

Technical Report # 2602

**Reliability Analyses for easyCBM[®] Measures in Grades 3-8: Total Scores
for Vocabulary, Proficient Reading, and Proficient Math**

Gerald Tindal

Joesph F. T. Nese

University of Oregon



behavioral research & teaching

Published by

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

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

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

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

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

Abstract

This technical report focuses on the reliability of three easyCBMs® (Vocabulary, Proficient Reading, and Proficient Math) used for benchmarking and screening students potentially at risk of failing to learn important reading and math skills. We document the samples of students drawn for analysis (demographics) and their performance on each of the measures, presenting traditional descriptive statistics. Finally, we present two types of reliability indices: Rasch marginal reliabilities and Cronbach's alpha for each measure and for fall and winter Benchmarks at each grade. The results indicate that all measures at both benchmark seasons, are reliable. Therefore, these three measures can be used for making important screening decisions.

***Reliability Analyses for easyCBM® Measures –
Grades 3-8: Vocabulary, Proficient Reading, and Proficient Math***

In this technical report we focus on two measures of reliability for three easyCBMs and document the results after initially presenting a description of the students sampled, both their demographics and their performance on these measures using traditional descriptive statistics. We then present reliability indices from two approaches: item response theory (IRT) and classical test theory (CTT).

Overview

Reliability describes the extent to which observed scores reflect true differences among students rather than random measurement error. Two common internal-consistency reliability approaches used in curriculum-based and interim assessment contexts are: (1) a model-based IRT/Rasch approach summarized as marginal reliability, and (2) a CTT approach summarized as Cronbach’s alpha. Both can be computed separately by grade and season (i.e., fall vs. winter) and used to evaluate whether scores are precise enough to support intended uses.

Rasch Marginal Reliability (IRT/EAP)

How it is calculated. Under a Rasch (1-parameter logistic) model for dichotomous items, each student is assumed to have an underlying ability (θ) that drives the probability of a correct response. The model yields an ability estimate for each student (often expected a posteriori, EAP) along with a standard error of that estimate (SE). A commonly used marginal reliability is computed as:

$$\text{Marginal reliability} = 1 - \text{mean}(SE^2) / \text{Var}(\theta)$$

This can be viewed as a ratio of error variance to total variance on the latent trait scale. The mean of SE^2 approximates average measurement error variance for θ estimates in the sample, and $\text{Var}(\theta)$ is the observed variance of the estimated abilities. Some workflows use bootstrapping to quantify uncertainty and report a median reliability with a 95% interval; the lower bound is useful for documenting minimum expected precision.

What it means. Rasch marginal reliability reflects the average precision of the latent trait (θ) estimates for the specific sample. If marginal reliability is 0.80, a useful interpretation is that most of the variability in θ across students represents real differences rather than estimation error—on average, given the fitted Rasch model and the sample’s score distribution.

How to interpret it. Because IRT provides conditional precision, marginal reliability summarizes (averages) information that varies across the score scale. Precision tends to be highest where test information is greatest (often around the middle of the distribution) and lower at extreme high/low performance. Marginal reliability is therefore especially informative when the tool is used for growth monitoring or when multiple forms are intended to be comparable.

Typical/acceptable values (rules of thumb):

- < 0.60: low—often insufficient for most decision-making; consider more items, better targeting, or model review.
- 0.60–0.70: marginal—may be usable for low-stakes group summaries; caution for individual decisions.
- ≥ 0.70 : commonly cited minimum for screening or interim use; some frameworks (e.g., NCII) use the lower confidence bound ≥ 0.70 .
- ≥ 0.80 : good for many instructional decisions; better support for individual-level interpretations.
- ≥ 0.90 : very high—often needed for high-stakes individual decisions; may require longer tests or strong item targeting.

Cautions for Rasch Marginal Reliability

Model dependence: the value is valid to the extent the Rasch model fits (unidimensionality, local independence, appropriate item functioning).

