aimsweb PLUS

# Technical Manual

ALWAYS LEARNING

PEARSON

# Table of Contents


Introduction ..... 1

Standardization Sample ..... 1

Norm Sample ..... 3

Standardization Sample Descriptives ..... 5

Norm Sample Descriptives ..... 9

Reliability ..... 14

- Early Numeracy Equivalency Studies ..... 16
- Early Literacy Equivalency Studies ..... 17
- Math Equivalency Studies ..... 18
- Reading Equivalency Studies ..... 18
- Equivalency of Progress Monitoring Forms ..... 23

Validity ..... 45

- Early Numeracy Criterion Validity ..... 46
- Early Literacy Criterion Validity ..... 46
- Math Criterion Validity ..... 47
- Reading Criterion Validity ..... 49

Classification Accuracy ..... 51

- Early Numeracy Classification Accuracy ..... 52
- Early Literacy Classification Accuracy ..... 53
- Math Classification Accuracy ..... 55
- Reading Classification Accuracy ..... 58

References ..... 61

Appendix: aimswebPlus Measures ..... 63

Tables

- Table 1 Demographic Characteristics of the Standardization Sample ..... 2
- Table 2 Demographic Characteristics of the Norm Sample, Kindergarten and Grade 1 ..... 3
- Table 3 Demographic Characteristics of the Norm Sample for Math, Grades 2 Through 8 ..... 4
- Table 4 Demographic Characteristics of the Norm Sample for Reading, Grades 2 Through 8 ..... 4
- Table 5 Descriptive Statistics for Early Numeracy Measures in the Standardization Sample, Kindergarten and Grade 1 ..... 6
- Table 6 Descriptive Statistics for Early Literacy Measures in the Standardization Sample, Kindergarten and Grade 1 ..... 6
- Table 7 Descriptive Statistics for Math Measures in the Standardization Sample, Grades 2 Through 8 ..... 7

# Table of Contents *cont.*

Tables *cont.*


Table 8 Descriptive Statistics for Reading Measures in the Standardization Sample, Grades 2 Through 8 ..... 8
Table 9 Descriptive Statistics for Early Numeracy Measures and Composites in the Norm Sample, Kindergarten and Grade 1 ..... 9
Table 10 Descriptive Statistics for Early Literacy Measures and Composites in the Norm Sample, Kindergarten and Grade 1 ..... 10
Table 11 Descriptive Statistics for Math Measures and Composites in the Norm Sample, Grades 2 Through 8 ..... 11
Table 12 Descriptive Statistics for Reading Measures and Composites in the Norm Sample, Grades 2 Through 8 ..... 12
Table 13 Concepts & Applications Total Score Means and Quantiles, by Grade and Season ..... 13
Table 14 Oral Reading Fluency Words Read Correctly (WRC) Means and Lexiles, by Grade and Season ..... 13
Table 15 Benchmark Composite Scoring Rules, by Subject, Grade, and Season ..... 15
Table 16 Reliability of Early Numeracy Measures and Composites, Kindergarten and Grade 1 ..... 19
Table 17 Reliability of Early Literacy Measures and Composites, Kindergarten and Grade 1 ..... 20
Table 18 Reliability of Math Measures and Composites, Grades 2 Through 8 ..... 21
Table 19 Reliability of Reading Measures and Composites, Grades 2 Through 8 ..... 22
Table 20 Average Alternate-Form Reliability of Single ORF Stories, by Grade and Season ..... 22
Table 21 Average Alternate-Form Reliability of Single SRF Stories, by Grade and Season ..... 23
Table 22 Reliability of the Median of Three SRF Story Reading Rates, by Grade ..... 23
Table 23 Statistical Equivalency of Progress Monitoring Forms, Number Naming Fluency ..... 24
Table 24 Statistical Equivalency of Progress Monitoring Forms, Quantity Total Fluency ..... 25
Table 25 Statistical Equivalency of Progress Monitoring Forms, Quantity Difference Fluency ..... 26
Table 26 Statistical Equivalency of Progress Monitoring Forms, Number Comparison Fluency–Pairs ..... 27
Table 27 Statistical Equivalency of Progress Monitoring Forms, Math Facts Fluency–1 Digit ..... 28
Table 28 Statistical Equivalency of Progress Monitoring Forms, Math Facts Fluency–Tens ..... 29
Table 29 Statistical Equivalency of Progress Monitoring Forms, Letter Word Sounds Fluency ..... 29
Table 30 Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 1) ..... 30
Table 31 Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 2) ..... 31
Table 32 Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 3) ..... 32
Table 33 Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 4) ..... 33

# Table of Contents *cont.*

Tables *cont.*


Table 34 Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 5) ..... 34
Table 35 Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 6) ..... 35
Table 36 Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 7) ..... 36
Table 37 Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 8) ..... 37
Table 38 Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 2) ..... 38
Table 39 Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 3) ..... 39
Table 40 Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 4) ..... 40
Table 41 Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 5) ..... 41
Table 42 Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 6) ..... 42
Table 43 Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 7) ..... 43
Table 44 Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 8) ..... 44
Table 45 Early Numeracy Composite Score Predictive Validity Coefficients, by Grade, Season, and Criterion Measure ..... 46
Table 46 Early Numeracy Spring Composite Score Concurrent Validity Coefficients, by Grade and Criterion Measure ..... 46
Table 47 Early Literacy Predictive Validity Coefficients, by Grade, Season, and Criterion Measure ..... 47
Table 48 Early Literacy Concurrent Validity Coefficients, by Grade and Criterion Measure ..... 47
Table 49 Math Composite Score Predictive Validity Coefficients, by Grade and Criterion Measure ..... 48
Table 50 Math Composite Score Concurrent Validity Coefficients, by Grade and Criterion Measure ..... 49
Table 51 Reading Composite Score Predictive Validity Coefficients, by Grade and Criterion Measure ..... 50
Table 52 Reading Composite Score Concurrent Validity Coefficients, by Grade and Criterion Measure ..... 50
Table 53 Classification Accuracy, Two-by-Two Model ..... 51

# Table of Contents *cont.*

Tables *cont.*


Table 54 Classification Accuracy Statistics ..... 52
Table 55 Classification Accuracy of Early Numeracy Composite Scores and TCAP ..... 53
Table 56 Classification Accuracy of Fall Letter Naming Fluency with Winter Early Literacy Composite Scores and Word Reading Fluency (Kindergarten) ..... 54
Table 57 Classification Accuracy of Fall Letter Naming Fluency with Winter Early Literacy Composite Scores and R–CBM (Kindergarten) ..... 54
Table 58 Classification Accuracy of Fall Oral Reading Fluency Scores and Spring ITBS (Grade 1) ..... 55
Table 59 Classification Accuracy of Math Composite Scores and ITBS (Grade 2) and ISAT (Grades 3–8) ..... 56
Table 60 Classification Accuracy of Math Composite Scores and NWEA–MAP (Grades 2–6) ..... 56
Table 61 Classification Accuracy of Math Composite Scores and STAAR (Grades 3–8) ..... 57
Table 62 Classification Accuracy of Math Composite Scores and NMSBA (Grades 6–8) ..... 57
Table 63 Classification Accuracy of Reading Composites Scores and ISAT (Grades 3–8) ..... 58
Table 64 Classification Accuracy of Reading Composites Scores and NWEA–MAP (Grades 2–6) ..... 59
Table 65 Classification Accuracy of Reading Composites Scores and MAP–GLA (Grades 3–8) ..... 59
Table 66 Classification Accuracy of Reading Composites Scores and STAAR (Grades 3–5) ..... 60
Table A1 Early Numeracy Measure Descriptions ..... 63
Table A2 Early Literacy Measure Descriptions ..... 65
Table A3 Math Measure Descriptions ..... 67
Table A4 Reading Measure Descriptions ..... 69

# Introduction

aimswebPlus® is an assessment, data management, and reporting system that combines standards-aligned assessments of math and reading achievement with brief curriculum-based measurement (CBM) of critical math and reading basic skills for Kindergarten through Grade 8 students. This system provides reliable, valid, and nationally normed scores for Fall, Winter, and Spring benchmark assessments, and provides all of the features and content for multi-tiered systems of support (MTSS). This manual discusses the technical characteristics of the aimswebPlus measures, including demographic characteristics, descriptive statistics, reliability and validity data, and classification accuracy information. In addition, an overview of the aimswebPlus measures available at each grade level is provided in the appendix.

Note that this manual is dynamic document—additional information will be added as new research is conducted and new evidence is collected. Also, this manual is designed to be utilized in conjunction with the *aimswebPlus Development Manual,* which provides detailed information regarding the rationale for the aimswebPlus measures, descriptions of the developmental stages, and the supporting scientific research.

# Standardization Sample

Over 31,000 students participated in the aimswebPlus standardization study, with data collected during the 2013–2014 school year. Most participating students completed testing in each of the Fall, Winter, and Spring test sessions. Table 1 provides a summary of the demographic characteristics of the standardization sample for the math and reading measures at each grade level. Characteristics are reported for sex, race/ethnicity, and English language learner status.

Sampling was conducted at the school level, by grade. Schools indicated the grade(s) that would participate in testing and were then assigned to reading, math, or both content areas. Participating schools were required to assess to *all* students in the selected grades except those with moderate to severe intellectual disabilities or moderate to severe motor impairment and those who are blind, deaf, or had an English Language Proficiency score of less than 3.

The standardization sample at each grade level reflects adequate representation across each demographic category, enabling the selection of normative samples that are representative of the U.S. population.

**Table 1** Demographic Characteristics of the Standardization Sample

| | | Sex | | | | Race/Ethnicity | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Female | | Male | | Black | | Hispanic | | Other | | White | | ELL |
| Subject | Grade | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % | % |
| Math | K | 701 | 49 | 741 | 51 | 229 | 16 | 234 | 16 | 46 | 3 | 933 | 65 | 10 |
| Math | 1 | 854 | 50 | 840 | 50 | 305 | 18 | 237 | 14 | 90 | 5 | 1062 | 63 | 10 |
| Math | 2 | 970 | 49 | 991 | 51 | 287 | 15 | 386 | 20 | 103 | 5 | 1185 | 60 | 10 |
| Math | 3 | 920 | 49 | 950 | 51 | 345 | 18 | 315 | 17 | 121 | 6 | 1089 | 58 | 10 |
| Math | 4 | 907 | 47 | 1018 | 53 | 313 | 16 | 302 | 16 | 158 | 8 | 1152 | 60 | 10 |
| Math | 5 | 952 | 48 | 1024 | 52 | 315 | 16 | 321 | 16 | 125 | 6 | 1215 | 61 | 10 |
| Math | 6 | 814 | 47 | 922 | 53 | 133 | 8 | 477 | 27 | 143 | 8 | 983 | 57 | 10 |
| Math | 7 | 678 | 48 | 731 | 52 | 146 | 10 | 445 | 32 | 59 | 4 | 759 | 54 | 10 |
| Math | 8 | 601 | 46 | 698 | 54 | 129 | 10 | 389 | 30 | 90 | 7 | 691 | 53 | 10 |
| Reading | K | 731 | 49 | 752 | 51 | 293 | 20 | 241 | 16 | 44 | 3 | 905 | 61 | 10 |
| Reading | 1 | 825 | 50 | 822 | 50 | 335 | 20 | 242 | 15 | 86 | 5 | 984 | 60 | 10 |
| Reading | 2 | 1107 | 50 | 1102 | 50 | 379 | 17 | 425 | 19 | 110 | 5 | 1295 | 59 | 10 |
| Reading | 3 | 1103 | 50 | 1120 | 50 | 446 | 20 | 337 | 15 | 138 | 6 | 1302 | 59 | 10 |
| Reading | 4 | 1036 | 47 | 1162 | 53 | 409 | 19 | 342 | 16 | 167 | 8 | 1281 | 58 | 10 |
| Reading | 5 | 1118 | 48 | 1192 | 52 | 432 | 19 | 356 | 15 | 149 | 6 | 1373 | 59 | 10 |
| Reading | 6 | 781 | 48 | 850 | 52 | 225 | 14 | 377 | 23 | 115 | 7 | 914 | 56 | 10 |
| Reading | 7 | 625 | 50 | 637 | 50 | 205 | 16 | 342 | 27 | 25 | 2 | 690 | 55 | 10 |
| Reading | 8 | 565 | 48 | 621 | 52 | 133 | 11 | 328 | 28 | 57 | 5 | 668 | 56 | 10 |

# Norm Sample

Tables 2 through 4 present the demographic characteristics of the normative samples for the math and reading measures at each grade level. To be included in the norm sample, students had to complete the set of measures assigned to them (reading, math, or both). The percentage of students completing all assigned measures in all three seasons generally exceeded 90% in Math (Grades 2–8) and Early Literacy (Kindergarten and Grade 1). Approximately 85% of students completed all Early Numeracy measures (Kindergarten and Grade 1) and all Reading measures (Grades 2–8) in all three seasons. The dropout pattern was unrelated to demographic characteristics and was generally consistent across participating schools, with two exceptions. First, one school dropped out after the Winter testing session in the Early Numeracy study. Second, Oral Reading Fluency was administered on two separate platforms during Fall testing, which then had to be combined by matching various student characteristics, including student name. About 15% of the cases could not be matched and were excluded from the remaining data analyses.

Although the standardization samples were reasonably representative of the U.S. student population across demographic categories (even after listwise deletion of students not completing all assigned measures), a resampling method was used to generate the final norm samples. By using this resampling method, a perfect match to the U.S. Census Bureau American Community Survey (U.S. Census Bureau, 2013) by sex, race/ethnicity, and ELL status was obtained. This matching was done to improve precision and reduce bias in the norms. The resampling algorithm used a target total sample size by grade and subject, resulting in target counts for each demographic based on U.S. census percentages. The total target sample size for each grade was identified such that it did not exceed more than twice the original sample and no student would be resampled more than eight times. This resampling technique is analogous to weighting each student.

Table 2 summarizes the characteristics of the Early Literacy and Early Numeracy norm samples (Kindergarten and Grade 1). Note that SES percentages are based on free and reduced lunch data for Tables 2 through 4.

**Table 2** Demographic Characteristics of the Norm Sample, Kindergarten and Grade 1

| | | Sex | | | | Race/Ethnicity | | | | | | | | | SES | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Female | | Male | | Black | | Hispanic | | Other | | White | | ELL | Low | Mod | High |
| Subject | Grade | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % | % | % | % | % |
| Early Numeracy | K | 1000 | 50 | 1000 | 50 | 279 | 14 | 504 | 25 | 204 | 10 | 1013 | 51 | 10 | 32 | 32 | 36 |
| Early Numeracy | 1 | 1000 | 50 | 1000 | 50 | 265 | 13 | 506 | 25 | 201 | 10 | 1028 | 51 | 10 | 32 | 32 | 36 |
| Early Literacy | K | 1000 | 50 | 1000 | 50 | 279 | 14 | 504 | 25 | 204 | 10 | 1013 | 51 | 10 | 32 | 32 | 36 |
| Early Literacy | 1 | 1000 | 50 | 1000 | 50 | 265 | 13 | 506 | 25 | 201 | 10 | 1028 | 51 | 10 | 32 | 32 | 36 |

Tables 3 and 4 summarize the characteristics of the Math (Grades 2–8) and Reading (Grades 2–8) norm samples, respectively. For Reading, note that Oral Reading Fluency sample selections based on sex are generally close to an even split between males and females, and characteristics based on race/ethnicity are closely matched to U.S. population estimates. The race/ethnicity sample characteristics for Reading Comprehension, Vocabulary, and Silent Reading Fluency exactly match those of the U.S. population estimates, per the previous weighting discussion.

**Table 3** Demographic Characteristics of the Norm Sample for Math, Grades 2 Through 8

| | | | Sex | | | | Race/Ethnicity | | | | | | | | SES | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Female | | Male | | Black | | Hispanic | | Other | | White | | ELL | Low | Mod | High |
| Subject | Grade | Measure | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % | % | % | % | % |
| Math | 2 | NCF–T, MCF, CA | 1500 | 50 | 1500 | 50 | 420 | 14 | 700 | 23 | 300 | 10 | 1580 | 53 | 10 | 30 | 40 | 30 |
| Math | 3 | NCF–T, MCF, CA | 1500 | 50 | 1500 | 50 | 420 | 14 | 700 | 23 | 292 | 10 | 1588 | 53 | 10 | 30 | 40 | 30 |
| Math | 4 | NCF–T, MCF, CA | 1500 | 50 | 1500 | 50 | 430 | 14 | 650 | 22 | 300 | 10 | 1620 | 54 | 10 | 30 | 40 | 30 |
| Math | 5 | NCF–T, MCF, CA | 1500 | 50 | 1500 | 50 | 414 | 14 | 693 | 23 | 293 | 10 | 1600 | 53 | 10 | 30 | 40 | 30 |
| Math | 6 | NCF–T, MCF, CA | 1000 | 50 | 1000 | 50 | 260 | 13 | 487 | 24 | 187 | 9 | 1066 | 53 | 10 | 30 | 40 | 30 |
| Math | 7 | NCF–T, MCF, CA | 1000 | 50 | 1000 | 50 | 275 | 14 | 456 | 23 | 100 | 5 | 1169 | 58 | 10 | 30 | 40 | 30 |
| Math | 8 | NCF–T, MCF, CA | 1000 | 50 | 1000 | 50 | 234 | 12 | 446 | 22 | 150 | 8 | 1170 | 58 | 10 | 30 | 40 | 30 |

**Table 4** Demographic Characteristics of the Norm Sample for Reading, Grades 2 Through 8

| | | | Sex | | | | Race/Ethnicity | | | | | | | | SES | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Female | | Male | | Black | | Hispanic | | Other | | White | | ELL | Low | Mod | High |
| Subject | Grade | Measure | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % | % | % | % | % |
| Reading | 2 | ORF | 1158 | 0.49 | 1187 | 0.51 | 355 | 0.15 | 614 | 0.26 | 246 | 0.10 | 1130 | 0.48 | 10 | 32 | 32 | 36 |
| Reading | 3 | ORF | 1180 | 0.51 | 1125 | 0.49 | 266 | 0.12 | 583 | 0.25 | 225 | 0.10 | 1231 | 0.53 | 10 | 32 | 32 | 36 |
| Reading | 4 | ORF | 1251 | 0.50 | 1233 | 0.50 | 307 | 0.12 | 591 | 0.24 | 250 | 0.10 | 1336 | 0.54 | 10 | 32 | 32 | 36 |
| Reading | 5 | ORF | 1138 | 0.52 | 1066 | 0.48 | 337 | 0.15 | 518 | 0.24 | 251 | 0.11 | 1098 | 0.50 | 10 | 32 | 32 | 36 |
| Reading | 6 | ORF | 842 | 0.50 | 826 | 0.50 | 260 | 0.16 | 396 | 0.24 | 153 | 0.09 | 859 | 0.51 | 10 | 32 | 32 | 36 |
| Reading | 7 | ORF | 814 | 0.51 | 790 | 0.49 | 255 | 0.16 | 413 | 0.26 | 137 | 0.09 | 799 | 0.50 | 10 | 32 | 32 | 36 |
| Reading | 8 | ORF | 790 | 0.53 | 688 | 0.47 | 194 | 0.13 | 392 | 0.27 | 142 | 0.10 | 750 | 0.51 | 10 | 32 | 32 | 36 |
| Reading | 2 | RC, VO | 1500 | 0.50 | 1500 | 0.50 | 413 | 0.14 | 740 | 0.25 | 300 | 0.10 | 1547 | 0.52 | 10 | 32 | 32 | 36 |
| Reading | 3 | RC, VO | 1500 | 0.50 | 1500 | 0.50 | 414 | 0.14 | 732 | 0.24 | 292 | 0.10 | 1562 | 0.52 | 10 | 32 | 32 | 36 |
| Reading | 4 | RC, VO, SRF | 1500 | 0.50 | 1500 | 0.50 | 407 | 0.14 | 717 | 0.24 | 289 | 0.10 | 1587 | 0.53 | 10 | 32 | 32 | 36 |
| Reading | 5 | RC, VO, SRF | 1500 | 0.50 | 1500 | 0.50 | 415 | 0.14 | 693 | 0.23 | 293 | 0.10 | 1599 | 0.53 | 10 | 32 | 32 | 36 |
| Reading | 6 | RC, VO, SRF | 1000 | 0.50 | 1000 | 0.50 | 285 | 0.14 | 462 | 0.23 | 187 | 0.09 | 1066 | 0.53 | 10 | 32 | 32 | 36 |
| Reading | 7 | RC, VO, SRF | 1000 | 0.50 | 1000 | 0.50 | 275 | 0.14 | 456 | 0.23 | 182 | 0.09 | 1087 | 0.54 | 10 | 32 | 32 | 36 |
| Reading | 8 | RC, VO, SRF | 1000 | 0.50 | 1000 | 0.50 | 202 | 0.10 | 446 | 0.22 | 184 | 0.09 | 1168 | 0.58 | 10 | 32 | 32 | 36 |

# Standardization Sample Descriptives

For most Early Numeracy and Early Literacy measures (Kindergarten and Grade 1), descriptive statistics are based on the total number correct score (e.g., number of correctly answered items). The two exceptions are Phoneme Segmentation (total number of correct phonemes) and Oral Reading Fluency (mean number of words read correctly in two stories, each read for 1 minute). The descriptive statistics for Math and Reading are based on the following:

- Oral Reading Fluency: Mean number of words read correctly in two stories, each read for 1 minute.
- Silent Reading Fluency: Words read silently per minute.
- Number Comparison Fluency–Triads and Mental Computation Fluency: Adjusted total scores, in which 0.5 points are subtracted for every item answered incorrectly and the result rounded to the nearest whole number.
- Concepts & Applications, Vocabulary, and Reading Comprehension: Scores reported on a vertical standardized scale that spans Grades 2 through 8, is centered on Spring of Grade 5, and has a mean of 200.

