Math and SB Math.

For predictive/concurrent validity, the authors report grade-specific correlations between easyCBM[®] Math and SB Math for each season (fall and winter as predictive; spring as near-concurrent). To support comparisons across grades, an easyCBM[®] Math composite is also computed by converting each student's Math score to a grade-level z score (subtract the grade mean and divide by the grade SD) in each season; correlations and regression plots are then presented between the Math composite and SB Math.

For classification accuracy, SB Math serves as the criterion and “risk” is defined primarily as performance at the 20th percentile. For each grade and season, the report provides ROC evidence (AUC with 95% confidence intervals) and grade/season cut scores, along with confusion-matrix counts (true/false positives and negatives). Summary tables report base rates, overall classification rates, sensitivity, specificity, false positive/negative rates, and positive/negative predictive power aligned to NCII screening guidance.

Results

Validity evidence indicates that easyCBM[®] Math benchmarks relate strongly to SB Math across Grades 3–8 and seasons. In the grade-level correlation tables, the Math–SB Math correlations are consistently high (generally in the upper .60s to high .80s). Specifically, the reported coefficients range from about .69 (Grade 3 fall) up to .88 (Grade 6 spring), with many values in the .74–.86 range. The composite summary further supports these relations: the easyCBM[®] Math composite correlates .57 (fall), .59 (winter), and .61 (spring) with SB Math, and the corresponding regression plots show clear positive linear trends, indicating that higher benchmark Math performance is associated with higher SB Math scores.

Descriptive statistics show expected seasonal growth. For example, Grade 3 mean Math scores increase from 24.52 (fall) to 28.76 (winter) to 31.82 (spring), and similar upward shifts appear across grades (with score ranges spanning much of each grade's possible scale). SB Math distributions by grade show substantial variability (SDs roughly mid-80s to ~120 scale-score points), supporting the use of continuous models and ROC analyses. Classification accuracy results for Math (risk defined at the SB 20th percentile) are generally strong. In fall, AUC values range from .81 to .92 across grades (highest in Grade 4 at .92), with specificity often very high (.79–.93) and sensitivity typically moderate (.63–.77). In winter, AUC values improve and cluster tightly between .88 and .92 across grades; sensitivity is about .76–.81 and specificity about .83–.91, yielding overall classification rates around .79–.82. In spring, AUC values are again very strong (.88–.94) with several grades at or above .92; sensitivity ranges from .71 to .91 and specificity from .81 to .93, and overall classification rates rise as high as .89 (Grade 5) and .86 (Grades 3 and 6).

Operating characteristics show the typical screening tradeoffs. Base rates for risk are modest (about .15–.24 depending on grade/season). False positive rates are usually low (often .07–.19), while false negative rates vary more (roughly .09–.37), with the larger false negative values concentrated in some upper grades or earlier seasons. Positive predictive power is consistently very high—most often .93–.98 across grades and seasons—meaning

students flagged “at risk” are very likely to fall below the SB cut. Negative predictive power is lower (often about .35–.66), reflecting the combination of base rates and the sensitivity/specificity balance. Overall, TR2401 concludes that easyCBM® Math benchmarks provide credible criterion validity with SB Math and strong classification accuracy for screening when grade- and season-specific cut scores are used.

Cut scores reflect growth: for example, Grade 3 uses an easyCBM® Math cut of 23 (fall), 27 (winter), and 30 (spring) for the SB 20th-percentile criterion, while Grade 8 uses 21, 24, and 25 across the same seasons. SB Math cut scores also rise with grade (2362 in Grade 3 to 2434 in Grade 8). Overall classification rates are typically in the mid-.70s to high-.80s (about .68–.89). AUC confidence intervals are extremely tight for large within-grade samples.

Reference

Tindal, G. & Nese, J. F. T. (2024). *Criterion validity and classification accuracy of easyCBM®: Grades 3-8. (Technical Report 2401)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Appendix A: Technical Report Table Titles