Sample dependence: restricted range (e.g., a very homogeneous grade sample) lowers $\text{Var}(\theta)$ and can reduce reliability even when items are good.

Targeting matters: if items are too easy/hard for the grade, SEs increase and marginal reliability drops.

Missingness: many missing responses can inflate SEs or reduce usable sample size; document n with valid item data.

Rasch constraint: equal discrimination is assumed; if items vary widely in discrimination, a 2PL model may fit better and change reliability.

Cronbach's Alpha (CTT Internal Consistency)

How it is calculated. Cronbach's alpha is based on the covariance structure among item scores and can be expressed in multiple equivalent ways. For a k-item test with item variances σ_i^2 and total-score variance σ_x^2 , alpha is:

$$\alpha = (k / (k - 1)) \times [1 - (\sum \sigma_i^2 / \sigma_x^2)]$$

With dichotomous items (0/1), alpha corresponds to KR-20. Conceptually, alpha estimates the reliability of a summed score under an internal-consistency framework: if items measure the same construct, students who do well on one item tend to do well on others, yielding a stable total score.

What it means. Alpha can be interpreted as the expected correlation between two parallel forms of the test (under strict assumptions), or as the degree to which the items hang together as a scale. An alpha of 0.80 suggests that about 80% of the observed total-score variance is attributable to true-score variance and 20% to random error, within the tested sample.

Typical/acceptable values (rules of thumb):

- < 0.60: low internal consistency; likely inadequate for most uses.
- 0.60–0.70: marginal; may suffice for exploratory/group-level summaries.
- ≥ 0.70 : commonly accepted minimum for screening and many instructional decisions.
- ≥ 0.80 : good for most individual instructional decisions.
- ≥ 0.90 : excellent; often expected for higher-stakes individual classification.

Cautions for Cronbach’s Alpha

Not a test of unidimensionality: alpha can be high even if a test measures multiple correlated dimensions.

Tau-equivalence assumption: alpha is most interpretable when items have similar true-score loadings; violations can under- or over-estimate reliability.

Length and redundancy: alpha increases with more items and with item redundancy; a very high alpha can indicate overly similar items rather than quality.

Sample dependence: restricted score variance reduces alpha; compare across grades/seasons with attention to score distributions.

Local dependence: item clusters (e.g., shared stimuli) can inflate alpha because errors correlate.

Conclusion

Reporting both Rasch marginal reliability and Cronbach’s alpha provides converging evidence. If both are ≥ 0.70 across grades and seasons, that supports a claim of adequate internal consistency for typical screening/interim purposes. If alpha is high but marginal reliability is notably lower, the test may produce consistent raw totals but be poorly targeted on the latent scale (large *SEs*) or violate IRT assumptions. If marginal reliability is high but alpha is lower, investigate item heterogeneity, missingness patterns, or whether the total score mixes item types that reduce internal consistency. In either case, follow up with dimensionality checks, item-fit review, and—when decisions involve cut scores—classification accuracy evidence.

Method

Settings and Subjects

Records from individual students who had taken any of the three measures were drawn from districts who allowed research to be conducted on their district data. The data draw included the following fields: Site, Student ID, Grade, disability, English language learner, Ethnicity, Race, Gender, Form (an internal check), Start, Finish, Test Grade, Use (Benchmark only), Benchmark Grade, and Academic Year. See descriptive statistics in tables. Following these fields, item responses were recorded in successive columns with data presented as 1 for correct and 0 for incorrect.

Measurement/Instrument Development

See technical reports provided on the BRT website (<http://brtprojects.org>) for information on the development of the Vocabulary, Proficient reading, and Proficient math measures.

Data Preparation and Analysis

The master file drawn from easyCBM® server contained all records for fall and winter benchmarks. This file was then separated into two separate files for Fall and Winter Benchmarks.