Each benchmark form within a given grade was developed from a common blueprint, with the resulting forms nearly equivalent in difficulty. Thus, score gains can be interpreted as actual achievement growth. One way to interpret the magnitude of the gain is to express it in Fall (or Winter) standard deviation (*SD*) units. Doing so enables direct comparison of gains across measures and grades.

As expected, scores for each grade tend to increase across seasons. Large annual gains (>0.7 *SD* units) are common in Kindergarten through Grade 3. In Grades 4 through 8, gains are more modest, generally ranging from about 0.3 to 0.5 *SD* units. Two measures—Initial Sounds and Auditory Vocabulary—show very small gains. These two Early Literacy measures were designed to support diagnostic interpretation of results for the lowest performing students; as such, they are relatively easy and are not expected to be sensitive to growth for the average performing student.

Tables 5 through 8 provide standardization sample sizes, means, and standard deviations by season and grade for all of the aimswebPlus reading and math measures. Note that the results shown are based on students with a valid score on each measure in a given season, while dashed lines indicate that a given measure is not administered in a particular grade or season.

**Table 5** Descriptive Statistics for Early Numeracy Measures in the Standardization Sample, Kindergarten and Grade 1

| | | | Fall | | | Winter | | | Spring | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | Grade | Measure | *n* | Mean | *SD* | *n* | Mean | *SD* | *n* | Mean | *SD* |
| Math | K | NNF | 1304 | 32.22 | 13.84 | 1195 | 40.17 | 14.58 | 1178 | 48.52 | 14.46 |
| Math | K | QTF | 1324 | 12.54 | 4.71 | 1197 | 15.33 | 4.66 | 1178 | 18.21 | 4.40 |
| Math | K | QDF | --- | --- | --- | 1149 | 7.27 | 4.55 | 1156 | 11.60 | 5.27 |
| Math | K | CA | 1376 | 12.33 | 4.74 | 1216 | 14.87 | 4.81 | 1192 | 18.35 | 4.41 |
| Math | 1 | NCF–P | 1395 | 23.54 | 7.70 | 1226 | 27.79 | 7.00 | 1321 | 28.93 | 6.91 |
| Math | 1 | MFF–1D | 1370 | 12.70 | 5.92 | 1222 | 15.65 | 5.94 | 1322 | 16.47 | 6.68 |
| Math | 1 | MFF–T | --- | --- | --- | 1118 | 4.40 | 4.23 | 1227 | 6.46 | 4.88 |
| Math | 1 | CA | 1453 | 13.06 | 5.29 | 1255 | 15.83 | 5.35 | 1360 | 18.00 | 5.35 |

**Table 6** Descriptive Statistics for Early Literacy Measures in the Standardization Sample, Kindergarten and Grade 1

| | | | Fall | | | Winter | | | Spring | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | Grade | Measure | *n* | Mean | *SD* | *n* | Mean | *SD* | *n* | Mean | *SD* |
| Reading | K | PC | 1240 | 7.12 | 1.71 | NA | --- | --- | --- | --- | --- |
| Reading | K | IS | 1240 | 9.35 | 3.28 | 1205 | 10.67 | 2.42 | --- | --- | --- |
| Reading | K | AV | 1240 | 21.10 | 3.96 | 1205 | 21.05 | 3.44 | 1222 | 22.88 | 2.85 |
| Reading | K | LNF | 1240 | 33.79 | 17.24 | 1205 | 45.34 | 19.79 | 1222 | 54.16 | 19.52 |
| Reading | K | LWSF | --- | --- | --- | 1205 | 32.52 | 13.73 | 1222 | 41.76 | 11.95 |
| Reading | K | PS | --- | --- | --- | 1205 | 33.25 | 13.68 | 1222 | 39.28 | 10.15 |
| Reading | K | WRF | --- | --- | --- | --- | --- | --- | 1222 | 18.95 | 15.75 |
| Reading | 1 | LWSF | 1342 | 47.07 | 11.34 | --- | --- | --- | --- | --- | --- |
| Reading | 1 | PS | 1342 | 39.01 | 10.04 | --- | --- | --- | --- | --- | --- |
| Reading | 1 | AV | 1342 | 22.80 | 3.55 | 1387 | 21.51 | 5.10 | 1498 | 23.51 | 3.35 |
| Reading | 1 | WRF | 1342 | 30.57 | 21.21 | 1387 | 42.70 | 22.69 | 1498 | 54.84 | 22.71 |
| Reading | 1 | ORF | 1342 | 41.18 | 29.18 | 1387 | 59.38 | 32.21 | 1498 | 74.92 | 36.73 |

**Table 7** Descriptive Statistics for Math Measures in the Standardization Sample, Grades 2 Through 8

| | | | Fall | | | Winter | | | Spring | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | Grade | Measure | *n* | Mean | *SD* | *n* | Mean | *SD* | *n* | Mean | *SD* |
| Math | 2 | NCF–T | 1961 | 6.73 | 7.08 | 1842 | 7.83 | 7.28 | 1763 | 10.86 | 8.89 |
| Math | 3 | NCF–T | 1870 | 14.00 | 9.73 | 1860 | 14.04 | 10.19 | 1803 | 17.12 | 10.73 |
| Math | 4 | NCF–T | 1925 | 13.54 | 10.47 | 1891 | 12.50 | 10.16 | 1726 | 17.11 | 11.08 |
| Math | 5 | NCF–T | 1976 | 15.05 | 10.54 | 1888 | 15.34 | 10.73 | 1794 | 17.49 | 11.16 |
| Math | 6 | NCF–T | 1736 | 9.96 | 8.09 | 1497 | 9.96 | 8.86 | 1514 | 13.12 | 9.99 |
| Math | 7 | NCF–T | 1409 | 8.60 | 7.55 | 1241 | 8.72 | 7.97 | 1144 | 11.30 | 9.65 |
| Math | 8 | NCF–T | 1299 | 9.70 | 8.62 | 1176 | 9.38 | 9.03 | 1078 | 11.04 | 10.11 |
| Math | 2 | MCF | 1961 | 8.38 | 7.85 | 1842 | 11.11 | 8.94 | 1763 | 14.48 | 9.42 |
| Math | 3 | MCF | 1870 | 8.87 | 7.57 | 1860 | 8.87 | 7.03 | 1803 | 13.87 | 9.33 |
| Math | 4 | MCF | 1925 | 10.96 | 6.19 | 1891 | 10.38 | 7.08 | 1726 | 16.07 | 8.62 |
| Math | 5 | MCF | 1976 | 9.37 | 6.51 | 1888 | 9.15 | 6.75 | 1794 | 12.06 | 8.42 |
| Math | 6 | MCF | 1736 | 10.94 | 8.29 | 1497 | 13.57 | 10.47 | 1514 | 15.49 | 11.67 |
| Math | 7 | MCF | 1409 | 9.37 | 8.82 | 1241 | 9.83 | 9.53 | 1144 | 12.29 | 10.40 |
| Math | 8 | MCF | 1299 | 10.08 | 10.46 | 1176 | 9.74 | 9.94 | 1078 | 12.04 | 11.42 |
| Math | 2 | CA | 1961 | 149.02 | 21.36 | 1842 | 158.60 | 23.59 | 1763 | 168.22 | 23.71 |
| Math | 3 | CA | 1870 | 166.96 | 21.89 | 1860 | 172.88 | 20.71 | 1803 | 182.75 | 22.34 |
| Math | 4 | CA | 1925 | 177.04 | 17.21 | 1891 | 183.34 | 16.61 | 1726 | 191.75 | 20.71 |
| Math | 5 | CA | 1976 | 192.02 | 18.36 | 1888 | 196.19 | 19.82 | 1794 | 202.56 | 19.48 |
| Math | 6 | CA | 1736 | 199.41 | 17.28 | 1497 | 203.55 | 20.50 | 1514 | 207.99 | 22.09 |
| Math | 7 | CA | 1409 | 202.95 | 18.57 | 1241 | 203.89 | 19.85 | 1144 | 206.30 | 20.14 |
| Math | 8 | CA | 1299 | 204.99 | 20.41 | 1176 | 209.78 | 18.08 | 1078 | 211.89 | 17.85 |

**Table 8** Descriptive Statistics for Reading Measures in the Standardization Sample, Grades 2 Through 8

| | | | Fall | | | Winter | | | Spring | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | Grade | Measure | *n* | Mean | *SD* | *n* | Mean | *SD* | *n* | Mean | *SD* |
| Reading | 2 | RC | 2055 | 155.80 | 27.90 | 1921 | 161.16 | 25.31 | 1870 | 173.61 | 25.63 |
| Reading | 3 | RC | 1927 | 175.35 | 25.75 | 1937 | 177.69 | 28.97 | 1869 | 185.08 | 26.82 |
| Reading | 4 | RC | 2007 | 189.22 | 22.18 | 1957 | 193.55 | 24.77 | 1855 | 194.29 | 25.29 |
| Reading | 5 | RC | 2122 | 201.06 | 23.55 | 1998 | 199.63 | 25.73 | 1905 | 205.73 | 25.01 |
| Reading | 6 | RC | 1538 | 213.25 | 24.01 | 1295 | 214.08 | 24.96 | 1359 | 213.31 | 24.84 |
| Reading | 7 | RC | 1191 | 220.86 | 25.28 | 1045 | 219.24 | 23.93 | 978 | 217.85 | 23.17 |
| Reading | 8 | RC | 1143 | 229.12 | 23.05 | 988 | 226.80 | 22.76 | 907 | 227.10 | 23.87 |
| Reading | 2 | VO | 2084 | 175.56 | 25.53 | 1991 | 186.75 | 25.88 | 1842 | 192.07 | 22.18 |
| Reading | 3 | VO | 1955 | 194.62 | 22.06 | 1952 | 197.58 | 22.21 | 1839 | 204.10 | 22.16 |
| Reading | 4 | VO | 2083 | 205.61 | 21.32 | 2024 | 204.89 | 21.63 | 1874 | 211.49 | 21.32 |
| Reading | 5 | VO | 2291 | 215.20 | 19.95 | 2042 | 215.41 | 21.02 | 1910 | 220.65 | 21.62 |
| Reading | 6 | VO | 1664 | 223.86 | 20.42 | 1358 | 227.18 | 20.49 | 1396 | 227.37 | 22.22 |
| Reading | 7 | VO | 1293 | 229.47 | 20.32 | 1094 | 231.59 | 20.58 | 1003 | 232.80 | 22.47 |
| Reading | 8 | VO | 1241 | 234.61 | 23.51 | 1027 | 237.85 | 24.99 | 950 | 236.61 | 24.24 |
| Reading | 4 | SRF | 1786 | 128.28 | 44.98 | 1783 | 137.68 | 48.99 | 1722 | 146.49 | 49.03 |
| Reading | 5 | SRF | 2062 | 130.31 | 40.04 | 1791 | 144.80 | 43.46 | 1796 | 145.10 | 39.24 |
| Reading | 6 | SRF | 1529 | 149.55 | 52.08 | 1257 | 151.71 | 48.68 | 1239 | 154.46 | 46.43 |
| Reading | 7 | SRF | 1132 | 147.83 | 48.13 | 962 | 158.40 | 48.76 | 871 | 166.53 | 46.38 |
| Reading | 8 | SRF | 1116 | 154.64 | 47.42 | 861 | 158.58 | 44.87 | 751 | 170.16 | 49.12 |
| Reading | 2 | ORF | 1787 | 68.79 | 35.69 | 1826 | 84.07 | 38.17 | 1981 | 100.27 | 41.84 |
| Reading | 3 | ORF | 1637 | 91.31 | 38.88 | 1885 | 107.10 | 36.73 | 1930 | 119.81 | 38.69 |
| Reading | 4 | ORF | 1657 | 110.18 | 37.01 | 1954 | 125.98 | 37.50 | 2021 | 134.22 | 41.59 |
| Reading | 5 | ORF | 1691 | 123.92 | 40.58 | 1927 | 144.42 | 41.94 | 2076 | 146.36 | 44.41 |
| Reading | 6 | ORF | 1332 | 143.29 | 37.99 | 1518 | 148.45 | 41.44 | 1366 | 154.78 | 40.19 |
| Reading | 7 | ORF | 967 | 142.45 | 39.42 | 1149 | 150.68 | 39.04 | 1024 | 164.88 | 40.08 |
| Reading | 8 | ORF | 859 | 145.96 | 37.83 | 1093 | 148.76 | 37.19 | 923 | 152.28 | 37.12 |

# Norm Sample Descriptives

Tables 9 through 12 provide norm sample sizes, means, and standard deviations by season and grade for all of the aimswebPlus reading and math measures. Scores for each measure are reported on the same scale described in the Standardization Sample Descriptives section of this manual.

Table 13 presents the total score means and corresponding Quantile® scores for Concepts & Applications, by grade and season. Similarly, Table 14 presents the words read correctly means and corresponding Lexile® levels for Oral Reading Fluency, by grade and season. Quantile and Lexile scores represent MetaMetrics's proprietary developmental math and reading scales, respectively, that span Kindergarten through Grade 12. These scores were obtained as part of extensive linking studies conducted by MetaMetrics (see *Linking aimswebPlus Concepts & Applications (Grades 2–8) With the Quantile® Framework for Mathematics* and *Linking aimswebPlus Oral Reading Fluency (Grade 1) With the Lexile® Framework for Reading* for descriptions of each study).

As previously discussed, the norm samples are based on a resampling method used to improve representation of certain student and school demographics as compared to U.S. census data. Note that the results shown in the following tables are based on norm sample students with valid scores in all three testing seasons, while dashed lines indicate that a given measure is not administered in a particular grade or season.

**Table 9** Descriptive Statistics for Early Numeracy Measures and Composites in the Norm Sample, Kindergarten and Grade 1

| | | | Fall | | | Winter | | | Spring | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | Grade | Measure | *n* | Mean | *SD* | *n* | Mean | *SD* | *n* | Mean | *SD* |
| Math | K | NNF | 2000 | 28.77 | 14.89 | 2000 | 40.73 | 14.47 | 2000 | 48.70 | 14.53 |
| Math | K | QTF | 2000 | 11.25 | 5.21 | 2000 | 15.22 | 4.29 | 2000 | 18.21 | 4.00 |
| Math | K | QDF | --- | --- | --- | 2000 | 7.58 | 4.48 | 2000 | 11.85 | 5.23 |
| Math | K | CA | 2000 | 10.78 | 5.06 | 2000 | 14.68 | 4.67 | 2000 | 18.26 | 4.28 |
| Math | K | Composite | 2000 | 34.67 | 11.84 | 2000 | 51.06 | 14.71 | 2000 | 64.54 | 14.70 |
| Math | 1 | NCF–P | 2000 | 22.20 | 8.15 | 2000 | 27.91 | 7.09 | 2000 | 29.76 | 6.18 |
| Math | 1 | MFF–1D | 2000 | 11.14 | 6.20 | 2000 | 15.37 | 5.84 | 2000 | 16.21 | 6.64 |
| Math | 1 | MFF–T | --- | --- | --- | 2000 | 4.42 | 4.34 | 2000 | 6.54 | 5.02 |
| Math | 1 | CA | 2000 | 11.87 | 5.67 | 2000 | 16.00 | 5.22 | 2000 | 18.13 | 5.06 |
| Math | 1 | Composite | 2000 | 47.97 | 16.49 | 2000 | 63.69 | 19.02 | 2000 | 70.64 | 19.52 |

**Table 10** Descriptive Statistics for Early Literacy Measures and Composites in the Norm Sample, Kindergarten and Grade 1

| | | | Fall | | | Winter | | | Spring | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | Grade | Measure | *n* | Mean | *SD* | *n* | Mean | *SD* | *n* | Mean | *SD* |
| Reading | K | PC | 2000 | 7.05 | 1.75 | --- | --- | --- | --- | --- | --- |
| Reading | K | IS | 2000 | 8.71 | 3.89 | 2000 | 10.78 | 2.27 | --- | --- | --- |
| Reading | K | AV | 2000 | 21.12 | 3.88 | 2000 | 21.42 | 2.91 | 2000 | 22.98 | 2.74 |
| Reading | K | LNF | 2000 | 28.39 | 18.69 | 2000 | 46.90 | 18.14 | 2000 | 55.45 | 19.05 |
| Reading | K | LWSF | --- | --- | --- | 2000 | 32.08 | 13.46 | 2000 | 41.95 | 12.38 |
| Reading | K | PS | --- | --- | --- | 2000 | 33.39 | 13.09 | 2000 | 39.06 | 10.60 |
| Reading | K | WRF | --- | --- | --- | --- | --- | --- | 2000 | 19.03 | 15.54 |
| Reading | K | Composite | 2000 | 31.72 | 17.70 | 2000 | 112.36 | 38.12 | 2000 | 136.46 | 34.71 |
| Reading | 1 | LWSF | 2000 | 43.75 | 10.64 | --- | --- | --- | --- | --- | --- |
| Reading | 1 | PS | 2000 | 38.48 | 9.31 | --- | --- | --- | --- | --- | --- |
| Reading | 1 | AV | 2000 | 23.06 | 2.47 | 2000 | 22.62 | 2.45 | 2000 | 23.97 | 1.67 |
| Reading | 1 | WRF | 2000 | 26.64 | 22.56 | 2000 | 42.65 | 22.93 | 2000 | 56.09 | 21.95 |
| Reading | 1 | ORF | 2000 | 35.25 | 30.07 | 2000 | 59.07 | 32.64 | 2000 | 74.37 | 35.54 |
| Reading | 1 | Composite | 2000 | 85.55 | 36.34 | --- | --- | --- | --- | --- | --- |

**Table 11** Descriptive Statistics for Math Measures and Composites in the Norm Sample, Grades 2 Through 8

| | | | Fall | | | Winter | | | Spring | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | Grade | Measure | *n* | Mean | *SD* | *n* | Mean | *SD* | *n* | Mean | *SD* |
| Math | 2 | NCF–T | 3000 | 7.06 | 7.34 | 3000 | 8.06 | 7.29 | 3000 | 10.98 | 9.05 |
| Math | 3 | NCF–T | 3000 | 15.11 | 9.94 | 3000 | 15.45 | 10.29 | 3000 | 18.40 | 10.75 |
| Math | 4 | NCF–T | 3000 | 14.44 | 10.56 | 3000 | 13.33 | 10.26 | 3000 | 17.54 | 11.15 |
| Math | 5 | NCF–T | 3000 | 15.90 | 10.63 | 3000 | 16.87 | 10.67 | 3000 | 18.51 | 11.21 |
| Math | 6 | NCF–T | 2000 | 10.72 | 8.28 | 2000 | 10.36 | 8.73 | 2000 | 13.60 | 9.98 |
| Math | 7 | NCF–T | 2000 | 9.74 | 8.24 | 2000 | 9.58 | 8.12 | 2000 | 11.92 | 9.53 |
| Math | 8 | NCF–T | 2000 | 11.40 | 8.99 | 2000 | 11.01 | 9.26 | 2000 | 12.01 | 10.02 |
| Math | 2 | MCF | 3000 | 8.91 | 8.15 | 3000 | 11.67 | 9.15 | 3000 | 14.68 | 9.64 |
| Math | 3 | MCF | 3000 | 9.64 | 7.85 | 3000 | 9.79 | 7.17 | 3000 | 15.05 | 9.58 |
| Math | 4 | MCF | 3000 | 11.37 | 6.42 | 3000 | 10.83 | 7.21 | 3000 | 16.57 | 8.73 |
| Math | 5 | MCF | 3000 | 9.94 | 6.64 | 3000 | 9.75 | 6.75 | 3000 | 12.74 | 8.59 |
| Math | 6 | MCF | 2000 | 11.61 | 8.37 | 2000 | 14.08 | 10.20 | 2000 | 16.43 | 11.77 |
| Math | 7 | MCF | 2000 | 11.19 | 9.60 | 2000 | 10.90 | 9.84 | 2000 | 13.15 | 10.51 |
| Math | 8 | MCF | 2000 | 11.73 | 10.85 | 2000 | 11.32 | 10.15 | 2000 | 13.59 | 11.60 |
| Math | 2 | NSF | 3000 | 14.82 | 13.75 | 3000 | 19.73 | 14.94 | 3000 | 25.66 | 17.16 |
| Math | 3 | NSF | 3000 | 23.12 | 16.05 | 3000 | 25.24 | 15.90 | 3000 | 33.45 | 18.46 |
| Math | 4 | NSF | 3000 | 24.16 | 14.92 | 3000 | 24.16 | 15.90 | 3000 | 34.12 | 17.81 |
| Math | 5 | NSF | 3000 | 24.41 | 15.77 | 3000 | 26.62 | 15.95 | 3000 | 31.25 | 18.23 |
| Math | 6 | NSF | 2000 | 20.77 | 14.94 | 2000 | 24.44 | 17.61 | 2000 | 30.03 | 20.35 |
| Math | 7 | NSF | 2000 | 19.50 | 16.73 | 2000 | 20.48 | 16.71 | 2000 | 25.07 | 18.79 |
| Math | 8 | NSF | 2000 | 21.55 | 18.29 | 2000 | 22.33 | 18.35 | 2000 | 25.60 | 20.62 |
| Math | 2 | CA | 3000 | 150.70 | 21.34 | 3000 | 160.25 | 23.37 | 3000 | 168.09 | 23.98 |
| Math | 3 | CA | 3000 | 168.66 | 22.23 | 3000 | 175.55 | 20.57 | 3000 | 184.28 | 21.94 |
| Math | 4 | CA | 3000 | 177.89 | 17.66 | 3000 | 184.28 | 16.88 | 3000 | 191.55 | 21.09 |
| Math | 5 | CA | 3000 | 193.63 | 19.06 | 3000 | 198.62 | 19.02 | 3000 | 203.21 | 19.06 |
| Math | 6 | CA | 2000 | 200.60 | 16.55 | 2000 | 203.90 | 20.07 | 2000 | 208.13 | 22.16 |
| Math | 7 | CA | 2000 | 205.85 | 18.76 | 2000 | 206.20 | 19.66 | 2000 | 207.14 | 19.81 |
| Math | 8 | CA | 2000 | 207.44 | 20.74 | 2000 | 211.95 | 18.03 | 2000 | 212.78 | 17.22 |
| Math | 2 | Composite | 3000 | 166.67 | 32.62 | 3000 | 179.97 | 35.23 | 3000 | 193.75 | 37.43 |
| Math | 3 | Composite | 3000 | 193.40 | 34.94 | 3000 | 200.79 | 33.38 | 3000 | 217.73 | 37.00 |
| Math | 4 | Composite | 3000 | 203.70 | 29.64 | 3000 | 208.44 | 29.51 | 3000 | 225.66 | 35.46 |
| Math | 5 | Composite | 3000 | 219.46 | 32.23 | 3000 | 225.25 | 32.09 | 3000 | 234.46 | 34.01 |
| Math | 6 | Composite | 2000 | 222.93 | 28.85 | 2000 | 228.34 | 34.86 | 2000 | 238.16 | 39.70 |
| Math | 7 | Composite | 2000 | 226.77 | 33.26 | 2000 | 226.68 | 34.28 | 2000 | 232.21 | 36.49 |
| Math | 8 | Composite | 2000 | 230.57 | 36.42 | 2000 | 234.28 | 33.66 | 2000 | 238.38 | 34.96 |