Table 1. Illustrative Results from Technical Report 1002

Table 2. Illustrative Table of Key Findings from Technical Report 1228

Table 3. Illustrative Table of Key Findings from Technical Report 1229

Table 4. Illustrative Table of Key Findings from Technical Report 1230

Table 5. Illustrative Table of Key Findings from Technical Report 1208

Table 6. Example Summary of Key Findings from Technical Report 1006

Table 7. Illustrative Table of Key Findings from Technical Report 1007

Table 8. Illustrative Table of Key Findings from Technical Report 1008

Table 9. Summary of Main Findings from Technical Report 1009

Table 10. Illustrative Table of Key Findings from Technical Report 1010

Table 11. Illustrative Table of Key Findings from Technical Report 1011

Table 12. Illustrative Table of Key Findings from Technical Report 1402

Table 13. Illustrative Table of Key Findings from Technical Report 1104

Table 14. Illustrative Table of Key Findings from Technical Report 1105

Table 15. Illustrative Table of Key Findings from Technical Report 1501

Appendix B: Guide to Spreadsheet Technical Report Value Displays

See Riverside Insights or BRT to access exact values for TR Summaries
2603-VK8M_ValidityMathTables.xlsx

- TR1002... See Technical Report 2603-AK8RM...Page 9
- TR1228...See Technical Report 2603-AK8RM...Page 13
- TR1229...See Technical Report 2603-AK8RM...Page 17
- TR1230...See Technical Report 2603-AK8RM...Page 20
- TR1208...See Technical Report 2603-AK8RM...Page 24
- TR2101...See Technical Report 2603-AK8RM...Page 6
- TR1006
- TR1007
- TR1008
- TR1009
- TR1010
- TR1011
- TR1402
- TR1104
- TR1105
- TR1501
- TR2401

Technical Report References

- Anderson, D., Alonzo, J., & Tindal, G. (2010a). *Diagnostic efficiency of easyCBM math: Oregon (Technical Report # 1009)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Anderson, D., Alonzo, J., & Tindal, G. (2010b). *Diagnostic efficiency of easyCBM mathematics: Washington state (Technical Report # 1008)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Anderson, D., Alonzo, J., & Tindal, G. (2010c). *easyCBM mathematics criterion related validity evidence: Oregon state test (Technical Report # 1011)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Anderson, D., Alonzo, J., & Tindal, G. (2010d). *easyCBM mathematics criterion related validity evidence: Washington state test (Technical Report # 1010)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Anderson, D., Alonzo, J., & Tindal, G. (2011a). *A cross-validation of easyCBM mathematics cut scores in Oregon: 2009-2010 (Technical Report # 1104)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Anderson, D., Alonzo, J., & Tindal, G. (2011b). *A cross-validation of easyCBM mathematics cut scores in Washington state: 2009-2010 Test (Technical Report # 1105)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Anderson, D., Irvin, P. S., Alonzo, J., & Tindal, G. (2012). *The alignment of the easyCBM middle school mathematics CCSS measures to the Common Core State Standards (Technical Report # 1208)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Anderson, D., Lai, C. F., Nese, J. F. T., Park, B. J., Sáez, L., Jamgochian, E. M., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM primary-level mathematics measures (Grades K-2), 2009-2010 version (Technical Report # 1006)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Anderson, D., Park, S., Alonzo, J., & Tindal, G. (2015). *An exploration of differential item functioning with the easyCBM middle school mathematics tests: Grades 6-8 (Technical Report # 1501)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Anderson, D., Rowley, B., Alonzo, J., & Tindal, G. (2014). *Criterion validity evidence for the easyCBM CCSS math measures: Grades 6-8 (Technical Report # 1402)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Irvin, P. S., Park, B. J., Alonzo, J., & Tindal, G. (2012a). *The alignment of the easyCBM Grades 6-8 math measures to the Common Core Standards (Technical Report # 1230)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Irvin, P. S., Park, B. J., Alonzo, J., & Tindal, G. (2012b). *The alignment of the easyCBM Grades K-2 math measures to the Common Core Standards (Technical Report # 1228)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Nese, J. F. T., Lai, C. F., Anderson, D., Jamgochian, E. M., Kamata, A., Saez, L., Park, B. J., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM mathematics measures: Grades 3-8: 2009-2010 version (Technical Report # 1007)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Nese, J. F. T., Lai, C. F., Anderson, D., Park, B. J., Tindal, G., & Alonzo, J. (2010). *The alignment of easyCBM math measures to curriculum standards (Technical Report #1002)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Park, B. J., Irvin, P. S., Alonzo, J., & Tindal, G. (2012). *The alignment of the easyCBM grades 3-5 math measures to the Common Core Standards (Technical Report # 1229)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Saez, L., Whitney, M., Swanson, D., & Alonzo, J. (2021). *The alignment between easyCBM® mathematics and literacy assessments and state and national standards (Technical Report # 2101)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.
- Tindal, G., & Nese, J. F. T. (2024). *Criterion validity and classification accuracy of easyCBM: Grades 3-8. (Technical Report No. 2401)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.