The analyses conducted for this study incorporated marginal reliability and Cronbach’s alpha as indices of score consistency. Within an item response theory (IRT) framework, marginal reliability evaluates how reliably the tests measure student ability across the population. Marginal reliability is calculated using IRT-based ability estimates and their associated standard errors to produce an overall reliability value (Samajima 1994). By weighting reliability by the distribution of student ability, marginal reliability provides a stable, interpretable index of overall

score consistency. Conceptually, marginal reliability reflects the proportion of observed variance in estimated latent ability that is attributable to true differences in ability rather than measurement error, defined as the variance of the ability estimates divided by the total of that variance plus the average error variance. In an IRT framework, error varies as a function of the latent ability, and the error variance can be integrated such that the marginal reliability for IRT scores is the ratio of the variance of the estimated latent abilities relative to the sum of the variance of the latent ability and the expected error variance.

Marginal reliability coefficients were estimated for each grade and season combination. We used the TAM (Robitzsch, Kiefer, & Wu, 2025) package in the R (R Core Team, 2025) environment to fit a Rasch model using marginal maximum likelihood, where the expected *a posteriori* (EAP) reliability estimate is defined as $v/(s + v)$ where v is the variance of theta estimates and s is the average of the squared error (Adams, 2005). To compute confidence intervals, a bootstrapping approach was applied with 500 bootstraps.

We also estimated Cronbach's alpha (Cronbach, 1951) as a measure of internal consistency, which estimates the proportion of observed score variance attributable to a common latent construct, under a tau-equivalence model. Alpha is estimated from the inter-item covariance structure and reflects the degree to which scale items measure a common underlying construct. We used the psych (Revelle, 2025) package to estimate Cronbach's alpha.

In addition to TAM (Robitzsch, Kiefer, & Wu, 2025) and psych (Revelle, 2025), the following R packages were also used in the creation of this report: data.table (Barrett et al., 2025), flextable (Gohel & Skintzos, 2025), here (Müller, 2025), janitor (Firke, 2024), tidymodels (Kuhn & Wickham, 2020), and tidyverse (Wickham et al., 2019).

Results

In the subsequent pages, results are presented for each of the measures and benchmark.

Reliability Analysis for Vocabulary Measures: Grades 3 - 8 Fall Benchmarks**Table 1. Demographics – Fall and Winter**

Key	Fall Count	Fall Percent	Winter Count	Winter Percent
Disability				
No	43,462	64.16	32,931	65.01
Yes	10,753	15.87	8,472	16.72
Missing	13,508	19.94	9,254	18.27
Other	20	.03	--	--
English Learner				
No	6,862	10.13	4,642	9.16
Yes	11,862	17.51	7,786	15.37
Missing	48,998	72.33	38,229	75.47
Other	21	.03	--	--
Gender				
Male	33,976	50.15	25,572	50.48
Female	31,989	47.22	23,796	46.97
NonBinary	8	.01	2	0
Other	1	0	0	0
Missing	1,769	2.61	1,287	2.54
Race				
White	39,457	58.25	29,428	58.09
Black/African American	8,845	13.06	7,545	14.89
Two or More Races	3,903	5.76	2,952	5.83
Asian	1,852	2.73	1,543	3.04
American Indian/Native Alaskan	652	0.96	381	0.75
Native Hawaiian/Pacific Islander	647	0.96	442	0.87
Other	481	0.71	0	0
Missing	11,906	17.58	8,017	15.83

Table 2. Descriptive Statistics – Fall

Grade	Count	Mean	SD	Min	Max	Median	Kurtosis	Skewness
3	7,657	14.30	4.49	0	20	16	2.43	-0.73
4	9,902	15.28	4.13	0	20	17	3.14	-0.98
5	9,601	15.81	3.58	0	20	17	4.24	-1.29
6	14,247	16.40	3.45	1	20	17	5.34	-1.60
7	13,264	16.47	3.63	0	20	18	4.62	-1.45
8	13,072	16.51	3.55	2	20	18	4.69	-1.44