**Table 12** Descriptive Statistics for Reading Measures and Composites in the Norm Sample, Grades 2 Through 8

| | | | Fall | | | Winter | | | Spring | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Subject | Grade | Measure | *n* | Mean | *SD* | *n* | Mean | *SD* | *n* | Mean | *SD* |
| Reading | 2 | ORF | 2580 | 70.50 | 36.10 | 2776 | 87.97 | 39.26 | 2968 | 104.71 | 41.30 |
| Reading | 3 | ORF | 2587 | 93.89 | 38.63 | 2752 | 109.30 | 36.35 | 2940 | 122.83 | 36.88 |
| Reading | 4 | ORF | 2634 | 112.63 | 36.46 | 2868 | 127.16 | 36.66 | 2960 | 136.06 | 39.55 |
| Reading | 5 | ORF | 2555 | 128.30 | 39.89 | 2673 | 148.45 | 42.37 | 2919 | 152.05 | 42.95 |
| Reading | 6 | ORF | 1775 | 148.06 | 37.14 | 1978 | 152.29 | 39.28 | 1894 | 158.04 | 38.13 |
| Reading | 7 | ORF | 1689 | 149.35 | 40.34 | 1931 | 156.94 | 37.25 | 1899 | 170.19 | 39.70 |
| Reading | 8 | ORF | 1603 | 153.52 | 36.37 | 1952 | 155.87 | 36.20 | 1672 | 158.35 | 36.47 |
| Reading | 4 | SRF | 2639 | 126.34 | 46.17 | 2729 | 136.54 | 47.39 | 2778 | 146.79 | 48.63 |
| Reading | 5 | SRF | 2820 | 131.01 | 42.21 | 2792 | 148.78 | 42.74 | 2815 | 149.41 | 39.03 |
| Reading | 6 | SRF | 1919 | 147.33 | 52.89 | 1907 | 153.62 | 47.91 | 1885 | 154.94 | 46.04 |
| Reading | 7 | SRF | 1889 | 146.78 | 52.51 | 1870 | 166.11 | 50.63 | 1799 | 170.96 | 44.71 |
| Reading | 8 | SRF | 1889 | 153.50 | 46.40 | 1800 | 161.18 | 44.11 | 1781 | 170.24 | 50.53 |
| Reading | 2 | VO | 3000 | 155.95 | 26.76 | 3000 | 170.59 | 23.96 | 3000 | 174.15 | 21.74 |
| Reading | 3 | VO | 3000 | 176.12 | 23.57 | 3000 | 181.43 | 20.33 | 3000 | 185.90 | 22.07 |
| Reading | 4 | VO | 3000 | 186.85 | 21.55 | 3000 | 187.24 | 21.23 | 3000 | 192.16 | 21.43 |
| Reading | 5 | VO | 3000 | 196.61 | 21.61 | 3000 | 198.61 | 20.41 | 3000 | 202.84 | 21.64 |
| Reading | 6 | VO | 2000 | 204.83 | 20.47 | 2000 | 208.68 | 20.71 | 2000 | 209.68 | 21.93 |
| Reading | 7 | VO | 2000 | 211.18 | 21.86 | 2000 | 214.97 | 20.23 | 2000 | 215.53 | 22.00 |
| Reading | 8 | VO | 2000 | 218.94 | 25.05 | 2000 | 222.70 | 24.12 | 2000 | 219.89 | 24.39 |
| Reading | 2 | RC | 3000 | 146.99 | 30.51 | 3000 | 154.56 | 24.79 | 3000 | 165.72 | 25.66 |
| Reading | 3 | RC | 3000 | 167.66 | 26.81 | 3000 | 173.98 | 28.37 | 3000 | 178.74 | 26.51 |
| Reading | 4 | RC | 3000 | 181.08 | 22.54 | 3000 | 185.76 | 24.20 | 3000 | 185.74 | 24.64 |
| Reading | 5 | RC | 3000 | 194.59 | 24.02 | 3000 | 194.28 | 24.87 | 3000 | 198.38 | 25.10 |
| Reading | 6 | RC | 2000 | 206.71 | 24.82 | 2000 | 205.25 | 25.68 | 2000 | 205.59 | 25.00 |
| Reading | 7 | RC | 2000 | 214.44 | 25.61 | 2000 | 213.35 | 23.72 | 2000 | 210.79 | 23.17 |
| Reading | 8 | RC | 2000 | 224.93 | 24.67 | 2000 | 222.10 | 22.73 | 2000 | 222.15 | 24.68 |
| Reading | 2 | Composite | 3000 | 339.11 | 64.20 | 3000 | 368.48 | 58.11 | 3000 | 392.47 | 57.50 |
| Reading | 3 | Composite | 3000 | 391.91 | 59.75 | 3000 | 411.04 | 57.35 | 3000 | 426.10 | 56.87 |
| Reading | 4 | Composite | 3000 | 432.21 | 53.11 | 3000 | 439.61 | 54.42 | 3000 | 449.55 | 56.86 |
| Reading | 5 | Composite | 3000 | 458.71 | 52.47 | 3000 | 466.39 | 54.82 | 3000 | 474.92 | 55.14 |
| Reading | 6 | Composite | 2000 | 485.56 | 55.44 | 2000 | 489.91 | 57.38 | 2000 | 492.41 | 56.31 |
| Reading | 7 | Composite | 2000 | 500.54 | 59.41 | 2000 | 510.15 | 55.69 | 2000 | 510.93 | 54.16 |
| Reading | 8 | Composite | 2000 | 520.91 | 60.24 | 2000 | 524.06 | 57.32 | 2000 | 526.40 | 60.82 |

**Table 13** Concepts & Applications Total Score Means and Quantiles, by Grade and Season

| | Fall | | Winter | | Spring | |
|---|---|---|---|---|---|---|
| Grade | Mean | Quantile | Mean | Quantile | Mean | Quantile |
| 2 | 150.70 | 268 | 160.25 | 360 | 168.09 | 442 |
| 3 | 168.66 | 452 | 175.55 | 524 | 184.28 | 605 |
| 4 | 177.89 | 544 | 184.28 | 605 | 191.55 | 687 |
| 5 | 193.63 | 708 | 198.62 | 759 | 203.21 | 800 |
| 6 | 200.60 | 779 | 203.90 | 810 | 208.13 | 851 |
| 7 | 205.85 | 831 | 206.20 | 831 | 207.14 | 841 |
| 8 | 207.44 | 841 | 211.95 | 892 | 212.78 | 902 |

**Table 14** Oral Reading Fluency Words Read Correctly (WRC) Means and Lexiles, by Grade and Season

| | Fall | | Winter | | Spring | |
|---|---|---|---|---|---|---|
| Grade | WRC | Lexile | WRC | Lexile | WRC | Lexile |
| 1 | 35.25 | BR | 59.07 | 70L | 74.37 | 145L |
| 2 | 70.50 | 250L | 87.97 | 350L | 104.71 | 450L |
| 3 | 93.89 | 455L | 109.30 | 535L | 122.83 | 600L |
| 4 | 112.63 | 560L | 127.16 | 635L | 136.06 | 680L |
| 5 | 128.30 | 695L | 148.45 | 795L | 152.05 | 815L |
| 6 | 148.06 | 920L | 152.29 | 940L | 158.04 | 970L |
| 7 | 149.35 | 990L | 156.94 | 1020L | 170.19 | 1100L |
| 8 | 153.52 | 1105L | 155.87 | 1120L | 158.35 | 1140L |

# Reliability

Tables 16 through 44 present test reliability results for each aimswebPlus measure. Two types of reliability are reported: alternate form and internal consistency. Alternate form reliability is reported for all *timed* measures, while internal consistency reliability is reported for *untimed* measures.

Reliability is an estimate of the consistency or stability of test scores. Consistency is affected by random error (which can be caused by many factors including variations in student motivation and attentiveness), imperfect and incomplete specification of the achievement domain, and guessing. The choice of reliability method depends on how the test is administered and scored, as well as how the results will be used. For untimed tests that assess student achievement at a single point in time, internal consistency reliability is most appropriate. Among the various internal consistency methods, Cronbach's alpha is the most commonly utilized and it is the one reported for all aimswebPlus untimed measures.

Note that for untimed measures, items that were skipped/unanswered were scored as zero. To be included in the analysis, a minimum of five valid item scores were required for any given measure. This number of items was chosen because the administration guidelines for standardization testing indicating that testing should be discontinued if the student failed each of the first five items of a given measure. This occurred, on average, during about 1% of test administrations.

Cronbach's alpha is not appropriate for aimswebPlus timed measures because this type of reliability requires a score on all items in a given measure. The time limits used for aimswebPlus fluency measures are designed to provide strong reliability and growth sensitivity; however, these time limits also have the effect of ensuring that most students will *not* complete all of the items in a given measure. As such, alternate form reliability is most appropriate for aimswebPlus timed measures.

Another important reason for using alternate form reliability for these measures is how scores from the timed measures are used. aimswebPlus timed measures are used for benchmark screening and for frequent (e.g., weekly) monitoring of student progress. The timed measures have either 12 or 23 alternate forms for each grade, depending on benchmark seasons administered. Two (fall/winter or winter/spring) or three (fall/winter/spring) of the forms are used for universal screening, with the remaining 10 or 20 used for progress monitoring. All alternate forms for each measure were constructed from a common test blueprint and are nearly equivalent in difficulty.

Progress monitoring scores are used to estimate rate of growth and to determine whether that rate is sufficient to meet the performance goal set for a student. Therefore, it is important to know how variations in test content and occasion affect score consistency. Alternate form reliability is designed for that purpose.

aimswebPlus uses composite scores (sums of scores from two or more measures) to determine risk classification. A measure's influence on the composite depends on the magnitude of its variance relative to the variances of the other tests in the composite. The greater the variance, the greater its influence on the composite. In order to equalize the contribution of each test to the overall composite, a weighting method was used. However, because this process can be perceived as complicating the interpretation of scores, weighting was applied only when a measure's variance was greater than twice the variance of the other measures in the composite.

To equalize the contribution of each test in a composite, the total score for a measure was either multiplied by 1.0 or by a fractional weight (see Table 15). An exception was made for the Grade 1 Early Literacy Fall composite because ORF is such a strong predictor of end-of-year reading performance.

**Table 15** Benchmark Composite Scoring Rules, by Subject, Grade, and Season

| Subject | Grade | Season | Composite |
|---|---|---|---|
| Early Literacy | K | W, S | LNF + LWSF + PS |
| Early Literacy | 1 | F | LWSF + ORF |
| Reading | 2–3 | F, W, S | (1/2*ORF) + VO + RC |
| Reading | 4–8 | F, W, S | (1/2*SRF) + VO + RC |
| Early Numeracy | K | F | (1/3*NNF) + QTF + CA |
| Early Numeracy | K | W, S | (1/3*NNF) + QTF + CA + QDF |
| Early Numeracy | 1 | F | NCF–P + MFF–1D + CA |
| Early Numeracy | 1 | W, S | NCF–P + MFF–1D + CA + MFF–T |
| Math | 2–8 | F, W, S | (NCF–T + MCF) + CA |

Composite reliabilities are based on Feldt & Brennan's (1989) stratified alpha method. Stratified alpha uses observed reliabilities and variances for each measure contributing to the composite to estimate the error variance of the composite. Using this method, reliability is computed as:

$$Stratified\ \alpha = 1 - \frac{\sum \sigma_i^2 (1-\alpha_i)}{\sigma_x^2}$$

Where *i* is a component (i.e., measure) in the composite, α is the reliability of each component, and the denominator is the total composite variance. Note that sample sizes are not shown in the tables reporting stratified alpha values because the individual measure reliabilities come from different studies with varying sample sizes. As such, no single sample size is appropriate.

Reliability results are presented in table organized by domain: Early Literacy, Early Numeracy, Reading, and Math. Reliability coefficients are provided for each measure, season, and grade within these domains. Cronbach's alpha coefficients were derived using all standardization cases with valid test scores for a given season. Alternate form reliability coefficients were derived from data collected in separate equivalency studies. These alternate form equivalency studies are briefly described below, followed by the tables reporting reliability for each measure, grade, and season.

## Early Numeracy Equivalency Studies

### *NNF, QTF, and QDF*

Alternate form reliability data of the Fall, Winter, and Spring benchmark forms were gathered as part of the national item calibration field test study. For this study, four alternate forms each of NNF and QTF and three alternate forms of QDF were evaluated, with all alternate forms developed from a common blueprint, summarized as follows:

- NNF: Each form consisted of 80 Arabic numerals, ranging from 0 to 20. The total score equaled the number of numerals correctly named in 1 minute.
- QTF: Each form consisted of 38 items presenting a box or a pair of boxes containing dots. Dots were arranged like the dots on dice, with up to six dots in each box and the maximum total number of dots displayed per item was 10. The student indicated the total number of dots for each item. The total score equaled the number of items answered correctly in 1 minute.
- QDF: Each form consisted of 24 items presenting two boxes containing dots, one with blue dots and one with red dots. Dots were arranged like the dots on dice, with the blue dots ranging between 1 and 5 and red dots ranging from 2 to 6. For each item, the box with blue dots always contained fewer dots than the box with red dots. The student indicated how many more blue dots were needed to match the number of red dots for each item. The total score equaled the number of items answered correctly in 1 minute.

For this study, six test sets were used. Each set consisted of two NNF, two QTF, and two QDF forms, as well as 25 Concepts & Applications items. The order of measures for each set was: NNF(1), QTF(1), QDF(1), CA, NNF(2), QTF(2), QDF(2). Each fluency measure was assigned to two test sets in counter-balanced sequence such that if a fluency form (e.g., NNF) appeared *before* CA in the first set, then it appeared *after* CA in the second set.

A spiraling approach was used to assign students to test sets. In total, 635 students completed all seven test forms in all the sets. With approximately 105 students completing each set, this resulted in about 210 students completed each of the four alternate forms per fluency measure. Note that an administration error occurred with two of the QTF forms, which resulted in a loss of about 100 cases.

### *NCF–P, MFF–1D, and MFF–T*

Alternate form reliability data of the Fall, Winter, and Spring benchmark forms were gathered as part of the national item calibration field test study. For this study, four alternate forms each of NCF–P and MFF–1D and three alternate forms of MFF–T were evaluated, with all alternate forms developed from a common blueprint, summarized as follows:

- NCF–P: Each form consisted of 50 pairs of Arabic numerals, with numbers ranging from 0 to 99. The total score equaled the number of items answered correctly in 1 minute.
- MFF–1D: Each form consisted of 40 addition and subtraction problems involving numbers 0 through 10. The total score equaled the number of items answered correctly in 1 minute.
- MFF–T: Each form consisted of 32 items involving the addition and subtraction of 10. The total score equaled the number of items answered correctly in 1 minute.

For this study, six test sets were used. Each set consisted of two NCF–P, two MFF–1D, and two MFF–T forms, as well as 25 Concepts & Applications items. The order of measures for each set was: NCF–P(1), MFF–1D(1), MFF–T(1), CA, NCF–P(2), MFF–1D(2), MFF–T(2). Each fluency measure was assigned to two test sets in counter-balanced sequence such that if a fluency form (e.g., NCF–P) appeared *before* CA in the first set, then it appeared *after* CA in the second set.

A spiraling approach was used to assign students to test sets. In total, 606 students completed all seven test forms in all the sets. With approximately 100 students completing each set, this resulted in about 200 students completed each of the four alternate forms per fluency measure.

## Early Literacy Equivalency Studies

### *LWSF*

In the Winter testing season, 536 Kindergarten students completed one set of four alternate LWSF forms. Each of the 10 sets included the Grade 1 Fall LWSF benchmark form as the anchor form, with the remaining three forms per set being drawn from the 14 alternate forms developed for LWSF. Note that the Grade 1 Fall LSWF benchmark form was developed from the same blueprint used in the Winter and Spring of Kindergarten. Each group of three alternate forms was assigned to two of the ten equivalency study sets, with the order of the first and third forms reversed across the sets. In each set, the anchor form was always administered first. This approach was used to control for order effects and sampling variation.

Approximately 50 students completed each LWSF set. The correlation of the scores from the anchor form and the alternate forms was used to estimate reliability. The coefficient reported was computed from the weighted mean of the Fisher's *z*-transformed correlation coefficients.

### *LNF*

Alternate form reliability for this measure was computed from Fall, Winter, and Spring LNF benchmark scores obtained during the 2007–2008 school year. Due to the 4-month interval between benchmark administrations (i.e., fall to winter, winter to spring), the correlation coefficients represent lower bound estimates of reliability.

### *WRF*

To assess the equivalency of the six WRF forms, an equivalency study was conducted in which each form was assigned to two sets and each set comprised three forms. The order of forms was counterbalanced such that if a form appeared in the first position in one set, then it appeared in the third position in another set and vice versa. Forms assigned to the second position were assigned to that position in both sets it appeared in. This approach was used to control for order effects and sampling variation. For this study, 355 Grade 1 students completed three forms during the Spring testing window.

### *ORF*

Alternate form reliability coefficients for ORF were derived from benchmark data obtained during standardization. In each season, Grade 1 students read two passages aloud, each for 1 minute. The Spearman-Brown Prophecy formula was used to estimate reliability of the ***mean*** reading rate from the correlation of reading rates for the two passages.

## Math Equivalency Studies

### *NCF–T and MCF*

Alternate form reliability data of the three NCF–T and MCF benchmark forms was obtained as part of a larger study of the equivalency of NCF–T and MCF progress monitoring forms. For this study, students in Grades 2 through 8 completed a set of three NCF–T forms and three MCF forms. Fifteen sets were used for this study, with each set randomly assigned to students by spiraling sets within grade at each testing site. Sets 13 through 15 each contained all three benchmark forms, with the order of the forms completely counterbalanced across these three sets to control for order effects and sampling variation.

### *NSF*

Number Sense Fluency (NSF) is a composite derived from the sum of NCF–T and MCF scores. As such, NSF alternate form reliabilities are based on this sum and include only students who had a valid score on both NCF–T and MCF. The NSF score is the basis for all progress monitoring decisions.

## Reading Equivalency Studies

### *ORF*

Alternate form reliability coefficients for ORF were derived from benchmark data obtained during standardization. In each season, Grade 2 through 8 students read two passages aloud, each for 1 minute. The Spearman-Brown Prophecy formula was used to estimate reliability of the *mean* reading rate from the correlation of reading rates for the two passages. Results of single story reliabilities are shown in Table 20.

### *SRF*

Silent Reading Fluency reading rates are based on the median rate from three stories. Because there is no formulaic approach to estimate the reliability of a median, a simulation study was conducted. For this study, 10 replications of 1,000 cases per grade were simulated, using the observed correlations between pairs of stories for each grade and benchmark period and variance. Six scores were simulated for each student using the MVTNORM package in R. The median score on variables 1 to 3 was then correlated with the median score on variables 4 to 6 to yield the alternate-form reliability of the median of three stories. The average pairwise correlation of reading rates among single stories in SRF across Grades 4 through 8 is 0.75 (see Table 21), while the average reliability of the median of three stories is 0.87 (see Table 22).