Table 3. Descriptive Statistics – Winter

Grade	Count	Mean	SD	Min	Max	Median	Kurtosis	Skewness
3	5,418	15.16	4.33	0	20	17	3.12	-1.03
4	7,478	15.93	3.74	1	20	17	4.10	-1.26
5	7,665	15.84	3.52	2	20	17	4.28	-1.27
6	11,210	16.46	3.57	1	20	18	5.20	-1.58
7	9,588	16.39	3.47	1	20	17	4.89	-1.46
8	9,298	16.46	3.48	2	20	18	5.11	-1.53

For the Vocabulary measure, the lower bound of the confidence interval around the median estimate of marginal reliability exceeded 0.80 for all grades and seasons.

Table 4. Marginal Reliability Estimates: Vocabulary

Type of Reliability	Grade/Measure	<i>n</i>	Median Coefficient	95% Confidence Interval	95% Confidence Interval
				Lower Bound	Upper Bound
Marginal Reliability	Grade 3 Fall	7,657	0.81	0.81	0.82
Marginal Reliability	Grade 3 Winter	5,418	0.80	0.79	0.80
Marginal Reliability	Grade 4 Fall	9,902	0.79	0.79	0.8
Marginal Reliability	Grade 4 Winter	7,478	0.76	0.76	0.77
Marginal Reliability	Grade 5 Fall	9,601	0.75	0.75	0.76
Marginal Reliability	Grade 5 Winter	7,665	0.74	0.73	0.75
Marginal Reliability	Grade 6 Fall	14,247	0.74	0.73	0.74
Marginal Reliability	Grade 6 Winter	11,210	0.75	0.75	0.76
Marginal Reliability	Grade 7 Fall	13,264	0.76	0.75	0.76
Marginal Reliability	Grade 7 Winter	9,588	0.75	0.74	0.76
Marginal Reliability	Grade 8 Fall	13,072	0.75	0.74	0.75
Marginal Reliability	Grade 8 Winter	9,298	0.75	0.75	0.76

The lower bound of the confidence interval around the median estimate of Cronbach's alpha met or exceeded 0.80 for all grades and seasons except Grade 5 Winter (.79).

Table 5 . Cronbach's Alpha Estimates: Vocabulary

Type of Reliability	Grade/Measure	<i>n</i>	Median Coefficient	95% Confidence Interval	95% Confidence Interval
				Lower Bound	Upper Bound
Cronbach's alpha	Grade 3 Fall	7,657	0.85	0.85	0.86
Cronbach's alpha	Grade 3 Winter	5,418	0.86	0.85	0.86
Cronbach's alpha	Grade 4 Fall	9,902	0.85	0.84	0.85
Cronbach's alpha	Grade 4 Winter	7,478	0.83	0.82	0.83
Cronbach's alpha	Grade 5 Fall	9,601	0.81	0.81	0.82
Cronbach's alpha	Grade 5 Winter	7,665	0.80	0.79	0.80
Cronbach's alpha	Grade 6 Fall	14,247	0.82	0.81	0.82
Cronbach's alpha	Grade 6 Winter	11,210	0.83	0.83	0.84
Cronbach's alpha	Grade 7 Fall	13,264	0.84	0.83	0.84
Cronbach's alpha	Grade 7 Winter	9,588	0.82	0.81	0.83
Cronbach's alpha	Grade 8 Fall	13,072	0.83	0.82	0.83
Cronbach's alpha	Grade 8 Winter	9,298	0.83	0.82	0.84

Reliability Analysis for Proficient Reading Measures: Grades 3 - 8 Fall Benchmarks**Table 6. Demographics – Fall and Winter**

Key	Fall Count	Fall Percent	Winter Count	Winter Percent
Disability				
No	44,578	62.4	33,199	65.35
Yes	10,889	15.24	8,316	16.37
Missing	15,971	22.36	9,284	18.28
English Learner				
No	7,403	10.36	4,685	9.22
Yes	12,143	17	7,880	15.51
Missing	51,892	72.64	38,234	75.27
Gender				
Male	35,806	50.12	25,641	50.48
Female	33,789	47.3	23,942	47.13
NonBinary	10	0.01	3	0.01
Missing	1,833	2.57	1,213	2.39
Race				
White	32,529	45.53	24,564	48.36
Black/African American	8,965	12.55	7,641	15.04
Two or More Races	4,215	5.9	3,104	6.11
Asian	1,723	2.41	1,393	2.74
American Indian/Alaska Native	685	0.96	391	0.77
Native Hawaiian/Pacific Islander	644	0.9	411	0.81
Missing	22,677	31.74	13,295	26.17

Table 7. Descriptive Statistics – Fall

Grade	Count	Mean	SD	Min	Max	Median	Kurtosis	Skewness
3	8,450	10.76	3.97	0	20	11	2.10	0.00
4	10,705	11.72	4.28	0	20	12	1.97	-0.02
5	10,436	13.28	3.67	0	20	14	2.92	-0.78
6	14,660	13.70	3.53	0	20	15	3.26	-0.89
7	13,416	12.90	3.88	0	20	14	2.49	-0.51
8	13,771	12.94	3.83	0	20	14	2.50	-0.59

Table 8. Descriptive Statistics – Winter

Grade	Count	Mean	SD	Min	Max	Median	Kurtosis	Skewness
3	5,423	9.87	3.21	0	19	10	2.36	-0.04
4	7,315	12.58	4.02	1	20	13	2.22	-0.36
5	7,578	14.71	3.98	0	20	16	2.89	-0.87
6	11,155	13.01	3.50	0	20	14	3.04	-0.74
7	9,451	13.62	3.77	1	20	15	3.02	-0.90
8	9,877	12.02	3.57	0	20	13	2.53	-0.44

For the Proficient Reading measure, the lower bound of the confidence interval around the median estimate of marginal reliability exceeded 0.60 for all grades and seasons except Grade 3 Winter (.59). The lower bound of the confidence interval around the median estimate of marginal reliability exceeded 0.70 for all grades and seasons except Grade 6 Fall (.68) and Winter (.69()) and Grade 8 Winter (.68).

Table 9. Marginal Reliability Estimates: Proficient Reading

Type of Reliability	Grade/Measure	<i>n</i>	Median Coefficient	95% Confidence Interval	95% Confidence Interval
				Lower Bound	Upper Bound
Marginal Reliability	Grade 3 Fall	8,450	0.73	0.72	0.74
Marginal Reliability	Grade 3 Winter	5,423	0.60	0.59	0.61
Marginal Reliability	Grade 4 Fall	10,705	0.78	0.77	0.78
Marginal Reliability	Grade 4 Winter	7,315	0.75	0.75	0.76
Marginal Reliability	Grade 5 Fall	10,436	0.73	0.73	0.74
Marginal Reliability	Grade 5 Winter	7,578	0.77	0.76	0.77
Marginal Reliability	Grade 6 Fall	14,660	0.68	0.68	0.69
Marginal Reliability	Grade 6 Winter	11,155	0.69	0.69	0.70
Marginal Reliability	Grade 7 Winter	9,451	0.74	0.73	0.74
Marginal Reliability	Grade 7 Fall	13,416	0.75	0.74	0.75
Marginal Reliability	Grade 8 Fall	13,771	0.74	0.73	0.74
Marginal Reliability	Grade 8 Winter	9,877	0.69	0.68	0.69

And the lower bound of the confidence interval around the median estimate of Cronbach's alpha met or exceeded 0.60 for all grades and seasons except Grade 3 Winter (.59). The lower bound of the confidence interval around the median estimate of Cronbach's alpha exceeded 0.70 for all grades and seasons except Grade 8 Winter (.69).