**Table 16** Reliability of Early Numeracy Measures and Composites, Kindergarten and Grade 1

| Measure | Grade | Season | Cronbach's alpha | | Alternate form | | | Stratified alpha | *SEM* |
|---|---|---|---|---|---|---|---|---|---|
| | | | *n* | Coefficient | *n* range | Coefficient mean | Range | | |
| NNF | K | F, W, S | -- | -- | 201–207 | 0.90 | 0.88–0.90 | -- | 4.63 |
| QTF | K | F, W, S | -- | -- | 93–206 | 0.80 | 0.77–0.81 | -- | 2.01 |
| QDF | K | W, S | -- | -- | 201–203 | 0.74 | 0.71–0.76 | -- | 2.48 |
| CA | K | F | 1378 | 0.83 | -- | -- | -- | -- | 2.09 |
| CA | K | W | 1217 | 0.83 | -- | -- | -- | -- | 1.93 |
| CA | K | S | 1193 | 0.83 | -- | -- | -- | -- | 1.76 |
| Composite | K | F | -- | -- | -- | -- | -- | 0.88 | 4.10 |
| Composite | K | W | -- | -- | -- | -- | -- | 0.91 | 4.41 |
| Composite | K | S | -- | -- | -- | -- | -- | 0.91 | 4.41 |
| NCF–P | 1 | F, W, S | -- | -- | 222–239 | 0.88 | 0.86–0.89 | -- | 2.47 |
| MFF–1D | 1 | F, W, S | -- | -- | 217–234 | 0.86 | 0.86–0.89 | -- | 2.33 |
| MFF–T | 1 | W, S | -- | -- | 167–175 | 0.93 | 0.93 | -- | 1.25 |
| CA | 1 | F | 1459 | 0.86 | -- | -- | -- | -- | 2.12 |
| CA | 1 | W | 1259 | 0.87 | -- | -- | -- | -- | 1.88 |
| CA | 1 | S | 1364 | 0.88 | -- | -- | -- | -- | 1.75 |
| Composite | 1 | F | -- | -- | -- | -- | -- | 0.96 | 3.30 |
| Composite | 1 | W | -- | -- | -- | -- | -- | 0.97 | 3.29 |
| Composite | 1 | S | -- | -- | -- | -- | -- | 0.97 | 3.38 |

**Table 17** Reliability of Early Literacy Measures and Composites, Kindergarten and Grade 1

| Measure | Grade | Season | Cronbach's alpha | | Alternate form | | | Stratified alpha | *SEM* |
|---|---|---|---|---|---|---|---|---|---|
| | | | *n* | Coefficient | *n* range | Coefficient mean | Range | | |
| LNF | K | F, W, S | -- | -- | 655–672 | 0.78 | 0.73–0.82 | -- | 8.74 |
| LWSF | K | PM | -- | -- | 90–217 | 0.87 | 0.84–0.90 | -- | 5.39 |
| IS | K | F | 1256 | 0.88 | -- | -- | -- | -- | 1.35 |
| IS | K | W | 1221 | 0.87 | -- | -- | -- | -- | 0.82 |
| PC | K | F | 1256 | 0.63 | -- | -- | -- | -- | 1.06 |
| PS | K | W | 1238 | 0.93 | -- | -- | -- | -- | 3.46 |
| PS | K | S | 1221 | 0.87 | -- | -- | -- | -- | 3.82 |
| AV | K | F | 1256 | 0.82 | -- | -- | -- | -- | 1.65 |
| AV | K | W | 1221 | 0.81 | -- | -- | -- | -- | 1.27 |
| AV | K | S | 1238 | 0.76 | -- | -- | -- | -- | 1.34 |
| Composite | K | W | -- | -- | -- | -- | -- | 0.93 | 10.09 |
| Composite | K | S | -- | -- | -- | -- | -- | 0.91 | 10.41 |
| WRF | 1 | F, W, S | -- | -- | 173–180 | 0.94 | 0.93–0.95 | -- | 5.51 |
| ORF | 1 | F | -- | -- | 1341 | 0.97 | -- | -- | 5.21 |
| ORF | 1 | W | -- | -- | 1389 | 0.96 | -- | -- | 6.53 |
| ORF | 1 | S | -- | -- | 1502 | 0.96 | -- | -- | 7.11 |
| PS | 1 | F | 1329 | 0.83 | -- | -- | -- | -- | 3.84 |
| AV | 1 | F | 1346 | 0.85 | -- | -- | -- | -- | 0.96 |
| AV | 1 | W | 1390 | 0.87 | -- | -- | -- | -- | 0.88 |
| AV | 1 | S | 1503 | 0.87 | -- | -- | -- | -- | 0.60 |
| Composite | 1 | F | -- | -- | -- | -- | -- | 0.95 | 8.13 |

**Table 18** Reliability of Math Measures and Composites, Grades 2 Through 8

| Measure | Grade | Season | Cronbach's alpha | | | Alternate form | | | Stratified alpha | | *SEM* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | *n* range | Coefficient mean | Range | *n* | Coefficient mean | Range | Coefficient mean | Range | |
| NCF–T | 2 | F, W, S | -- | -- | -- | 128 | 0.84 | 0.82–0.85 | -- | -- | 3.16 |
| NCF–T | 3 | F, W, S | -- | -- | -- | 140 | 0.91 | 0.91–0.92 | -- | -- | 3.10 |
| NCF–T | 4 | F, W, S | -- | -- | -- | 148 | 0.89 | 0.88–0.91 | -- | -- | 3.53 |
| NCF–T | 5 | F, W, S | -- | -- | -- | 145 | 0.86 | 0.85–0.87 | -- | -- | 4.05 |
| NCF–T | 6 | F, W, S | -- | -- | -- | 121 | 0.78 | 0.78–0.80 | -- | -- | 4.22 |
| NCF–T | 7 | F, W, S | -- | -- | -- | 115 | 0.78 | 0.76–0.80 | -- | -- | 4.05 |
| NCF–T | 8 | F, W, S | -- | -- | -- | 153 | 0.80 | 0.80–0.81 | -- | -- | 4.21 |
| MCF | 2 | F, W, S | -- | -- | -- | 122 | 0.85 | 0.84–0.87 | -- | -- | 3.48 |
| MCF | 3 | F, W, S | -- | -- | -- | 140 | 0.83 | 0.82–0.84 | -- | -- | 3.38 |
| MCF | 4 | F, W, S | -- | -- | -- | 140 | 0.87 | 0.87–0.88 | -- | -- | 2.69 |
| MCF | 5 | F, W, S | -- | -- | -- | 136 | 0.85 | 0.84–0.87 | -- | -- | 2.84 |
| MCF | 6 | F, W, S | -- | -- | -- | 120 | 0.87 | 0.86–0.89 | -- | -- | 3.65 |
| MCF | 7 | F, W, S | -- | -- | -- | 79 | 0.87 | 0.86–0.88 | -- | -- | 3.60 |
| MCF | 8 | F, W, S | -- | -- | -- | 124 | 0.91 | 0.90–0.92 | -- | -- | 3.26 |
| NSF | 2 | F, W, S | -- | -- | -- | 113 | 0.92 | 0.90–0.93 | -- | -- | 4.32 |
| NSF | 3 | F, W, S | -- | -- | -- | 131 | 0.93 | 0.92–0.94 | -- | -- | 4.45 |
| NSF | 4 | F, W, S | -- | -- | -- | 137 | 0.93 | 0.91–0.94 | -- | -- | 4.29 |
| NSF | 5 | F, W, S | -- | -- | -- | 132 | 0.91 | 0.91–0.92 | -- | -- | 5.00 |
| NSF | 6 | F, W, S | -- | -- | -- | 115 | 0.86 | 0.83–0.88 | -- | -- | 6.60 |
| NSF | 7 | F, W, S | -- | -- | -- | 77 | 0.88 | 0.87–0.89 | -- | -- | 6.03 |
| NSF | 8 | F, W, S | -- | -- | -- | 123 | 0.90 | 0.89–0.91 | -- | -- | 6.04 |
| CA | 2 | F, W, S | 1763–1962 | 0.85 | 0.85–0.86 | -- | -- | -- | -- | -- | 8.87 |
| CA | 3 | F, W, S | 1803–1875 | 0.85 | 0.83–0.86 | -- | -- | -- | -- | -- | 8.36 |
| CA | 4 | F, W, S | 1726–1925 | 0.77 | 0.74–0.82 | -- | -- | -- | -- | -- | 8.89 |
| CA | 5 | F, W, S | 1795–1977 | 0.80 | 0.80–0.81 | -- | -- | -- | -- | -- | 8.52 |
| CA | 6 | F, W, S | 1514–1736 | 0.81 | 0.77–0.84 | -- | -- | -- | -- | -- | 8.54 |
| CA | 7 | F, W, S | 1144–1409 | 0.82 | 0.81–0.83 | -- | -- | -- | -- | -- | 8.23 |
| CA | 8 | F, W, S | 1078–1299 | 0.79 | 0.77–0.82 | -- | -- | -- | -- | -- | 8.55 |
| Composite | 2 | F, W, S | -- | -- | -- | -- | -- | -- | 0.92 | 0.92–0.93 | 9.93 |
| Composite | 3 | F, W, S | -- | -- | -- | -- | -- | -- | 0.92 | 0.92–0.93 | 9.93 |
| Composite | 4 | F, W, S | -- | -- | -- | -- | -- | -- | 0.90 | 0.89–0.92 | 9.97 |
| Composite | 5 | F, W, S | -- | -- | -- | -- | -- | -- | 0.91 | 0.91 | 9.83 |
| Composite | 6 | F, W, S | -- | -- | -- | -- | -- | -- | 0.90 | 0.88–0.91 | 10.90 |
| Composite | 7 | F, W, S | -- | -- | -- | -- | -- | -- | 0.91 | 0.91–0.92 | 10.40 |
| Composite | 8 | F, W, S | -- | -- | -- | -- | -- | -- | 0.91 | 0.91–0.92 | 10.50 |

**Table 19** Reliability of Reading Measures and Composites, Grades 2 Through 8

| Measure | Grade | Season | Cronbach's alpha | | | Alternate form | | | Stratified alpha | | *SEM* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | *n* range | Coefficient mean | Range | *n* range | Coefficient mean | Range | Coefficient mean | Range | |
| ORF | 2 | F, W, S | -- | -- | -- | 1719–1900 | 0.96 | 0.95–0.97 | -- | -- | 7.78 |
| ORF | 3 | F, W, S | -- | -- | -- | 1580–1902 | 0.96 | 0.95–0.96 | -- | -- | 7.46 |
| ORF | 4 | F, W, S | -- | -- | -- | 1633–2014 | 0.95 | 0.95–0.96 | -- | -- | 8.40 |
| ORF | 5 | F, W, S | -- | -- | -- | 1649–2009 | 0.95 | 0.95 | -- | -- | 9.33 |
| ORF | 6 | F, W, S | -- | -- | -- | 1271–1449 | 0.95 | 0.94–0.96 | -- | -- | 8.54 |
| ORF | 7 | F, W, S | -- | -- | -- | 959–1097 | 0.94 | 0.94–0.95 | -- | -- | 9.58 |
| ORF | 8 | F, W, S | -- | -- | -- | 850–1051 | 0.95 | 0.94–0.96 | -- | -- | 8.13 |
| SRF | 4 | F, W, S | -- | -- | -- | 1857–2022 | 0.87 | -- | -- | -- | 17.09 |
| SRF | 5 | F, W, S | -- | -- | -- | 1926–2212 | 0.87 | -- | -- | -- | 14.90 |
| SRF | 6 | F, W, S | -- | -- | -- | 1322–1632 | 0.86 | -- | -- | -- | 18.31 |
| SRF | 7 | F, W, S | -- | -- | -- | 985–1238 | 0.87 | -- | -- | -- | 17.77 |
| SRF | 8 | F, W, S | -- | -- | -- | 939–1207 | 0.86 | -- | -- | -- | 17.59 |
| VO | 2 | F, W, S | 1842–2084 | 0.67 | 0.63–0.71 | -- | -- | -- | -- | -- | 13.88 |
| VO | 3 | F, W, S | 1839–1955 | 0.73 | 0.72–0.74 | -- | -- | -- | -- | -- | 11.43 |
| VO | 4 | F, W, S | 1874–2083 | 0.74 | 0.73–0.74 | -- | -- | -- | -- | -- | 10.91 |
| VO | 5 | F, W, S | 1910–2291 | 0.73 | 0.70–0.75 | -- | -- | -- | -- | -- | 11.03 |
| VO | 6 | F, W, S | 1358–1664 | 0.73 | 0.72–0.76 | -- | -- | -- | -- | -- | 10.93 |
| VO | 7 | F, W, S | 1003–1293 | 0.75 | 0.73–0.77 | -- | -- | -- | -- | -- | 10.68 |
| VO | 8 | F, W, S | 950–1241 | 0.82 | 0.80–0.83 | -- | -- | -- | -- | -- | 10.40 |
| RC | 2 | F, W, S | 1870–2053 | 0.86 | 0.85–0.88 | -- | -- | -- | -- | -- | 10.10 |
| RC | 3 | F, W, S | 1868–1937 | 0.87 | 0.86–0.89 | -- | -- | -- | -- | -- | 9.82 |
| RC | 4 | F, W, S | 1853–2002 | 0.84 | 0.82–0.86 | -- | -- | -- | -- | -- | 9.52 |
| RC | 5 | F, W, S | 1903–2117 | 0.85 | 0.84–0.87 | -- | -- | -- | -- | -- | 9.55 |
| RC | 6 | F, W, S | 1292–1535 | 0.84 | 0.84–0.85 | -- | -- | -- | -- | -- | 10.07 |
| RC | 7 | F, W, S | 978–1191 | 0.85 | 0.83–0.86 | -- | -- | -- | -- | -- | 9.36 |
| RC | 8 | F, W, S | 907–1143 | 0.84 | 0.83–0.85 | -- | -- | -- | -- | -- | 9.61 |
| Composite | 2 | F, W, S | -- | -- | -- | -- | -- | -- | 0.91 | 0.91–0.92 | 17.98 |
| Composite | 3 | F, W, S | -- | -- | -- | -- | -- | -- | 0.92 | 0.92–0.93 | 16.40 |
| Composite | 4 | F, W, S | -- | -- | -- | -- | -- | -- | 0.88 | 0.87–0.89 | 18.98 |
| Composite | 5 | F, W, S | -- | -- | -- | -- | -- | -- | 0.88 | 0.87–0.89 | 18.76 |
| Composite | 6 | F, W, S | -- | -- | -- | -- | -- | -- | 0.87 | 0.86–0.88 | 20.33 |
| Composite | 7 | F, W, S | -- | -- | -- | -- | -- | -- | 0.88 | 0.88 | 19.54 |
| Composite | 8 | F, W, S | -- | -- | -- | -- | -- | -- | 0.89 | 0.89–0.90 | 19.72 |

**Table 20** Average Alternate-Form Reliability of Single ORF Stories, by Grade and Season

| Grade | Fall | | Winter | | Spring | |
|---|---|---|---|---|---|---|
| | *n* | Correlation | *n* | Correlation | *n* | Correlation |
| 2 | 1741 | 0.94 | 1719 | 0.92 | 1900 | 0.91 |
| 3 | 1580 | 0.92 | 1692 | 0.90 | 1902 | 0.93 |
| 4 | 1633 | 0.93 | 1774 | 0.90 | 2014 | 0.91 |
| 5 | 1643 | 0.90 | 1812 | 0.90 | 2009 | 0.90 |
| 6 | 1317 | 0.92 | 1449 | 0.92 | 1271 | 0.89 |
| 7 | 959 | 0.88 | 1097 | 0.90 | 1011 | 0.88 |
| 8 | 850 | 0.88 | 1051 | 0.92 | 919 | 0.91 |
| Mean | | 0.91 | | 0.91 | | 0.90 |

**Table 21** Average Alternate-Form Reliability of Single SRF Stories, by Grade and Season

| | Fall | | Winter | | Spring | | |
|---|---|---|---|---|---|---|---|
| Grade | *n* | Reliability | *n* | Reliability | *n* | Reliability | Mean |
| 4 | 2022 | 0.77 | 1986 | 0.77 | 1857 | 0.71 | 0.75 |
| 5 | 2212 | 0.73 | 2015 | 0.76 | 1926 | 0.76 | 0.75 |
| 6 | 1632 | 0.79 | 1322 | 0.71 | 1374 | 0.74 | 0.75 |
| 7 | 1238 | 0.78 | 1066 | 0.75 | 985 | 0.70 | 0.74 |
| 8 | 1207 | 0.73 | 993 | 0.75 | 939 | 0.74 | 0.74 |
| Mean | | 0.76 | | 0.75 | | 0.73 | |

**Table 22** Reliability of the Median of Three SRF Story Reading Rates, by Grade

| | Average intercorrelation | | Reliability of the median |
|---|---|---|---|
| Grade | Stories 1–3 | Stories 4–6 | |
| 4 | 0.75 | 0.75 | 0.87 |
| 5 | 0.75 | 0.76 | 0.87 |
| 6 | 0.75 | 0.75 | 0.86 |
| 7 | 0.74 | 0.74 | 0.87 |
| 8 | 0.74 | 0.73 | 0.86 |
| Mean | 0.75 | 0.75 | 0.87 |

## Equivalency of Progress Monitoring Forms

Tables 23 through 44 report the average difficulty of each progress monitoring form obtained from data collected during the equivalency studies described above. Each table includes the measure abbreviation, grade level, the progress monitoring form number that is used in the aimswebPlus system, the sample size of students taking each form in the equivalency study, the mean score as a measure of difficulty, the standard deviation (*SD*), and the effect size (ES).

For each form, the reported effect size is the standard deviation unit difference between the form's mean and the overall mean presented below the table. This method is computed as follows:

$$\frac{(Form\ mean - Total\ Mean)}{SD}$$

Where the *SD* is the average *SD* reported below each measure's table.

Each table also reports the percentage of variation in test scores attributed to form. This percentage is the ratio of the variance of the means divided by the total score variance. The ratio is multiplied by 100 to generate the reported percentage provided in each table.

For all measures, test form variance accounts for less than 5% of the total variance and most are less than 1%. Most effect sizes are less than 0.1 and nearly all are less than 0.3, which is the commonly used threshold indicating a small effect.

**Table 23** Statistical Equivalency of Progress Monitoring Forms, Number Naming Fluency

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| NNF | K | 4 | 43 | 46.8 | 10.45 | 0.04 |
| NNF | K | 5 | 132 | 45.5 | 12.08 | 0.06 |
| NNF | K | 6 | 114 | 45.4 | 16.34 | 0.07 |
| NNF | K | 7 | 132 | 45.9 | 11.41 | 0.03 |
| NNF | K | 8 | 44 | 48.4 | 13.00 | 0.16 |
| NNF | K | 9 | 135 | 48.4 | 15.75 | 0.16 |
| NNF | K | 10 | 43 | 45.5 | 12.09 | 0.06 |
| NNF | K | 11 | 51 | 45.6 | 11.99 | 0.05 |
| NNF | K | 12 | 58 | 47.8 | 15.28 | 0.12 |
| NNF | K | 13 | 136 | 46.2 | 13.50 | 0.01 |
| NNF | K | 14 | 92 | 45.0 | 12.24 | 0.10 |
| NNF | K | 15 | 67 | 46.4 | 13.50 | 0.01 |
| NNF | K | 16 | 135 | 47.0 | 15.07 | 0.05 |
| NNF | K | 17 | 67 | 45.1 | 13.00 | 0.09 |
| NNF | K | 18 | 106 | 45.9 | 12.18 | 0.03 |
| NNF | K | 19 | 136 | 45.0 | 13.09 | 0.10 |
| NNF | K | 20 | 54 | 46.5 | 13.13 | 0.02 |
| NNF | K | 21 | 44 | 47.1 | 12.46 | 0.06 |
| NNF | K | 22 | 51 | 46.5 | 11.82 | 0.02 |
| NNF | K | 23 | 54 | 45.9 | 12.77 | 0.03 |
| | | | Mean | 46.3 | 13.1 | 0.06 |
| | | | *SD* | 1.04 | 1.51 | |
| | | | Variance | 0.59% | | |

**Table 24** Statistical Equivalency of Progress Monitoring Forms, Quantity Total Fluency

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| QTF | K | 4 | 54 | 16.3 | 4.47 | 0.16 |
| QTF | K | 5 | 135 | 17.0 | 4.56 | 0.01 |
| QTF | K | 6 | 136 | 17.4 | 3.92 | 0.10 |
| QTF | K | 7 | 43 | 17.2 | 4.16 | 0.05 |
| QTF | K | 8 | 106 | 17.0 | 5.08 | 0.01 |
| QTF | K | 9 | 51 | 17.3 | 3.49 | 0.08 |
| QTF | K | 10 | 51 | 17.2 | 3.26 | 0.05 |
| QTF | K | 11 | 92 | 16.9 | 5.00 | 0.02 |
| QTF | K | 12 | 58 | 16.8 | 4.87 | 0.04 |
| QTF | K | 13 | 58 | 17.4 | 4.81 | 0.10 |
| QTF | K | 14 | 136 | 17.1 | 4.53 | 0.03 |
| QTF | K | 15 | 114 | 17.2 | 5.07 | 0.05 |
| QTF | K | 16 | 132 | 16.3 | 3.62 | 0.16 |
| QTF | K | 17 | 67 | 17.2 | 4.32 | 0.05 |
| QTF | K | 18 | 135 | 16.9 | 5.06 | 0.02 |
| QTF | K | 19 | 114 | 16.6 | 4.45 | 0.09 |
| QTF | K | 20 | 44 | 17.3 | 2.98 | 0.08 |
| QTF | K | 21 | 132 | 16.8 | 3.71 | 0.04 |
| QTF | K | 22 | 92 | 16.4 | 4.89 | 0.14 |
| QTF | K | 23 | 44 | 17.2 | 2.83 | 0.05 |
| | | | Mean | 17.0 | 4.3 | 0.07 |
| | | | *SD* | 0.35 | 0.72 | |
| | | | Variance | 0.58% | | |