Table 10. Cronbach's Alpha Estimates: Proficient Reading

Type of Reliability	Grade/Measure	<i>n</i>	Median Coefficient	95% Confidence Interval	95% Confidence Interval
				Lower Bound	Upper Bound
Cronbach's alpha	Grade 3 Fall	8,450	0.74	0.73	0.74
Cronbach's alpha	Grade 3 Winter	5,423	0.60	0.59	0.62
Cronbach's alpha	Grade 4 Fall	10,705	0.79	0.78	0.79
Cronbach's alpha	Grade 4 Winter	7,315	0.77	0.76	0.78
Cronbach's alpha	Grade 5 Fall	10,436	0.76	0.75	0.76
Cronbach's alpha	Grade 5 Winter	7,578	0.81	0.80	0.81
Cronbach's alpha	Grade 6 Fall	14,660	0.71	0.71	0.72
Cronbach's alpha	Grade 6 Winter	11,155	0.71	0.71	0.72
Cronbach's alpha	Grade 7 Winter	9,451	0.77	0.76	0.78
Cronbach's alpha	Grade 7 Fall	13,416	0.76	0.76	0.77
Cronbach's alpha	Grade 8 Fall	13,771	0.76	0.75	0.76
Cronbach's alpha	Grade 8 Winter	9,877	0.70	0.69	0.70

Reliability Analysis for Proficient Math Measures: Grades 3 - 8 Fall Benchmarks**Table 11. Demographics – Fall and Winter**

Key	Fall Count	Fall Percent	Winter Count	Winter Percent
Disability				
No	29,210	54.82	17,834	37.96
Yes	9,507	17.84	6,835	14.55
Missing	14,567	27.34	22,306	47.48
English Learner				
No	7,292	13.69	4,372	9.31
Yes	10,104	18.96	6,393	13.61
Missing	35,888	67.35	36,210	77.08
Gender				
Male	26,753	50.21	17,083	36.37
Female	25,341	47.56	15,938	33.93
NonBinary	9	0.02	3	0.01
Missing	1,181	2.21	13,951	29.7
Race				
White	27,333	51.3	17,232	36.68
Black/African American	5,774	10.84	4,953	10.54
Two or More Races	2,773	5.2	1,840	3.92
Asian	1,054	1.98	624	1.33
American Indian/Alaska Native	544	1.02	238	0.51
Native Hawaiian/Pacific Islander	176	0.33	101	0.22
Missing	15,630	29.33	21,987	46.81

Table 12. Descriptive Statistics – Fall

Grade	Count	N Items	Mean	SD	Min	Max	Median	Kurtosis	Skewness
3	8,032	40	24.51	5.98	1	40	24	2.88	0.03
4	9,600	40	25.01	7.45	3	40	25	2.13	-0.11
5	9,449	40	24.20	6.67	1	40	24	2.33	0.04
6	8,760	45	23.64	7.29	1	45	23	2.58	0.36
7	8,473	45	22.13	7.26	3	45	21	2.91	0.65
8	8,969	45	22.17	8.07	4	45	20	2.77	0.68

Table 13. Descriptive Statistics – Winter

Grade	Count	N Items	Mean	SD	Min	Max	Median	Kurtosis	Skewness
3	5,767	40	28.07	6.29	3	40	29	2.80	-0.54
4	7,121	40	26.26	6.53	3	40	27	2.52	-0.34
5	6,941	40	26.08	6.75	2	40	27	2.35	-0.24
6	5,157	45	25.86	8.26	6	45	26	2.10	0.08
7	3,942	45	23.03	8.60	4	45	21	2.45	0.57
8	4,734	45	24.23	8.43	4	45	23	2.27	0.36

For the Proficient Math measure, the lower bound of the confidence interval around the median estimate of marginal reliability exceeded 0.80 for all grades and seasons.