**Table 25** Statistical Equivalency of Progress Monitoring Forms, Quantity Difference Fluency

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| QDF | K | 4 | 136 | 9.7 | 3.86 | 0.06 |
| QDF | K | 5 | 114 | 11.0 | 5.30 | 0.24 |
| QDF | K | 6 | 114 | 10.1 | 4.45 | 0.03 |
| QDF | K | 7 | 132 | 11.3 | 4.22 | 0.30 |
| QDF | K | 8 | 106 | 9.2 | 4.60 | 0.17 |
| QDF | K | 9 | 136 | 9.2 | 4.26 | 0.17 |
| QDF | K | 10 | 92 | 9.7 | 3.85 | 0.06 |
| QDF | K | 11 | 135 | 9.4 | 5.10 | 0.13 |
| QDF | K | 12 | 132 | 10.1 | 4.66 | 0.03 |
| QDF | K | 13 | 92 | 9.9 | 3.85 | 0.01 |
| | | | Mean | 10.0 | 4.4 | 0.12 |
| | | | *SD* | 0.71 | 0.51 | |
| | | | Variance | 2.33% | | |

**Table 26** Statistical Equivalency of Progress Monitoring Forms, Number Comparison Fluency–Pairs

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| NCF–P | 1 | 4 | 57 | 27.2 | 6.28 | 0.18 |
| NCF–P | 1 | 5 | 28 | 28.6 | 9.61 | 0.02 |
| NCF–P | 1 | 6 | 62 | 29.6 | 4.26 | 0.17 |
| NCF–P | 1 | 7 | 37 | 28.6 | 6.14 | 0.02 |
| NCF–P | 1 | 8 | 47 | 28.5 | 7.40 | 0.01 |
| NCF–P | 1 | 9 | 53 | 29.4 | 6.66 | 0.14 |
| NCF–P | 1 | 10 | 71 | 28.9 | 5.65 | 0.07 |
| NCF–P | 1 | 11 | 50 | 28.1 | 7.71 | 0.05 |
| NCF–P | 1 | 12 | 63 | 27.7 | 6.17 | 0.11 |
| NCF–P | 1 | 13 | 39 | 29.5 | 7.01 | 0.16 |
| NCF–P | 1 | 14 | 39 | 28.6 | 5.46 | 0.02 |
| NCF–P | 1 | 15 | 62 | 29.2 | 4.40 | 0.11 |
| NCF–P | 1 | 16 | 47 | 27.2 | 6.85 | 0.18 |
| NCF–P | 1 | 17 | 37 | 28.6 | 7.18 | 0.02 |
| NCF–P | 1 | 18 | 70 | 28.2 | 6.66 | 0.04 |
| NCF–P | 1 | 19 | 57 | 27.7 | 7.76 | 0.11 |
| NCF–P | 1 | 20 | 28 | 29.3 | 9.41 | 0.13 |
| NCF–P | 1 | 21 | 52 | 28.0 | 6.47 | 0.07 |
| NCF–P | 1 | 22 | 50 | 27.3 | 7.12 | 0.17 |
| NCF–P | 1 | 23 | 63 | 28.8 | 7.07 | 0.05 |
| | | | Mean | 28.5 | 6.8 | 0.09 |
| | | | *SD* | 0.76 | 1.30 | |
| | | | Variance | 1.25% | | |

**Table 27** Statistical Equivalency of Progress Monitoring Forms, Math Facts Fluency–1 Digit

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| MFF–1D | 1 | 4 | 63 | 16.0 | 5.32 | 0.10 |
| MFF–1D | 1 | 5 | 39 | 14.9 | 5.10 | 0.10 |
| MFF–1D | 1 | 6 | 70 | 14.8 | 6.51 | 0.11 |
| MFF–1D | 1 | 7 | 39 | 15.0 | 5.81 | 0.08 |
| MFF–1D | 1 | 8 | 50 | 16.7 | 5.81 | 0.22 |
| MFF–1D | 1 | 9 | 63 | 16.9 | 5.24 | 0.26 |
| MFF–1D | 1 | 10 | 57 | 14.8 | 6.30 | 0.11 |
| MFF–1D | 1 | 11 | 52 | 14.9 | 5.56 | 0.10 |
| MFF–1D | 1 | 12 | 71 | 16.6 | 5.46 | 0.20 |
| MFF–1D | 1 | 13 | 37 | 15.1 | 5.50 | 0.06 |
| MFF–1D | 1 | 14 | 47 | 14.7 | 6.95 | 0.13 |
| MFF–1D | 1 | 15 | 52 | 15.3 | 5.57 | 0.02 |
| MFF–1D | 1 | 16 | 70 | 16.1 | 7.55 | 0.12 |
| MFF–1D | 1 | 17 | 53 | 14.7 | 5.56 | 0.13 |
| MFF–1D | 1 | 18 | 50 | 15.0 | 6.21 | 0.08 |
| MFF–1D | 1 | 19 | 37 | 14.7 | 5.59 | 0.13 |
| MFF–1D | 1 | 20 | 62 | 15.6 | 4.20 | 0.03 |
| MFF–1D | 1 | 21 | 71 | 16.4 | 4.60 | 0.17 |
| MFF–1D | 1 | 22 | 57 | 15.6 | 6.17 | 0.03 |
| MFF–1D | 1 | 23 | 62 | 15.0 | 4.43 | 0.08 |
| | | | Mean | 15.4 | 5.7 | 0.11 |
| | | | *SD* | 0.75 | 0.81 | |
| | | | Variance | 1.67% | | |

**Table 28** Statistical Equivalency of Progress Monitoring Forms, Math Facts Fluency–Tens

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| MFF–T | 1 | 4 | 37 | 6.7 | 4.30 | 0.02 |
| MFF–T | 1 | 5 | 28 | 7.1 | 6.05 | 0.07 |
| MFF–T | 1 | 6 | 39 | 6.8 | 4.40 | 0.00 |
| MFF–T | 1 | 7 | 28 | 7.3 | 5.85 | 0.11 |
| MFF–T | 1 | 8 | 52 | 6.3 | 4.26 | 0.10 |
| MFF–T | 1 | 9 | 47 | 6.5 | 4.85 | 0.06 |
| MFF–T | 1 | 10 | 37 | 6.7 | 4.10 | 0.02 |
| MFF–T | 1 | 11 | 53 | 6.6 | 4.63 | 0.04 |
| MFF–T | 1 | 12 | 47 | 7.1 | 4.90 | 0.07 |
| MFF–T | 1 | 13 | 39 | 6.8 | 4.22 | 0.00 |
| | | | Mean | 6.8 | 4.8 | 0.05 |
| | | | *SD* | 0.30 | 0.68 | |
| | | | Variance | 0.37% | | |

**Table 29** Statistical Equivalency of Progress Monitoring Forms, Letter Word Sounds Fluency

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| LWSF | K | 4 | 124 | 43.8 | 14.80 | 0.04 |
| LWSF | K | 5 | 105 | 44.4 | 14.30 | 0.08 |
| LWSF | K | 6 | 90 | 42.5 | 14.10 | 0.05 |
| LWSF | K | 7 | 217 | 44.2 | 15.70 | 0.06 |
| LWSF | K | 8 | 90 | 40.9 | 15.00 | 0.16 |
| LWSF | K | 9 | 104 | 43.5 | 14.60 | 0.02 |
| LWSF | K | 10 | 90 | 41.4 | 14.80 | 0.13 |
| LWSF | K | 11 | 105 | 43.8 | 14.00 | 0.04 |
| LWSF | K | 12 | 124 | 44.8 | 15.10 | 0.10 |
| LWSF | K | 13 | 105 | 43.4 | 14.80 | 0.01 |
| | | | Mean | 43.3 | 14.7 | 0.07 |
| | | | *SD* | 1.28 | 0.51 | |
| | | | Variance | 0.61% | | |

**Table 30** Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 1)

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| ORF | 1 | 7 | 55 | 64.7 | 37.9 | 0.05 |
| ORF | 1 | 8 | 45 | 68.9 | 39.0 | 0.07 |
| ORF | 1 | 9 | 44 | 67.8 | 32.9 | 0.04 |
| ORF | 1 | 10 | 44 | 67.0 | 36.6 | 0.02 |
| ORF | 1 | 11 | 42 | 62.3 | 35.7 | 0.12 |
| ORF | 1 | 12 | 22 | 61.8 | 27.7 | 0.13 |
| ORF | 1 | 13 | 42 | 69.9 | 24.7 | 0.10 |
| ORF | 1 | 14 | 22 | 70.9 | 38.4 | 0.13 |
| ORF | 1 | 15 | 20 | 68.2 | 32.7 | 0.05 |
| ORF | 1 | 16 | 20 | 68.9 | 32.3 | 0.07 |
| ORF | 1 | 17 | 44 | 68.2 | 36.6 | 0.05 |
| ORF | 1 | 18 | 41 | 68.4 | 36.5 | 0.06 |
| ORF | 1 | 19 | 21 | 65.8 | 27.7 | 0.01 |
| ORF | 1 | 20 | 55 | 60.2 | 38.8 | 0.18 |
| ORF | 1 | 21 | 54 | 68.2 | 39.3 | 0.05 |
| ORF | 1 | 22 | 20 | 63.4 | 34.2 | 0.08 |
| ORF | 1 | 23 | 54 | 62.5 | 37.7 | 0.11 |
| ORF | 1 | 24 | 42 | 68.9 | 34.0 | 0.07 |
| ORF | 1 | 25 | 63 | 62.5 | 30.8 | 0.11 |
| ORF | 1 | 26 | 42 | 67.8 | 36.2 | 0.04 |
| | | | Mean | 66.3 | 34.5 | 0.08 |
| | | | *SD* | 3.15 | 4.16 | |
| | | | Variance | 0.78% | | |

**Table 31** Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 2)

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| ORF | 2 | 7 | 63 | 73.8 | 39.1 | 0.06 |
| ORF | 2 | 8 | 63 | 67.1 | 35.9 | 0.23 |
| ORF | 2 | 9 | 63 | 75.8 | 37.5 | 0.01 |
| ORF | 2 | 10 | 64 | 66.6 | 44.5 | 0.24 |
| ORF | 2 | 11 | 64 | 83.9 | 43.2 | 0.20 |
| ORF | 2 | 12 | 62 | 81.6 | 37.9 | 0.14 |
| ORF | 2 | 13 | 62 | 69.7 | 36.4 | 0.16 |
| ORF | 2 | 14 | 63 | 84.4 | 41.4 | 0.21 |
| ORF | 2 | 15 | 59 | 67.1 | 36.1 | 0.23 |
| ORF | 2 | 16 | 62 | 78.0 | 36.3 | 0.05 |
| ORF | 2 | 17 | 63 | 71.1 | 41.3 | 0.13 |
| ORF | 2 | 18 | 62 | 85.2 | 39.5 | 0.23 |
| ORF | 2 | 19 | 62 | 84.7 | 43.8 | 0.22 |
| ORF | 2 | 20 | 59 | 74.9 | 33.6 | 0.03 |
| ORF | 2 | 21 | 62 | 75.0 | 38.5 | 0.03 |
| ORF | 2 | 22 | 59 | 70.3 | 33.6 | 0.15 |
| ORF | 2 | 23 | 64 | 78.2 | 42.3 | 0.06 |
| ORF | 2 | 24 | 63 | 81.2 | 37.3 | 0.13 |
| ORF | 2 | 25 | 64 | 65.9 | 43.2 | 0.26 |
| ORF | 2 | 26 | 63 | 86.1 | 40.8 | 0.26 |
| | | | Mean | 76.0 | 39.1 | 0.15 |
| | | | *SD* | 6.94 | 3.33 | |
| | | | Variance | 3.77% | | |

**Table 32** Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 3)

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| ORF | 3 | 7 | 65 | 106.0 | 40.8 | 0.13 |
| ORF | 3 | 8 | 68 | 104.7 | 34.8 | 0.09 |
| ORF | 3 | 9 | 68 | 106.1 | 26.7 | 0.13 |
| ORF | 3 | 10 | 65 | 103.5 | 37.1 | 0.06 |
| ORF | 3 | 11 | 65 | 98.5 | 38.6 | 0.08 |
| ORF | 3 | 12 | 64 | 95.8 | 35.7 | 0.16 |
| ORF | 3 | 13 | 68 | 96.2 | 33.5 | 0.15 |
| ORF | 3 | 14 | 66 | 105.9 | 33.8 | 0.12 |
| ORF | 3 | 15 | 61 | 101.0 | 35.7 | 0.01 |
| ORF | 3 | 16 | 65 | 101.4 | 37.1 | 0.00 |
| ORF | 3 | 17 | 66 | 108.3 | 34.8 | 0.19 |
| ORF | 3 | 18 | 64 | 104.9 | 39.7 | 0.10 |
| ORF | 3 | 19 | 66 | 100.8 | 35.5 | 0.02 |
| ORF | 3 | 20 | 61 | 95.0 | 34.6 | 0.18 |
| ORF | 3 | 21 | 64 | 109.0 | 38.4 | 0.21 |
| ORF | 3 | 22 | 61 | 88.7 | 36.4 | 0.36 |
| ORF | 3 | 23 | 66 | 96.5 | 32.6 | 0.14 |
| ORF | 3 | 24 | 64 | 112.6 | 36.7 | 0.31 |
| ORF | 3 | 25 | 61 | 93.4 | 36.8 | 0.22 |
| ORF | 3 | 26 | 68 | 100.5 | 37.2 | 0.03 |
| | | | Mean | 101.4 | 35.8 | 0.13 |
| | | | *SD* | 5.98 | 2.98 | |
| | | | Variance | 2.60% | | |

**Table 33** Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 4)

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| ORF | 4 | 7 | 69 | 113.2 | 38.7 | 0.20 |
| ORF | 4 | 8 | 65 | 115.5 | 37.3 | 0.14 |
| ORF | 4 | 9 | 70 | 128.8 | 29.4 | 0.19 |
| ORF | 4 | 10 | 65 | 125.3 | 44.1 | 0.11 |
| ORF | 4 | 11 | 65 | 121.4 | 45.9 | 0.01 |
| ORF | 4 | 12 | 65 | 126.5 | 44.0 | 0.14 |
| ORF | 4 | 13 | 65 | 116.4 | 40.8 | 0.12 |
| ORF | 4 | 14 | 65 | 115.5 | 39.4 | 0.14 |
| ORF | 4 | 15 | 70 | 120.2 | 35.3 | 0.02 |
| ORF | 4 | 16 | 65 | 118.7 | 41.4 | 0.06 |
| ORF | 4 | 17 | 65 | 117.1 | 38.1 | 0.10 |
| ORF | 4 | 18 | 61 | 125.2 | 38.4 | 0.10 |
| ORF | 4 | 19 | 65 | 130.0 | 43.4 | 0.22 |
| ORF | 4 | 20 | 69 | 119.4 | 42.3 | 0.05 |
| ORF | 4 | 21 | 65 | 125.5 | 36.2 | 0.11 |
| ORF | 4 | 22 | 70 | 114.5 | 37.8 | 0.17 |
| ORF | 4 | 23 | 61 | 123.1 | 37.1 | 0.05 |
| ORF | 4 | 24 | 70 | 122.9 | 34.4 | 0.04 |
| ORF | 4 | 25 | 61 | 115.8 | 37.4 | 0.14 |
| ORF | 4 | 26 | 65 | 128.5 | 43.5 | 0.19 |
| | | | Mean | 121.2 | 39.2 | 0.12 |
| | | | *SD* | 5.27 | 4.00 | |
| | | | Variance | 1.68% | | |

**Table 34** Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 5)

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| ORF | 5 | 7 | 52 | 141.3 | 38.9 | 0.30 |
| ORF | 5 | 8 | 54 | 126.2 | 35.8 | 0.11 |
| ORF | 5 | 9 | 56 | 125.4 | 36.3 | 0.13 |
| ORF | 5 | 10 | 54 | 124.3 | 32.6 | 0.16 |
| ORF | 5 | 11 | 54 | 141.1 | 35.7 | 0.29 |
| ORF | 5 | 12 | 52 | 135.7 | 37.4 | 0.15 |
| ORF | 5 | 13 | 52 | 119.8 | 33.8 | 0.28 |
| ORF | 5 | 14 | 56 | 129.4 | 38.9 | 0.02 |
| ORF | 5 | 15 | 52 | 137.6 | 37.3 | 0.20 |
| ORF | 5 | 16 | 56 | 125.7 | 32.9 | 0.12 |
| ORF | 5 | 17 | 52 | 123.6 | 35.3 | 0.18 |
| ORF | 5 | 18 | 52 | 137.7 | 40.8 | 0.20 |
| ORF | 5 | 19 | 57 | 125.7 | 41.5 | 0.12 |
| ORF | 5 | 20 | 56 | 120.3 | 36.1 | 0.27 |
| ORF | 5 | 21 | 54 | 139.8 | 39.1 | 0.26 |
| ORF | 5 | 22 | 56 | 133.4 | 37.6 | 0.09 |
| ORF | 5 | 23 | 57 | 140.2 | 40.1 | 0.27 |
| ORF | 5 | 24 | 56 | 129.5 | 38.0 | 0.02 |
| ORF | 5 | 25 | 52 | 126.8 | 36.4 | 0.09 |
| ORF | 5 | 26 | 56 | 120.8 | 35.9 | 0.25 |
| | | | Mean | 130.2 | 37.0 | 0.18 |
| | | | *SD* | 7.44 | 2.44 | |
| | | | Variance | 3.65% | | |

**Table 35** Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 6)

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| ORF | 6 | 7 | 66 | 156.2 | 41.6 | 0.09 |
| ORF | 6 | 8 | 68 | 147.3 | 33.9 | 0.15 |
| ORF | 6 | 9 | 66 | 157.4 | 39.8 | 0.13 |
| ORF | 6 | 10 | 68 | 155.1 | 35.6 | 0.06 |
| ORF | 6 | 11 | 64 | 157.8 | 34.0 | 0.14 |
| ORF | 6 | 12 | 64 | 145.4 | 29.3 | 0.20 |
| ORF | 6 | 13 | 66 | 152.8 | 42.3 | 0.00 |
| ORF | 6 | 14 | 65 | 147.8 | 37.7 | 0.13 |
| ORF | 6 | 15 | 66 | 158.4 | 40.1 | 0.15 |
| ORF | 6 | 16 | 66 | 143.8 | 39.3 | 0.24 |
| ORF | 6 | 17 | 66 | 154.3 | 44.0 | 0.04 |
| ORF | 6 | 18 | 68 | 151.7 | 38.9 | 0.03 |
| ORF | 6 | 19 | 67 | 158.1 | 32.3 | 0.14 |
| ORF | 6 | 20 | 64 | 159.4 | 32.1 | 0.18 |
| ORF | 6 | 21 | 67 | 150.3 | 38.6 | 0.07 |
| ORF | 6 | 22 | 66 | 144.1 | 38.7 | 0.23 |
| ORF | 6 | 23 | 64 | 146.5 | 34.7 | 0.17 |
| ORF | 6 | 24 | 66 | 145.5 | 36.0 | 0.20 |
| ORF | 6 | 25 | 67 | 160.5 | 35.0 | 0.21 |
| ORF | 6 | 26 | 65 | 162.4 | 37.3 | 0.26 |
| | | | Mean | 152.7 | 37.1 | 0.14 |
| | | | *SD* | 6.01 | 3.75 | |
| | | | Variance | 2.40% | | |

**Table 36** Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 7)

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| ORF | 7 | 7 | 51 | 145.5 | 38.4 | 0.19 |
| ORF | 7 | 8 | 55 | 139.3 | 26.6 | 0.35 |
| ORF | 7 | 9 | 51 | 153.7 | 39.7 | 0.02 |
| ORF | 7 | 10 | 55 | 150.5 | 44.1 | 0.06 |
| ORF | 7 | 11 | 55 | 156.5 | 38.3 | 0.09 |
| ORF | 7 | 12 | 51 | 153.5 | 38.9 | 0.01 |
| ORF | 7 | 13 | 51 | 162.1 | 39.9 | 0.24 |
| ORF | 7 | 14 | 51 | 165.2 | 42.9 | 0.32 |
| ORF | 7 | 15 | 57 | 157.8 | 43.5 | 0.13 |
| ORF | 7 | 16 | 55 | 144.0 | 41.8 | 0.23 |
| ORF | 7 | 17 | 55 | 151.9 | 47.9 | 0.03 |
| ORF | 7 | 18 | 55 | 151.3 | 46.8 | 0.04 |
| ORF | 7 | 19 | 55 | 147.5 | 28.3 | 0.14 |
| ORF | 7 | 20 | 55 | 165.2 | 40.0 | 0.32 |
| ORF | 7 | 21 | 51 | 158.1 | 38.6 | 0.13 |
| ORF | 7 | 22 | 55 | 137.2 | 34.8 | 0.41 |
| ORF | 7 | 23 | 55 | 156.8 | 35.8 | 0.10 |
| ORF | 7 | 24 | 51 | 153.4 | 34.5 | 0.01 |
| ORF | 7 | 25 | 51 | 153.0 | 39.7 | 0.00 |
| ORF | 7 | 26 | 57 | 156.2 | 34.6 | 0.08 |
| | | | Mean | 152.9 | 38.8 | 0.15 |
| | | | *SD* | 7.55 | 5.37 | |
| | | | Variance | 3.44% | | |

**Table 37** Statistical Equivalency of Progress Monitoring Forms, Oral Reading Fluency (Grade 8)

| Measure | Grade | Form | *n* | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| ORF | 8 | 7 | 58 | 161.8 | 42.4 | 0.25 |
| ORF | 8 | 8 | 56 | 152.1 | 32.5 | 0.01 |
| ORF | 8 | 9 | 58 | 160.2 | 42.8 | 0.21 |
| ORF | 8 | 10 | 56 | 143.3 | 31.2 | 0.20 |
| ORF | 8 | 11 | 56 | 150.8 | 42.2 | 0.02 |
| ORF | 8 | 12 | 56 | 151.1 | 42.2 | 0.01 |
| ORF | 8 | 13 | 56 | 140.3 | 26.4 | 0.28 |
| ORF | 8 | 14 | 58 | 152.3 | 38.3 | 0.02 |
| ORF | 8 | 15 | 53 | 159.3 | 47.2 | 0.19 |
| ORF | 8 | 16 | 56 | 152.6 | 43.2 | 0.03 |
| ORF | 8 | 17 | 58 | 151.4 | 40.5 | 0.00 |
| ORF | 8 | 18 | 52 | 146.7 | 38.6 | 0.12 |
| ORF | 8 | 19 | 52 | 150.6 | 47.7 | 0.02 |
| ORF | 8 | 20 | 56 | 138.3 | 41.5 | 0.33 |
| ORF | 8 | 21 | 53 | 161.8 | 48.2 | 0.25 |
| ORF | 8 | 22 | 52 | 151.7 | 43.4 | 0.00 |
| ORF | 8 | 23 | 56 | 153.9 | 52.9 | 0.06 |
| ORF | 8 | 24 | 56 | 154.1 | 33.2 | 0.06 |
| ORF | 8 | 25 | 53 | 144.5 | 43.8 | 0.17 |
| ORF | 8 | 26 | 56 | 154.6 | 37.2 | 0.07 |
| | | | Mean | 151.6 | 40.8 | 0.12 |
| | | | *SD* | 6.5 | 6.4 | |
| | | | Variance | 2.35% | | |

**Table 38** Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 2)

| Measure | Grade | Form | *n* range | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| NSF | 2 | 4 | 59–63 | 17.4 | 16.97 | 0.02 |
| NSF | 2 | 5 | 58–59 | 17.5 | 18.75 | 0.03 |
| NSF | 2 | 6 | 57–60 | 18.9 | 18.53 | 0.11 |
| NSF | 2 | 7 | 51–56 | 15.8 | 15.16 | 0.07 |
| NSF | 2 | 8 | 59 | 16.7 | 17.00 | 0.02 |
| NSF | 2 | 9 | 57–61 | 17.3 | 16.40 | 0.02 |
| NSF | 2 | 10 | 56–63 | 17.8 | 17.70 | 0.05 |
| NSF | 2 | 11 | 59–62 | 16.1 | 15.44 | 0.05 |
| NSF | 2 | 12 | 58 | 16.8 | 17.36 | 0.01 |
| NSF | 2 | 13 | 50–58 | 15.9 | 15.04 | 0.06 |
| NSF | 2 | 14 | 51–58 | 18.7 | 15.08 | 0.10 |
| NSF | 2 | 15 | 59–60 | 17.1 | 15.70 | 0.01 |
| NSF | 2 | 16 | 57–58 | 17.0 | 19.29 | 0.00 |
| NSF | 2 | 17 | 58–61 | 16.7 | 17.91 | 0.02 |
| NSF | 2 | 18 | 57–58 | 16.4 | 20.76 | 0.03 |
| NSF | 2 | 19 | 59–60 | 17.2 | 15.57 | 0.01 |
| NSF | 2 | 20 | 50–62 | 15.6 | 16.97 | 0.08 |
| NSF | 2 | 21 | 57–58 | 16.5 | 15.97 | 0.03 |
| NSF | 2 | 22 | 56–59 | 18.4 | 15.26 | 0.08 |
| NSF | 2 | 23 | 56–59 | 15.8 | 16.57 | 0.07 |
| | | | Mean | 17.0 | 16.9 | 0.04 |
| | | | *SD* | 0.96 | 1.58 | |
| | | | Variance | 0.29% | | |

**Table 39** Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 3)

| Measure | Grade | Form | *n* range | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| NSF | 3 | 4 | 52–59 | 24.6 | 19.77 | 0.02 |
| NSF | 3 | 5 | 63–64 | 25.5 | 18.60 | 0.06 |
| NSF | 3 | 6 | 58–64 | 25.7 | 20.22 | 0.07 |
| NSF | 3 | 7 | 51–62 | 23.9 | 20.14 | 0.02 |
| NSF | 3 | 8 | 53–62 | 23.8 | 16.01 | 0.03 |
| NSF | 3 | 9 | 54 | 24.3 | 23.13 | 0.00 |
| NSF | 3 | 10 | 55–57 | 25.5 | 19.17 | 0.06 |
| NSF | 3 | 11 | 54–57 | 24.0 | 18.40 | 0.02 |
| NSF | 3 | 12 | 54 | 24.1 | 23.42 | 0.01 |
| NSF | 3 | 13 | 55–63 | 22.4 | 18.01 | 0.10 |
| NSF | 3 | 14 | 51–62 | 26.2 | 20.21 | 0.10 |
| NSF | 3 | 15 | 55–61 | 24.6 | 17.40 | 0.02 |
| NSF | 3 | 16 | 58–70 | 24.1 | 18.63 | 0.01 |
| NSF | 3 | 17 | 52–67 | 23.4 | 20.17 | 0.05 |
| NSF | 3 | 18 | 61–63 | 23.5 | 16.06 | 0.04 |
| NSF | 3 | 19 | 53–60 | 24.9 | 20.81 | 0.03 |
| NSF | 3 | 20 | 53–63 | 23.7 | 16.79 | 0.03 |
| NSF | 3 | 21 | 56–62 | 24.0 | 16.42 | 0.02 |
| NSF | 3 | 22 | 55–60 | 25.1 | 17.81 | 0.04 |
| NSF | 3 | 23 | 57–61 | 22.7 | 17.28 | 0.08 |
| | | | Mean | 24.3 | 18.9 | 0.04 |
| | | | *SD* | 0.98 | 2.10 | |
| | | | Variance | 0.25% | | |

**Table 40** Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 4)

| Measure | Grade | Form | *n* range | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| NSF | 4 | 4 | 58–68 | 27.6 | 19.90 | 0.02 |
| NSF | 4 | 5 | 58–70 | 29.2 | 17.98 | 0.07 |
| NSF | 4 | 6 | 63–66 | 30.2 | 19.72 | 0.12 |
| NSF | 4 | 7 | 65 | 27.1 | 18.79 | 0.04 |
| NSF | 4 | 8 | 58–63 | 26.6 | 17.97 | 0.07 |
| NSF | 4 | 9 | 58–68 | 28.6 | 20.48 | 0.04 |
| NSF | 4 | 10 | 67–70 | 29.1 | 18.55 | 0.06 |
| NSF | 4 | 11 | 61–67 | 27.7 | 18.93 | 0.01 |
| NSF | 4 | 12 | 57–70 | 28.4 | 22.13 | 0.03 |
| NSF | 4 | 13 | 65–67 | 25.8 | 18.41 | 0.11 |
| NSF | 4 | 14 | 61–67 | 31.1 | 22.57 | 0.16 |
| NSF | 4 | 15 | 66–72 | 27.2 | 18.62 | 0.04 |
| NSF | 4 | 16 | 57–70 | 27.7 | 21.70 | 0.01 |
| NSF | 4 | 17 | 66–68 | 26.4 | 18.10 | 0.08 |
| NSF | 4 | 18 | 58–64 | 27.3 | 20.62 | 0.03 |
| NSF | 4 | 19 | 67–68 | 27.8 | 20.49 | 0.01 |
| NSF | 4 | 20 | 58–72 | 26.7 | 16.91 | 0.06 |
| NSF | 4 | 21 | 61–67 | 28.2 | 17.59 | 0.02 |
| NSF | 4 | 22 | 61–65 | 29.2 | 20.64 | 0.07 |
| NSF | 4 | 23 | 65–74 | 26.1 | 18.28 | 0.09 |
| | | | Mean | 27.9 | 19.4 | 0.06 |
| | | | *SD* | 1.37 | 1.59 | |
| | | | Variance | 0.49% | | |

**Table 41** Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 5)

| Measure | Grade | Form | *n* range | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| NSF | 5 | 4 | 57–67 | 29.0 | 18.19 | 0.02 |
| NSF | 5 | 5 | 57–60 | 29.9 | 17.74 | 0.06 |
| NSF | 5 | 6 | 57–64 | 31.0 | 18.83 | 0.12 |
| NSF | 5 | 7 | 70–71 | 27.9 | 19.76 | 0.04 |
| NSF | 5 | 8 | 59–70 | 27.2 | 18.41 | 0.08 |
| NSF | 5 | 9 | 61–62 | 29.4 | 22.14 | 0.04 |
| NSF | 5 | 10 | 63–71 | 30.1 | 20.10 | 0.07 |
| NSF | 5 | 11 | 63–64 | 27.6 | 20.17 | 0.06 |
| NSF | 5 | 12 | 64–70 | 30.0 | 22.05 | 0.07 |
| NSF | 5 | 13 | 61–62 | 26.6 | 19.08 | 0.11 |
| NSF | 5 | 14 | 58–63 | 32.6 | 20.54 | 0.20 |
| NSF | 5 | 15 | 61–64 | 28.4 | 20.20 | 0.01 |
| NSF | 5 | 16 | 57–67 | 28.1 | 19.69 | 0.03 |
| NSF | 5 | 17 | 58–68 | 27.9 | 18.74 | 0.04 |
| NSF | 5 | 18 | 64–66 | 28.2 | 18.38 | 0.03 |
| NSF | 5 | 19 | 66 | 28.6 | 19.91 | 0.00 |
| NSF | 5 | 20 | 58–63 | 26.9 | 18.12 | 0.09 |
| NSF | 5 | 21 | 60–71 | 27.9 | 19.40 | 0.04 |
| NSF | 5 | 22 | 59–60 | 29.8 | 19.71 | 0.06 |
| NSF | 5 | 23 | 61–66 | 26.6 | 16.77 | 0.11 |
| | | | Mean | 28.7 | 19.4 | 0.06 |
| | | | *SD* | 1.55 | 1.33 | |
| | | | Variance | 0.59% | | |

**Table 42** Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 6)

| Measure | Grade | Form | *n* range | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| NSF | 6 | 4 | 65–75 | 25.2 | 20.28 | 0.02 |
| NSF | 6 | 5 | 53–70 | 26.0 | 17.82 | 0.06 |
| NSF | 6 | 6 | 51–80 | 26.5 | 19.78 | 0.09 |
| NSF | 6 | 7 | 48–49 | 23.9 | 16.89 | 0.06 |
| NSF | 6 | 8 | 52–77 | 24.0 | 17.16 | 0.05 |
| NSF | 6 | 9 | 47–51 | 25.8 | 15.47 | 0.05 |
| NSF | 6 | 10 | 48–80 | 26.2 | 18.18 | 0.07 |
| NSF | 6 | 11 | 79–80 | 23.8 | 19.67 | 0.06 |
| NSF | 6 | 12 | 69–86 | 25.0 | 19.18 | 0.01 |
| NSF | 6 | 13 | 43–80 | 23.3 | 17.54 | 0.09 |
| NSF | 6 | 14 | 51–52 | 27.3 | 17.30 | 0.14 |
| NSF | 6 | 15 | 56–86 | 24.7 | 16.25 | 0.01 |
| NSF | 6 | 16 | 56–73 | 24.9 | 20.08 | 0.00 |
| NSF | 6 | 17 | 43–58 | 24.1 | 15.15 | 0.04 |
| NSF | 6 | 18 | 69–77 | 25.0 | 19.66 | 0.01 |
| NSF | 6 | 19 | 56–82 | 25.6 | 16.10 | 0.04 |
| NSF | 6 | 20 | 65–82 | 23.4 | 19.14 | 0.08 |
| NSF | 6 | 21 | 49–73 | 24.0 | 19.45 | 0.05 |
| NSF | 6 | 22 | 46–47 | 25.8 | 15.81 | 0.05 |
| NSF | 6 | 23 | 51–75 | 23.2 | 16.75 | 0.09 |
| | | | Mean | 24.9 | 17.9 | 0.05 |
| | | | *SD* | 1.16 | 1.67 | |
| | | | Variance | 0.37% | | |

**Table 43** Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 7)

| Measure | Grade | Form | *n* range | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| NSF | 7 | 4 | 51–66 | 21.4 | 18.25 | 0.05 |
| NSF | 7 | 5 | 60–68 | 22.0 | 19.89 | 0.08 |
| NSF | 7 | 6 | 35–68 | 22.5 | 18.25 | 0.11 |
| NSF | 7 | 7 | 60–71 | 20.0 | 16.40 | 0.03 |
| NSF | 7 | 8 | 64–66 | 19.9 | 17.41 | 0.03 |
| NSF | 7 | 9 | 42–54 | 20.8 | 17.45 | 0.02 |
| NSF | 7 | 10 | 62–66 | 21.9 | 19.51 | 0.08 |
| NSF | 7 | 11 | 62–71 | 19.4 | 17.56 | 0.06 |
| NSF | 7 | 12 | 42–61 | 20.1 | 16.83 | 0.02 |
| NSF | 7 | 13 | 41–61 | 18.1 | 16.92 | 0.13 |
| NSF | 7 | 14 | 39–66 | 23.0 | 18.87 | 0.14 |
| NSF | 7 | 15 | 54–59 | 20.9 | 17.53 | 0.02 |
| NSF | 7 | 16 | 51–64 | 20.8 | 19.93 | 0.02 |
| NSF | 7 | 17 | 39–45 | 19.8 | 17.60 | 0.04 |
| NSF | 7 | 18 | 41–49 | 19.2 | 18.36 | 0.07 |
| NSF | 7 | 19 | 39–64 | 21.2 | 18.98 | 0.04 |
| NSF | 7 | 20 | 39–54 | 19.3 | 18.43 | 0.07 |
| NSF | 7 | 21 | 50–60 | 19.8 | 17.82 | 0.04 |
| NSF | 7 | 22 | 59–61 | 21.9 | 17.28 | 0.08 |
| NSF | 7 | 23 | 49–64 | 18.2 | 17.81 | 0.13 |
| | | | Mean | 20.5 | 18.1 | 0.06 |
| | | | *SD* | 1.36 | 0.99 | |
| | | | Variance | 0.50% | | |

**Table 44** Statistical Equivalency of Progress Monitoring Forms, Number Sense Fluency (Grade 8)

| Measure | Grade | Form | *n* range | Mean | *SD* | ES |
|---|---|---|---|---|---|---|
| NSF | 8 | 4 | 64–65 | 20.6 | 21.05 | 0.02 |
| NSF | 8 | 5 | 55–73 | 21.0 | 20.80 | 0.04 |
| NSF | 8 | 6 | 61–67 | 21.8 | 21.62 | 0.08 |
| NSF | 8 | 7 | 55–72 | 19.7 | 21.45 | 0.02 |
| NSF | 8 | 8 | 53–64 | 19.5 | 22.21 | 0.03 |
| NSF | 8 | 9 | 52–67 | 20.7 | 18.78 | 0.03 |
| NSF | 8 | 10 | 59–67 | 21.3 | 19.12 | 0.06 |
| NSF | 8 | 11 | 58–72 | 19.3 | 19.59 | 0.04 |
| NSF | 8 | 12 | 61–74 | 20.3 | 18.61 | 0.01 |
| NSF | 8 | 13 | 58–64 | 18.9 | 20.31 | 0.06 |
| NSF | 8 | 14 | 53–64 | 22.1 | 19.20 | 0.10 |
| NSF | 8 | 15 | 58–59 | 20.3 | 19.44 | 0.01 |
| NSF | 8 | 16 | 58–65 | 20.2 | 19.75 | 0.00 |
| NSF | 8 | 17 | 53–64 | 19.8 | 18.98 | 0.02 |
| NSF | 8 | 18 | 61–62 | 19.6 | 18.29 | 0.03 |
| NSF | 8 | 19 | 62–64 | 20.3 | 20.99 | 0.01 |
| NSF | 8 | 20 | 55–74 | 18.5 | 21.02 | 0.08 |
| NSF | 8 | 21 | 55–62 | 19.3 | 19.05 | 0.04 |
| NSF | 8 | 22 | 65 | 21.4 | 22.44 | 0.06 |
| NSF | 8 | 23 | 52–73 | 18.7 | 18.63 | 0.07 |
| | | | Mean | 20.2 | 20.1 | 0.04 |
| | | | *SD* | 1.02 | 1.29 | |
| | | | Variance | 0.24% | | |

# Validity

During the 2013–2014 standardization study, Pearson obtained achievement scores for participating students from other reading and math tests used by each school. As a condition of participation, schools provided spring test scores from interim assessments, state NCLB tests or other formative assessments. A secure file transfer protocol was used to share data, with test scores being provided to Pearson without individually identifiable information. A unique, randomly derived student ID assigned by Pearson was used to match each participant's scores to standardization data.

This section presents the concurrent and predictive validity coefficients obtained from these data from criterion measures and aimswebPlus. Concurrent validity represents the correlation of aimswebPlus composite scores and criterion measure scores, both from the Spring testing season. Predictive validity represents the correlation of Fall aimswebPlus composite scores and Spring scores from the criterion measures.

Predicting student achievement in the Spring from Fall benchmark scores is the basis for determining a student's risk status. The National Center on Intensive Intervention (NCII) requires predictive validity coefficients of 0.70 or higher to obtain the maximum rating (i.e., providing *convincing evidence*) for screeners. However, there is not a single universally accepted standard for defining success and many different tests are used across U.S. schools; thus, it is important to evaluate predictive validity with several criterion measures.

When a test shows strong prediction with several different criterion measures, there is greater confidence that results can be generalized to other standardized and validated measures of student achievement. In the sections that follow, concurrent and predictive validity coefficients for aimswebPlus Early Numeracy, Early Literacy, Math, and Reading benchmark composites are provided.

Each validity table presented shows the unadjusted and adjusted validity coefficients, as well as the mean adjusted coefficients by grade. The adjusted coefficients represent an estimate of the true population coefficient, which takes into account the effects that variation of sample characteristics has on the score variance of the predictor. All things being equal, an increase in score variance will result in larger coefficients. As such, the adjusted validity coefficient is a more accurate estimate of the true population coefficient. This adjustment is computed as:

$$Adjusted\ r_{xy} = \left(\frac{r_{xy}SD_u}{SD_r}\right) / \sqrt{1 - r_{xy}^2 + r_{xy}^2\left(\frac{SD_u^2}{SD_r^2}\right)}$$

Where $SD_u$ is the population standard deviation of the aimswebPlus composite from the national norms, and $SD_r$ is the sample standard deviation of the aimswebPlus composite. The average adjusted validity coefficient is the mean of the adjusted validity coefficients, by grade. The mean is weighted by the sample size of each coefficient.

## Early Numeracy Criterion Validity

Table 45 shows the predictive validity coefficients of the aimswebPlus Early Numeracy composite scores with the Tennessee Comprehensive Achievement Program (TCAP) math scores. TCAP assesses math skills aligned to Tennessee's state learning standards. The characteristics of the sample upon which the coefficient was obtained are also provided.

Table 46 shows the concurrent validity coefficients for the aimswebPlus Early Numeracy composites with TCAP math scores. The aimswebPlus Early Numeracy scores were collected in May 2014, while TCAP scores were obtained in late April 2014.

**Table 45** Early Numeracy Composite Score Predictive Validity Coefficients, by Grade, Season, and Criterion Measure

| | | | Predictive | | Sex | | Race/Ethnicity | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Criterion | Grade | *n* | Unadjusted | Adjusted | % Female | % Male | % Black | % Hispanic | % Other | % White |
| TCAP | K (Fall) | 68 | 0.62 | 0.70 | 41 | 59 | 9 | 12 | 0 | 79 |
| TCAP | K (Winter) | 68 | 0.70 | 0.76 | 41 | 59 | 9 | 12 | 0 | 79 |
| TCAP | 1 (Fall) | 55 | 0.79 | 0.86 | 53 | 47 | 2 | 25 | 0 | 73 |
| TCAP | 1 (Winter) | 55 | 0.80 | 0.87 | 53 | 47 | 2 | 25 | 0 | 73 |

**Table 46** Early Numeracy Spring Composite Score Concurrent Validity Coefficients, by Grade and Criterion Measure

| | | | Predictive | | Sex | | Race/Ethnicity | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Criterion | Grade | *n* | Unadjusted | Adjusted | % Female | % Male | % Black | % Hispanic | % Other | % White |
| TCAP | K (Spring) | 68 | 0.66 | 0.73 | 41 | 59 | 9 | 12 | 0 | 79 |
| TCAP | 1 (Spring) | 55 | 0.68 | 0.79 | 53 | 47 | 2 | 25 | 0 | 73 |

## Early Literacy Criterion Validity

An important outcome of Kindergarten early literacy instruction is to move students from elementary phonological awareness, such as letter identification and letter sounds, to word reading and eventually to reading connected text in the form of sentences and short stories. Thus, the aimswebPlus measure Word Reading Fluency is used as the predictive criterion measure of Fall and Winter Kindergarten scores. Word Reading Fluency assesses a student's automaticity with reading high frequency and highly decodable words. Students are given 1 minute to read as many words as possible.