Table 14. Marginal Reliability Estimates: Proficient Math

Type of Reliability	Grade/Measure	<i>n</i>	Median Coefficient	95% Confidence Interval	95% Confidence Interval
				Lower Bound	Upper Bound
Marginal Reliability	Grade 3 Fall	8,032	0.81	0.81	0.82
Marginal Reliability	Grade 3 Winter	5,767	0.84	0.83	0.84
Marginal Reliability	Grade 4 Fall	9,600	0.86	0.86	0.87
Marginal Reliability	Grade 4 Winter	7,121	0.83	0.83	0.84
Marginal Reliability	Grade 5 Fall	9,449	0.83	0.82	0.83
Marginal Reliability	Grade 5 Winter	6,941	0.85	0.84	0.85
Marginal Reliability	Grade 6 Fall	8,760	0.83	0.82	0.83
Marginal Reliability	Grade 6 Winter	5,157	0.87	0.87	0.88
Marginal Reliability	Grade 7 Fall	8,473	0.82	0.82	0.83
Marginal Reliability	Grade 7 Winter	3,942	0.87	0.87	0.88
Marginal Reliability	Grade 8 Fall	8,969	0.86	0.85	0.86
Marginal Reliability	Grade 8 Winter	4,734	0.87	0.87	0.87

The lower bound of the confidence interval around the median estimate of Cronbach's alpha met or exceeded 0.80 for all grades and seasons.

Table 15. Cronbach's Alpha Estimates: Proficient Math

Type of Reliability	Grade/Measure	<i>n</i>	Median Coefficient	95% Confidence Interval	95% Confidence Interval
				Lower Bound	Upper Bound
Cronbach's alpha	Grade 3 Fall	8,032	0.81	0.80	0.81
Cronbach's alpha	Grade 3 Winter	5,767	0.85	0.84	0.85
Cronbach's alpha	Grade 4 Fall	9,600	0.87	0.87	0.87
Cronbach's alpha	Grade 4 Winter	7,121	0.84	0.83	0.84
Cronbach's alpha	Grade 5 Fall	9,449	0.83	0.82	0.83
Cronbach's alpha	Grade 5 Winter	6,941	0.85	0.85	0.86
Cronbach's alpha	Grade 6 Fall	8,760	0.83	0.82	0.83
Cronbach's alpha	Grade 6 Winter	5,157	0.88	0.87	0.88
Cronbach's alpha	Grade 7 Fall	8,473	0.82	0.82	0.83
Cronbach's alpha	Grade 7 Winter	3,942	0.88	0.87	0.88
Cronbach's alpha	Grade 8 Fall	8,969	0.86	0.85	0.86
Cronbach's alpha	Grade 8 Winter	4,734	0.87	0.87	0.88

References

- Adams, Raymond J. 2005. "Reliability as a Measurement Design Effect." *Studies in Educational Evaluation* 31 (2-3): 162–72.
- Barrett, Tyson, Matt Dowle, Arun Srinivasan, Jan Gorecki, Michael Chirico, Toby Hocking, Benjamin Schwendinger, and Ivan Krylov. 2025. *Data.table: Extension of 'Data.frame'*. <https://doi.org/10.32614/CRAN.package.data.table>.
- Cronbach, Lee J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16 (3): 297–334.
- Firke, Sam. 2024. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://doi.org/10.32614/CRAN.package.janitor>.
- Gohel, David, and Panagiotis Skintzos. 2025. *Flextable: Functions for Tabular Reporting*. <https://doi.org/10.32614/CRAN.package.flextable>.
- Kuhn, Max, and Hadley Wickham. 2020. *Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles*. <https://www.tidymodels.org>.
- Müller, Kirill. 2025. *Here: A Simpler Way to Find Your Files*. <https://doi.org/10.32614/CRAN.package.here>.
- R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Robitzsch, Alexander, Thomas Kiefer, and Margaret Wu. 2025. *TAM: Test Analysis Modules*. <https://doi.org/10.32614/CRAN.package.TAM>.
- Samajima, Fumiko. 1994. "Estimation of Reliability Coefficients Using the Test Information Function and Its Modifications." *Applied Psychological Measurement* 18 (3): 229–44.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- William Revelle. 2025. *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois: Northwestern University. <https://CRAN.R-project.org/package=psych>.