In the Fall testing season of Kindergarten, aimswebPlus requires only Letter Naming Fluency for assessing risk status. This measure was selected because research shows it to be a strong predictor of end-of-year oral reading fluency ability (Clemens et al., 2015) and because it is a very appropriate measure of foundational reading skills in the beginning Kindergarten. By midyear, Kindergarten students typically have had formal instruction on letter identification, letters sounds, and parsing simple words into phonemes. As such, the aimwebPlus Early Literacy Winter composite for Kindergarten also includes Letter Word Sounds Fluency and Phoneme Segmentation. The composite of these three measures is used to identify risk and predict end-of-grade performance on Word Reading Fluency.

In Grade 1, early literacy instruction continues with a greater emphasis on word reading, as well as reading and comprehending connected text. For Grade 1 students, Oral Reading Fluency has been shown to provide strong prediction of end-of-grade performance on broad measures of reading. The Iowa Test of Basic Skills Level 6 measures vocabulary, word reading, and reading comprehension at the end of Grade 1, making it an appropriate criterion measure for ORF.

Table 47 shows the unadjusted and adjusted predictive validity coefficients of aimswebPlus LNF (Kindergarten, Fall), the composite comprised of LNF, LWSF, and PSF (Kindergarten, Winter), and ORF (Grade 1, Fall). The characteristics of the sample upon which the coefficient was obtained are also provided. Because WRF was administered to all Kindergarten students in the Spring testing season, data from this measure were used to obtain the validity coefficient.

Table 48 shows the concurrent validity coefficients for the composite comprised of LNF, LWSF, and PSF (Kindergarten, Spring) and ORF (Grade 1, Spring). ITBS scores were obtained in April 2014.

**Table 47** Early Literacy Predictive Validity Coefficients, by Grade, Season, and Criterion Measure

| | | | Predictive | | Sex | | Race/Ethnicity | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Criterion | Grade | *n* | Unadjusted | Adjusted | % Female | % Male | % Black | % Hispanic | % Other | % White |
| WRF | K (Fall) | 1075 | 0.58 | 0.58 | 50 | 50 | 14 | 25 | 10 | 51 |
| WRF | K (Winter) | 1075 | 0.63 | 0.63 | 50 | 50 | 14 | 25 | 10 | 51 |
| ITBS | 1 (Fall) | 61 | 0.57 | 0.72 | 41 | 59 | 25 | 25 | 17 | 33 |

**Table 48** Early Literacy Concurrent Validity Coefficients, by Grade and Criterion Measure

| | | | Predictive | | Sex | | Race/Ethnicity | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Criterion | Grade | *n* | Unadjusted | Adjusted | % Female | % Male | % Black | % Hispanic | % Other | % White |
| WRF | K (Spring) | 1075 | 0.57 | 0.57 | 50 | 50 | 14 | 25 | 10 | 51 |
| ITBS | 1 (Spring) | 61 | 0.67 | 0.74 | 41 | 59 | 25 | 25 | 17 | 33 |

## Math Criterion Validity

Five criterion measures were used to calculate criterion validity for aimswebPlus Math:

- Iowa Tests of Basic Skills®–Total Math (ITBS®)
- Illinois Standards Achievement Test (ISAT)
- New Mexico Standards Based Assessment (NMSBA)
- Northwest Evaluation Association Measures of Academic Progress® (NWEA–MAP®)
- State of Texas Academic Assessment of Readiness (STAAR)

The ITBS is a comprehensive, group-administered, paper-based assessment of reading and math achievement. ITBS's Total Math score reflects performance on standards-based math concepts, problem solving, and computation. The ISAT is the end-of-year achievement test assessing Illinois learning standards covering five math strands: Number Sense, Measurement, Algebra, Geometry, and Data Analysis and Probability. The NMSBA is used to measure student proficiency on New Mexico's reading and math learning standards. NWEA–MAP is a computer-adaptive test that assesses achievement in reading and

mathematics. Results are reported on an RIT scale, which is then linked to each state's performance standards. Finally, the STAAR assesses student performance on Texas's mathematics and reading learning standards.

Table 49 shows the predictive validity coefficients of the aimswebPlus Math composite with each criterion measure. Weighted mean validity coefficients, by grade, are also shown, which provides an estimate of the overall predictive validity. The characteristics of the sample upon which the coefficient was obtained are also provided.

Table 50 shows the concurrent validity coefficients for the aimswebPlus Math composite with each criterion measure, as well as the mean adjusted coefficients by grade. aimswebPlus Math scores were collected in May 2014, while the criterion measures scores were obtained in March through May 2014.

**Table 49** Math Composite Score Predictive Validity Coefficients, by Grade and Criterion Measure

| | | | Correlation | | | Sex | | Race/Ethnicity | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Criterion | Grade | *n* | Unadjusted | Adjusted | Mean | % Female | % Male | % Black | % Hispanic | % Other | % White |
| ITBS | 2 | 179 | 0.79 | 0.81 | 0.69 | 60 | 40 | 19 | 42 | 21 | 17 |
| NWEA–MAP | 2 | 218 | 0.62 | 0.56 | | 48 | 52 | 5 | 31 | 12 | 53 |
| ISAT | 3 | 69 | 0.85 | 0.81 | | 49 | 51 | 1 | 25 | 13 | 61 |
| NWEA–MAP | 3 | 101 | 0.83 | 0.79 | 0.79 | 46 | 54 | 1 | 40 | 14 | 44 |
| STAAR | 3 | 146 | 0.74 | 0.77 | | 55 | 45 | 10 | 39 | 37 | 14 |
| ISAT | 4 | 175 | 0.80 | 0.79 | | 51 | 49 | 4 | 28 | 9 | 58 |
| NWEA–MAP | 4 | 95 | 0.76 | 0.75 | 0.76 | 59 | 41 | 5 | 35 | 10 | 49 |
| STAAR | 4 | 207 | 0.75 | 0.73 | | 51 | 49 | 8 | 46 | 32 | 14 |
| ISAT | 5 | 189 | 0.86 | 0.84 | | 53 | 47 | 2 | 21 | 9 | 68 |
| NWEA–MAP | 5 | 81 | 0.89 | 0.86 | 0.83 | 47 | 53 | 3 | 43 | 11 | 43 |
| STAAR | 5 | 91 | 0.70 | 0.79 | | 49 | 51 | 2 | 52 | 41 | 6 |
| ISAT | 6 | 273 | 0.84 | 0.89 | | 59 | 41 | 22 | 6 | 8 | 64 |
| NMSBA | 6 | 210 | 0.75 | 0.80 | 0.85 | 52 | 48 | 3 | 64 | 1 | 32 |
| NWEA–MAP | 6 | 86 | 0.79 | 0.83 | | 55 | 45 | 22 | 9 | 10 | 59 |
| STAAR | 6 | 61 | 0.63 | 0.75 | | 55 | 45 | 5 | 44 | 48 | 3 |
| ISAT | 7 | 130 | 0.84 | 0.90 | | 45 | 55 | 13 | 2 | 3 | 82 |
| NMSBA | 7 | 220 | 0.78 | 0.78 | 0.85 | 47 | 53 | 2 | 62 | 0 | 36 |
| STAAR | 7 | 61 | 0.80 | 0.90 | | 40 | 60 | 5 | 43 | 49 | 4 |
| ISAT | 8 | 122 | 0.62 | 0.74 | | 37 | 63 | 5 | 1 | 3 | 91 |
| NMSBA | 8 | 223 | 0.84 | 0.87 | 0.83 | 44 | 56 | 6 | 67 | 1 | 26 |
| STAAR | 8 | 75 | 0.61 | 0.79 | | 61 | 39 | 15 | 53 | 32 | 0 |

**Table 50** Math Composite Score Concurrent Validity Coefficients, by Grade and Criterion Measure

| Criterion | Grade | *n* | Correlation | | | Sex | | Race/Ethnicity | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Unadjusted | Adjusted | Mean | % Female | % Male | % Black | % Hispanic | % Other | % White |
| ITBS | 2 | 218 | 0.82 | 0.81 | 0.77 | 60 | 40 | 19 | 42 | 21 | 17 |
| NWEA–MAP | 2 | 179 | 0.73 | 0.71 | | 48 | 52 | 5 | 31 | 12 | 53 |
| ISAT | 3 | 46 | 0.84 | 0.82 | 0.83 | 49 | 51 | 1 | 25 | 13 | 61 |
| NWEA–MAP | 3 | 101 | 0.87 | 0.85 | | 46 | 54 | 1 | 40 | 14 | 44 |
| STAAR | 3 | 211 | 0.76 | 0.82 | | 55 | 45 | 10 | 39 | 37 | 14 |
| ISAT | 4 | 126 | 0.85 | 0.83 | 0.79 | 51 | 49 | 4 | 28 | 9 | 58 |
| NWEA–MAP | 4 | 95 | 0.82 | 0.80 | | 59 | 41 | 5 | 35 | 10 | 49 |
| STAAR | 4 | 277 | 0.77 | 0.76 | | 51 | 49 | 8 | 46 | 32 | 14 |
| ISAT | 5 | 154 | 0.85 | 0.84 | 0.82 | 53 | 47 | 2 | 21 | 9 | 68 |
| NWEA–MAP | 5 | 81 | 0.84 | 0.84 | | 47 | 53 | 3 | 43 | 11 | 43 |
| STAAR | 5 | 157 | 0.72 | 0.80 | | 49 | 51 | 2 | 52 | 41 | 6 |
| ISAT | 6 | 231 | 0.85 | 0.88 | 0.85 | 59 | 41 | 22 | 6 | 8 | 64 |
| NMSBA | 6 | 210 | 0.77 | 0.85 | | 52 | 48 | 3 | 64 | 1 | 32 |
| NWEA–MAP | 6 | 86 | 0.74 | 0.76 | | 55 | 45 | 22 | 9 | 10 | 59 |
| STAAR | 6 | 61 | 0.68 | 0.79 | | 55 | 45 | 5 | 44 | 48 | 3 |
| ISAT | 7 | 130 | 0.78 | 0.83 | 0.84 | 45 | 55 | 13 | 2 | 3 | 82 |
| NMSBA | 7 | 220 | 0.76 | 0.85 | | 47 | 53 | 2 | 62 | 0 | 36 |
| STAAR | 7 | 61 | 0.74 | 0.84 | | 40 | 60 | 5 | 43 | 49 | 4 |
| ISAT | 8 | 122 | 0.68 | 0.73 | 0.82 | 37 | 63 | 5 | 1 | 3 | 91 |
| NMSBA | 8 | 223 | 0.80 | 0.87 | | 44 | 56 | 6 | 67 | 1 | 26 |
| STAAR | 8 | 75 | 0.56 | 0.77 | | 61 | 39 | 15 | 53 | 32 | 0 |

## Reading Criterion Validity

Four criterion measures were used to calculate criterion validity for aimswebPlus Reading:

- Illinois Standards Achievement Test (ISAT)
- Missouri Assessment Program Grade Level Assessment (MAP–GLA)
- Northwest Evaluation Association Measures of Academic Progress (NWEA–MAP)
- State of Texas Academic Assessment of Readiness (STAAR)

The ISAT is the end-of-year achievement test assessing Illinois learning standards, including reading comprehension. The MAP–GLA is the end-of-year achievement test that assesses Missouri reading and math standards, including reading comprehension. NWEA–MAP is a computer-adaptive test that assesses achievement in reading and mathematics. Results are reported on an RIT scale, which is then linked to each state's performance standards. Finally, the STAAR assesses student performance on Texas's mathematics and reading learning standards.

Table 51 shows the predictive validity coefficients of the aimswebPlus Reading composite with each criterion measure. Weighted mean validity coefficients, by grade, are also shown, which provides an estimate of the overall predictive validity. The characteristics of the sample upon which the coefficient was obtained are also provided.

Table 52 shows the concurrent validity coefficients for the aimswebPlus Reading composite with each criterion measure, as well as the mean adjusted coefficients by grade. aimswebPlus Math scores were collected in May 2014, while the criterion measures scores were obtained in March through May 2014.

**Table 51** Reading Composite Score Predictive Validity Coefficients, by Grade and Criterion Measure

| Criterion | Grade | *n* | Correlation | | | Sex | | Race/Ethnicity | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Unadjusted | Adjusted | Mean | % Female | % Male | % Black | % Hispanic | % Other | % White |
| NWEA–MAP | 2 | 128 | 0.83 | 0.83 | 0.83 | 52 | 48 | 2 | 23 | 21 | 53 |
| ISAT | 3 | 113 | 0.80 | 0.84 | 0.77 | 47 | 53 | 2 | 28 | 20 | 49 |
| MAP–GLA | 3 | 317 | 0.71 | 0.69 | | 55 | 45 | 24 | 2 | 2 | 72 |
| NWEA–MAP | 3 | 150 | 0.78 | 0.79 | | 45 | 55 | 2 | 25 | 20 | 52 |
| STAAR | 3 | 208 | 0.70 | 0.74 | | 56 | 44 | 10 | 49 | 14 | 27 |
| ISAT | 4 | 230 | 0.77 | 0.79 | 0.69 | 56 | 44 | 4 | 39 | 10 | 47 |
| MAP–GLA | 4 | 292 | 0.62 | 0.58 | | 49 | 51 | 32 | 1 | 5 | 62 |
| NWEA–MAP | 4 | 125 | 0.76 | 0.77 | | 53 | 47 | 4 | 28 | 16 | 52 |
| STAAR | 4 | 277 | 0.60 | 0.61 | | 44 | 56 | 8 | 52 | 10 | 29 |
| ISAT | 5 | 250 | 0.73 | 0.75 | 0.73 | 48 | 52 | 4 | 22 | 13 | 61 |
| MAP–GLA | 5 | 222 | 0.65 | 0.65 | | 50 | 50 | 42 | 0 | 7 | 50 |
| NWEA–MAP | 5 | 141 | 0.81 | 0.79 | | 48 | 52 | 3 | 30 | 18 | 48 |
| STAAR | 5 | 157 | 0.66 | 0.71 | | 53 | 47 | 9 | 57 | 3 | 31 |
| ISAT | 6 | 332 | 0.74 | 0.77 | 0.75 | 58 | 42 | 9 | 14 | 12 | 65 |
| NWEA–MAP | 6 | 124 | 0.67 | 0.73 | | 52 | 48 | 4 | 21 | 12 | 63 |
| ISAT | 7 | 179 | 0.78 | 0.81 | 0.73 | 44 | 56 | 12 | 12 | 7 | 68 |
| MAP–GLA | 7 | 101 | 0.71 | 0.78 | | 46 | 54 | 41 | 4 | 0 | 55 |
| NWEA–MAP | 7 | 207 | 0.51 | 0.61 | | 51 | 49 | 9 | 24 | 12 | 55 |
| ISAT | 8 | 202 | 0.72 | 0.80 | 0.78 | 46 | 54 | 10 | 11 | 6 | 74 |
| MAP–GLA | 8 | 218 | 0.69 | 0.76 | | 57 | 43 | 28 | 3 | 1 | 68 |

**Table 52** Reading Composite Score Concurrent Validity Coefficients, by Grade and Criterion Measure

| Criterion | Grade | *n* | Correlation | | | Sex | | Race/Ethnicity | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Unadjusted | Adjusted | Mean | % Female | % Male | % Black | % Hispanic | % Other | % White |
| NWEA–MAP | 2 | 128 | 0.80 | 0.80 | 0.80 | 52 | 48 | 2 | 23 | 21 | 53 |
| ISAT | 3 | 113 | 0.85 | 0.88 | 0.77 | 47 | 53 | 2 | 28 | 20 | 49 |
| MAP–GLA | 3 | 317 | 0.69 | 0.69 | | 55 | 45 | 24 | 2 | 2 | 72 |
| NWEA–MAP | 3 | 150 | 0.80 | 0.80 | | 45 | 55 | 2 | 25 | 20 | 52 |
| STAAR | 3 | 208 | 0.70 | 0.72 | | 56 | 44 | 10 | 49 | 14 | 27 |
| ISAT | 4 | 230 | 0.73 | 0.76 | 0.70 | 56 | 44 | 4 | 39 | 10 | 47 |
| MAP–GLA | 4 | 292 | 0.70 | 0.68 | | 49 | 51 | 32 | 1 | 5 | 62 |
| NWEA–MAP | 4 | 125 | 0.67 | 0.71 | | 53 | 47 | 4 | 28 | 16 | 52 |
| STAAR | 4 | 277 | 0.67 | 0.66 | | 44 | 56 | 8 | 52 | 10 | 29 |
| ISAT | 5 | 250 | 0.79 | 0.80 | 0.73 | 48 | 52 | 4 | 22 | 13 | 61 |
| MAP–GLA | 5 | 222 | 0.64 | 0.67 | | 50 | 50 | 42 | 0 | 7 | 50 |
| NWEA–MAP | 5 | 141 | 0.77 | 0.76 | | 48 | 52 | 3 | 30 | 18 | 48 |
| STAAR | 5 | 157 | 0.65 | 0.69 | | 53 | 47 | 9 | 57 | 3 | 31 |
| ISAT | 6 | 332 | 0.79 | 0.81 | 0.78 | 58 | 42 | 9 | 14 | 12 | 65 |
| NWEA–MAP | 6 | 124 | 0.72 | 0.74 | | 52 | 48 | 4 | 21 | 12 | 63 |
| ISAT | 7 | 179 | 0.78 | 0.80 | 0.68 | 44 | 56 | 12 | 12 | 7 | 68 |
| MAP–GLA | 7 | 101 | 0.64 | 0.67 | | 46 | 54 | 41 | 4 | 0 | 55 |
| NWEA–MAP | 7 | 207 | 0.50 | 0.57 | | 51 | 49 | 9 | 24 | 12 | 55 |
| ISAT | 8 | 202 | 0.72 | 0.79 | 0.76 | 46 | 54 | 10 | 11 | 6 | 74 |
| MAP–GLA | 8 | 218 | 0.69 | 0.72 | | 57 | 43 | 28 | 3 | 1 | 68 |

# Classification Accuracy

Educators want to know how well scores collected in the fall identify who is at risk of not attaining proficiency in the spring, so that they can provide those students with the resources and interventions to improve learning, close achievement gaps, and ultimately move them to proficiency by the end of the school year. Classification accuracy is a way to quantify how accurately scores on one test predict scores on a different, criterion test. More specifically, it refers to how accurately the predictor test classifies students as proficient or not proficient, according to the criterion test. In this section, classification accuracy results, based on the same data used for predictive validity, are provided.

Classification accuracy is an alternative means of expressing criterion validity that is appropriate when there is interest in predicting a dichotomous criterion (e.g., passing or not passing an end-of-year state test). A cut score on the predictor test (in this case, a given aimswebPlus measure) is chosen such that those who score at or above the cut score are considered likely to pass the criterion, while those who score below the cut score are likely to fail. A classification accuracy analysis indicates how frequently these expectations prove correct, and the results are reported in a variety of statistics.

Table 53 shows a two-by-two classification table. The columns indicate classification of proficiency based on the criterion (e.g., spring reading achievement test), and the rows indicate classification of proficiency based on the predictor (e.g., fall or winter aimswebPlus Reading composite). The four possible outcomes listed (TP, FP, FN, and TN) are defined as follows:

- TP is a true positive, meaning a student who passed the test was correctly predicted to pass.
- FP is a false positive, meaning a student who failed the test was incorrectly predicted to pass.
- FN is a false negative, meaning a student who passed the test was incorrectly predicted to fail.
- TN is a true negative, meaning a student who failed the test was correctly predicted to fail.

**Table 53** Classification Accuracy, Two-by-Two Model

| | | **Criterion Proficiency** | | |
|---|---|---|---|---|
| | | Yes (positive) | No (negative) | Row totals |
| **Predictor Proficiency** | Yes | TP | FP | **R1** |
| | No | FN | TN | **R2** |
| | Column totals | **P** | **N** | **Total** |

From each of these four prediction outcomes, several statistics can be derived and used to evaluate the accuracy of prediction. Table 54 lists the various classification accuracy statistics reported for aimswebPlus Reading and Math composite scores.

One key statistic is the overall accuracy rates, representing the percentage of students *correctly* classified by the predictor. This statistic directly answers the question of how accurately a test score classifies a student; however, overall accuracy rates depend on other statistics, such as base rate and the cut score chosen for the predictor, and even small changes in these values can significantly change overall accuracy rates.

Another statistic, known as the area under the curve (AUC), does *not* depend on base rates and cut scores; as such, this statistic can be used to compare the predictive accuracy of different predictors. AUC represents the total area under a curve formed from the relationship between the false positive rate and the true positive rate at each point from 0 to 1.0. (Note that AUC cannot be described with a simple formula.) AUCs greater than or equal to 0.85 are considered strong evidence of classification accuracy by the National Center on Intensive Intervention.

**Table 54** Classification Accuracy Statistics

| Statistic | Formula |
|---|---|
| False positive rate | FP ÷ N |
| False negative rate | FN ÷ P |
| Sensitivity | TP ÷ P |
| Specificity | TN ÷ N |
| Positive predictive power | TP ÷ R1 |
| Negative predictive power | TN ÷ R2 |
| Overall accuracy rate | (TP + TN) ÷ Total |
| Base rate | N ÷ Total |

## Early Numeracy Classification Accuracy

This section describes the classification accuracy of the aimswebPlus Fall and Winter Early Numeracy composite scores in Kindergarten and Grade 1 with Tennessee Comprehensive Achievement Program (TCAP) performance in the spring of 2014. TCAP assesses math and reading skills aligned to Tennessee's state learning standards. Because TCAP does not report proficiency levels below Grade 3, a cut score was defined such that the proficiency rate would approximate the proficiency rate observed in Grade 3, which was approximately 40% of students. Using aimswebPlus national percentiles to approximate this rate, the 40$^{th}$ national percentile was selected. Students scoring below the 40$^{th}$ national percentile were considered not proficient.

Classification accuracy results are shown in Table 55, by grade level. The base rate indicates the percentage of students not proficient. Using the criterion described above, 35% of students were not proficient in Kindergarten and 16% were not proficient in Grade 1. The relatively low base rate observed in Grade 1 indicates that the overall ability of the sample was above average. The overall classification accuracy rates range from 81% to 100%. The AUC is also very high, ranging from 0.90 to 1.00.

**Table 55** Classification Accuracy of Early Numeracy Composite Scores and TCAP

| | Kindergarten | | Grade 1 | |
|---|---|---|---|---|
| | Fall | Winter | Fall | Winter |
| False positive rate | 0.20 | 0.16 | 0 | 0.07 |
| False negative rate | 0.17 | 0.21 | 0 | 0 |
| Sensitivity | 0.83 | 0.79 | 1.00 | 1.00 |
| Specificity | 0.80 | 0.84 | 1.00 | 0.93 |
| Positive predictive power | 0.69 | 0.73 | 1.00 | 0.75 |
| Negative predictive power | 0.90 | 0.88 | 1.00 | 1.00 |
| Overall accuracy rate | 0.81 | 0.82 | 1.00 | 0.95 |
| Area under the curve | 0.90 | 0.92 | 1.00 | 0.99 |
| Base rate | 0.35 | 0.35 | 0.16 | 0.16 |
| aimswebPlus Fall cut score | 29 | 36 | 37 | 50 |
| Criterion Spring cut score | 491 | 491 | 534 | 534 |
| At 90% sensitivity, specificity equals | 0.68 | 0.73 | 0.89 | 0.89 |
| At 80% sensitivity, specificity equals | 0.80 | 0.84 | 0.78 | 0.78 |
| At 70% sensitivity, specificity equals | 0.86 | 0.91 | 0.67 | 0.67 |

## Early Literacy Classification Accuracy

This section describes the classification accuracy of the aimswebPlus Fall and Winter Early Literacy composite scores in Kindergarten, as well as for Fall Grade 1 Oral Reading Fluency (ORF) scores. In Kindergarten, two criterion measures were used: Spring Word Reading Fluency (WRF) scores and Spring R–CBM scores. WRF is a new aimswebPlus word reading CBM, while R–CBM is the original **aims**web oral reading fluency CBM. In Grade 1, spring scores on the Iowa Test of Basic Skills (ITBS) reading composite were used as the criterion. This total reading composite score includes foundational reading skills, listening comprehension, and reading comprehension.

The 25th Spring national percentile on WRF and the 25th Fall national percentile on R–CBM were defined as the criterion cut scores designating proficiency. For ITBS, a grade equivalent score of 1.5 was defined as the criterion cut score designating proficiency. This grade equivalent was chosen because it represents the median performance of students at the end of Grade 1.

Tables 56 and 57 show classification accuracy results for Kindergarten. In the Fall testing window, Letter Naming Fluency (LNF) is the predictor; meanwhile, in Winter, the predictor is a composite based on the sum of LNF, Letter Word Sounds Fluency (LWSF), and Phoneme Segmentation (PS) scores. Overall, classification accuracy rates range from 76% to 97%. The AUC is also very high, ranging from 0.82 to 0.99. Table 58 shows results for Grade 1. The overall accuracy rate for Fall ORF scores is 75%, with an AUC of 0.85.

**Table 56** Classification Accuracy of Fall Letter Naming Fluency with Winter Early Literacy Composite Scores and Word Reading Fluency (Kindergarten)

| | Season | |
|---|---|---|
| | Fall | Winter |
| False positive rate | 0.24 | 0.17 |
| False negative rate | 0.24 | 0.23 |
| Sensitivity | 0.76 | 0.77 |
| Specificity | 0.76 | 0.83 |
| Positive predictive power | 0.46 | 0.54 |
| Negative predictive power | 0.92 | 0.93 |
| Overall accuracy rate | 0.76 | 0.82 |
| Area under the curve | 0.82 | 0.87 |
| Base rate | 0.21 | 0.21 |
| aimswebPlus Fall cut score | 26 | 95 |
| Criterion Spring cut score | 8 | 8 |
| At 90% sensitivity, specificity equals | 0.56 | 0.62 |
| At 80% sensitivity, specificity equals | 0.70 | 0.76 |
| At 70% sensitivity, specificity equals | 0.80 | 0.89 |

**Table 57** Classification Accuracy of Fall Letter Naming Fluency with Winter Early Literacy Composite Scores and R–CBM (Kindergarten)

| | Season | |
|---|---|---|
| | Fall | Winter |
| False positive rate | 0.02 | 0.04 |
| False negative rate | 0.11 | 0.00 |
| Sensitivity | 0.89 | 1.00 |
| Specificity | 0.98 | 0.96 |
| Positive predictive power | 0.84 | 0.75 |
| Negative predictive power | 0.99 | 1.00 |
| Overall accuracy rate | 0.97 | 0.96 |
| Area under the curve | 0.98 | 0.99 |
| Base rate | 0.11 | 0.11 |
| aimswebPlus Fall cut score | 3 | 34 |
| Criterion Spring cut score | 18 | 18 |
| At 90% sensitivity, specificity equals | 0.98 | 0.96 |
| At 80% sensitivity, specificity equals | 0.98 | 1.00 |
| At 70% sensitivity, specificity equals | 1.00 | 1.00 |

**Table 58** Classification Accuracy of Fall Oral Reading Fluency Scores and Spring ITBS (Grade 1)

| | Season |
|---|---|
| | Fall |
| False positive rate | 0.28 |
| False negative rate | 0 |
| Sensitivity | 1.00 |
| Specificity | 0.72 |
| Positive predictive power | 0.29 |
| Negative predictive power | 1.00 |
| Overall accuracy rate | 0.75 |
| Area under the curve | 0.85 |
| Base rate | 0.10 |
| aimswebPlus Fall cut score | 26 |
| Criterion Spring cut score | 1.5 |
| At 90% sensitivity, specificity equals | 0.70 |
| At 80% sensitivity, specificity equals | 0.72 |
| At 70% sensitivity, specificity equals | 0.75 |

## Math Classification Accuracy

This section describes the classification accuracy of the aimswebPlus Fall Math composite score for Grades 2 through 8. To extend the generalizability of results, classification accuracy was evaluated with the following five different criterion measures: ITBS, ISAT, NMSBA, NWEA–MAP, and STAAR. Note that these measures are described in the Validity section of this manual.

Tables 59 through 62 show the classification accuracy results, by grade level. For the three state accountability criterion assessments (ISAT, STAAR, and NMSBA), spring benchmark performance levels were based on the cut score at or above which a student was designated as proficient in that state assessment system during the 2013–2014 school year, by grade level. Base rates, which range from the mid-0.20s to mid-0.40s, indicate the percentage of students who were not proficient on the state test.

The NWEA–MAP math cut scores were based on results provided in the NWEA linking study reports. NWEA conducts linking studies using data from students with MAP scores and state test scores. The linking study aligns NWEA's Rasch Unit (RIT) scale to the state test scale using equipercentile equating. For each state proficiency level, a RIT cut score is defined.

AUC values range from the upper-0.70s to the mid-0.90s. Approximately half of the AUCs exceed 0.85 and 85% exceed 0.80, the threshold for good classification accuracy.

**Table 59** Classification Accuracy of Math Composite Scores and ITBS (Grade 2) and ISAT (Grades 3–8)

| | Grade | | | | | | |
|---|---|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| False positive rate | 0.18 | 0.16 | 0.18 | 0.10 | 0.22 | 0.16 | 0.24 |
| False negative rate | 0.31 | 0.11 | 0.12 | 0.13 | 0.23 | 0.04 | 0.19 |
| Sensitivity | 0.69 | 0.89 | 0.88 | 0.87 | 0.77 | 0.96 | 0.81 |
| Specificity | 0.82 | 0.84 | 0.82 | 0.90 | 0.78 | 0.84 | 0.76 |
| Positive predictive power | 0.74 | 0.59 | 0.64 | 0.78 | 0.64 | 0.82 | 0.55 |
| Negative predictive power | 0.78 | 0.97 | 0.95 | 0.94 | 0.87 | 0.96 | 0.92 |
| Overall accuracy rate | 0.77 | 0.85 | 0.83 | 0.89 | 0.78 | 0.89 | 0.77 |
| Area under the curve | 0.83 | 0.88 | 0.88 | 0.94 | 0.88 | 0.92 | 0.82 |
| Base rate | 0.42 | 0.21 | 0.27 | 0.29 | 0.33 | 0.44 | 0.27 |
| aimswebPlus Fall cut score | 157 | 191 | 197 | 211 | 221 | 222 | 228 |
| Criterion Spring cut score | 165 | 214 | 224 | 235 | 247 | 257 | 267 |
| At 90% sensitivity, specificity equals | 0.62 | 0.84 | 0.75 | 0.79 | 0.64 | 0.87 | 0.48 |
| At 80% sensitivity, specificity equals | 0.73 | 0.84 | 0.82 | 0.91 | 0.65 | 0.88 | 0.76 |
| At 70% sensitivity, specificity equals | 0.82 | 0.84 | 0.84 | 0.94 | 0.88 | 0.89 | 0.84 |

**Table 60** Classification Accuracy of Math Composite Scores and NWEA–MAP (Grades 2–6)

| | Grade | | | | |
|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 |
| False positive rate | 0.19 | 0.15 | 0.15 | 0.13 | 0.15 |
| False negative rate | 0.19 | 0.05 | 0.18 | 0.06 | 0.23 |
| Sensitivity | 0.81 | 0.95 | 0.82 | 0.94 | 0.77 |
| Specificity | 0.81 | 0.85 | 0.85 | 0.87 | 0.85 |
| Positive predictive power | 0.56 | 0.70 | 0.77 | 0.77 | 0.66 |
| Negative predictive power | 0.93 | 0.98 | 0.89 | 0.97 | 0.90 |
| Overall accuracy rate | 0.81 | 0.88 | 0.84 | 0.89 | 0.83 |
| Area under the curve | 0.86 | 0.93 | 0.87 | 0.95 | 0.87 |
| Base rate | 0.23 | 0.27 | 0.38 | 0.31 | 0.28 |
| aimswebPlus Fall cut score | 156 | 196 | 202 | 214 | 212 |
| Criterion Spring cut score | 191 | 203 | 208 | 216 | 222 |
| At 90% sensitivity, specificity equals | 0.88 | 0.94 | 0.89 | 0.94 | 0.88 |
| At 80% sensitivity, specificity equals | 0.83 | 0.93 | 0.85 | 0.94 | 0.66 |
| At 70% sensitivity, specificity equals | 0.67 | 0.85 | 0.64 | 0.87 | 0.62 |

**Table 61** Classification Accuracy of Math Composite Scores and STAAR (Grades 3–8)

| | Grade | | | | | |
|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 6 | 7 | 8 |
| False positive rate | 0.26 | 0.27 | 0.37 | 0.30 | 0.17 | 0.27 |
| False negative rate | 0.10 | 0.15 | 0.13 | 0.17 | 0.06 | 0.31 |
| Sensitivity | 0.90 | 0.85 | 0.87 | 0.83 | 0.94 | 0.69 |
| Specificity | 0.74 | 0.73 | 0.63 | 0.70 | 0.83 | 0.73 |
| Positive predictive power | 0.69 | 0.64 | 0.66 | 0.73 | 0.83 | 0.59 |
| Negative predictive power | 0.92 | 0.90 | 0.85 | 0.81 | 0.94 | 0.80 |
| Overall accuracy rate | 0.80 | 0.78 | 0.74 | 0.76 | 0.88 | 0.71 |
| Area under the curve | 0.90 | 0.85 | 0.79 | 0.78 | 0.91 | 0.79 |
| Base rate | 0.39 | 0.36 | 0.45 | 0.49 | 0.46 | 0.37 |
| aimswebPlus Fall cut score | 192 | 196 | 217 | 217 | 212 | 211 |
| Criterion Spring cut score | 1460 | 1535 | 1558 | 1584 | 1615 | 1641 |
| At 90% sensitivity, specificity equals | 0.97 | 0.82 | 0.65 | 0.77 | 0.85 | 0.73 |
| At 80% sensitivity, specificity equals | 0.76 | 0.74 | 0.63 | 0.70 | 0.85 | 0.53 |
| At 70% sensitivity, specificity equals | 0.74 | 0.61 | 0.57 | 0.52 | 0.85 | 0.42 |

**Table 62** Classification Accuracy of Math Composite Scores and NMSBA (Grades 6–8)

| | Grade | | |
|---|---|---|---|
| | 6 | 7 | 8 |
| False positive rate | 0.22 | 0.27 | 0.25 |
| False negative rate | 0.12 | 0.12 | 0.20 |
| Sensitivity | 0.88 | 0.88 | 0.80 |
| Specificity | 0.78 | 0.73 | 0.75 |
| Positive predictive power | 0.42 | 0.44 | 0.39 |
| Negative predictive power | 0.97 | 0.96 | 0.95 |
| Overall accuracy rate | 0.79 | 0.76 | 0.76 |
| Area under the curve | 0.89 | 0.85 | 0.82 |
| Base rate | 0.15 | 0.19 | 0.17 |
| aimswebPlus Fall cut score | 198 | 204 | 200 |
| Criterion Spring cut score | 630 | 730 | 827 |
| At 90% sensitivity, specificity equals | 0.84 | 0.77 | 0.79 |
| At 80% sensitivity, specificity equals | 0.83 | 0.73 | 0.75 |
| At 70% sensitivity, specificity equals | 0.78 | 0.73 | 0.64 |

## Reading Classification Accuracy

This section describes the classification accuracy of the aimswebPlus Fall Reading composite score for Grades 2 through 8. To extend the generalizability of results, classification accuracy was evaluated with the following four different criterion measures: ISAT, MAP–GLA, NWEA–MAP, and STAAR. Note that these measures are described in the Validity section of this manual.

Tables 63 through 66 show the classification accuracy results, by grade level. For the three state accountability criterion assessments (ISAT, STAAR, and NMSBA), spring benchmark performance levels were based on the cut score at or above which a student was designated as proficient in that state assessment system during the 2013–2014 school year, by grade level. Base rates, which range from the low-0.20s to mid-0.60s, indicate the percentage of students who were not proficient on the state test.

The NWEA–MAP reading cut scores were based on results provided in the NWEA linking study reports. NWEA conducts linking studies using data from students with MAP scores and state test scores. The linking study aligns NWEA's RIT scale to the state test scale using equipercentile equating. For each state proficiency level, a RIT cut score is defined.

AUC values range from the upper-0.70s to the mid-0.90s, with 14 of the 20 reported AUCs exceeding 0.85 and 19 exceeding 0.80, the threshold for good classification accuracy.

**Table 63** Classification Accuracy of Reading Composites Scores and ISAT (Grades 3–8)

| | Grade | | | | | |
|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 6 | 7 | 8 |
| False positive rate | 0.16 | 0.17 | 0.25 | 0.10 | 0.07 | 0.15 |
| False negative rate | 0.05 | 0.12 | 0.14 | 0.22 | 0.21 | 0.14 |
| Sensitivity | 0.95 | 0.88 | 0.86 | 0.78 | 0.79 | 0.86 |
| Specificity | 0.84 | 0.83 | 0.75 | 0.90 | 0.93 | 0.85 |
| Positive predictive power | 0.60 | 0.70 | 0.55 | 0.75 | 0.82 | 0.72 |
| Negative predictive power | 0.99 | 0.94 | 0.94 | 0.91 | 0.91 | 0.93 |
| Overall accuracy rate | 0.86 | 0.85 | 0.78 | 0.86 | 0.88 | 0.86 |
| Area under the curve | 0.96 | 0.90 | 0.87 | 0.90 | 0.92 | 0.93 |
| Base rate | 0.20 | 0.31 | 0.26 | 0.28 | 0.31 | 0.31 |
| aimswebPlus Fall cut score | 396 | 430 | 470 | 473 | 499 | 523 |
| Criterion Spring cut score | 207 | 217 | 228 | 237 | 239 | 248 |
| At 90% sensitivity, specificity equals | 0.84 | 0.84 | 0.76 | 0.73 | 0.65 | 0.82 |
| At 80% sensitivity, specificity equals | 0.84 | 0.95 | 0.86 | 0.86 | 0.83 | 0.91 |
| At 70% sensitivity, specificity equals | 0.84 | 0.99 | 0.87 | 0.93 | 0.97 | 0.93 |

**Table 64** Classification Accuracy of Reading Composites Scores and NWEA–MAP (Grades 2–6)

| | Grade | | | | |
|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 |
| False positive rate | 0.20 | 0.13 | 0.27 | 0.17 | 0.20 |
| False negative rate | 0.12 | 0.15 | 0.14 | 0.12 | 0.26 |
| Sensitivity | 0.88 | 0.85 | 0.86 | 0.88 | 0.74 |
| Specificity | 0.80 | 0.87 | 0.73 | 0.83 | 0.80 |
| Positive predictive power | 0.64 | 0.70 | 0.53 | 0.73 | 0.52 |
| Negative predictive power | 0.94 | 0.94 | 0.94 | 0.93 | 0.91 |
| Overall accuracy rate | 0.82 | 0.86 | 0.76 | 0.85 | 0.79 |
| Area under the curve | 0.88 | 0.92 | 0.86 | 0.93 | 0.82 |
| Base rate | 0.29 | 0.26 | 0.26 | 0.34 | 0.23 |
| aimswebPlus Fall cut score | 342 | 378 | 432 | 463 | 483 |
| Criterion Spring cut score | 189 | 199 | 205 | 211 | 215 |
| At 90% sensitivity, specificity equals | 0.75 | 0.80 | 0.67 | 0.80 | 0.58 |
| At 80% sensitivity, specificity equals | 0.83 | 0.87 | 0.74 | 0.88 | 0.74 |
| At 70% sensitivity, specificity equals | 0.84 | 0.93 | 0.83 | 0.96 | 0.81 |

**Table 65** Classification Accuracy of Reading Composites Scores and MAP–GLA (Grades 3–8)

| | Grade | | | | | |
|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 6 | 7 | 8 |
| False positive rate | 0.11 | 0.28 | 0.23 | 0.10 | 0.28 | 0.27 |
| False negative rate | 0.19 | 0.23 | 0.19 | 0.24 | 0.29 | 0.24 |
| Sensitivity | 0.81 | 0.77 | 0.81 | 0.76 | 0.71 | 0.76 |
| Specificity | 0.89 | 0.72 | 0.77 | 0.90 | 0.72 | 0.73 |
| Positive predictive power | 0.93 | 0.80 | 0.85 | 0.90 | 0.80 | 0.69 |
| Negative predictive power | 0.73 | 0.68 | 0.71 | 0.74 | 0.61 | 0.80 |
| Overall accuracy rate | 0.84 | 0.75 | 0.79 | 0.82 | 0.71 | 0.75 |
| Area under the curve | 0.89 | 0.82 | 0.85 | 0.91 | 0.80 | 0.84 |
| Base rate | 0.63 | 0.59 | 0.62 | 0.56 | 0.61 | 0.44 |
| aimswebPlus Fall cut score | 393 | 439 | 461 | 471 | 474 | 515 |
| Criterion Spring cut score | 648 | 662 | 675 | 676 | 680 | 696 |
| At 90% sensitivity, specificity equals | 0.93 | 0.75 | 0.79 | 0.90 | 0.72 | 0.83 |
| At 80% sensitivity, specificity equals | 0.89 | 0.68 | 0.78 | 0.74 | 0.59 | 0.73 |
| At 70% sensitivity, specificity equals | 0.65 | 0.45 | 0.70 | 0.67 | 0.51 | 0.51 |

**Table 66** Classification Accuracy of Reading Composites Scores and STAAR (Grades 3–5)

| | Grade | | |
|---|---|---|---|
| | 3 | 4 | 5 |
| False positive rate | 0.27 | 0.31 | 0.24 |
| False negative rate | 0.10 | 0.17 | 0.21 |
| Sensitivity | 0.90 | 0.83 | 0.79 |
| Specificity | 0.73 | 0.69 | 0.76 |
| Positive predictive power | 0.55 | 0.60 | 0.70 |
| Negative predictive power | 0.95 | 0.87 | 0.84 |
| Overall accuracy rate | 0.78 | 0.74 | 0.77 |
| Area under the curve | 0.85 | 0.79 | 0.82 |
| Base rate | 0.27 | 0.36 | 0.41 |
| aimswebPlus Fall cut score | 376 | 419 | 438 |
| Criterion Spring cut score | 1400 | 1486 | 1520 |
| At 90% sensitivity, specificity equals | 0.82 | 0.72 | 0.77 |
| At 80% sensitivity, specificity equals | 0.79 | 0.69 | 0.74 |
| At 70% sensitivity, specificity equals | 0.73 | 0.52 | 0.58 |

# References

Clemens, N.H., Hagan-Burke, S., Luo, W., Cerda, C., Blakely, A., Frosch, J., … Jones, M. (2015). The predictive validity of kindergarten and first-grade reading skills. *School Psychology Review, 44*(1), 76–97.

Feldt, L.S., & Brennan, R.L. (1989). Reliability. In R.L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 105–146). New York: Macmillan.

Hoover, H. D., Dunbar, S. B., & Frisbie, D. A. (2007). Iowa Tests of Basic Skills. Itasca, IL: Houghton Mifflin Harcourt.

Northwest Evaluation Association. *Measures of Academic Progress.* Portland, OR: Author.

U.S. Census Bureau (2013). *American community survey, 2013: 1-year period estimates.* Washington, DC: Author.

# Appendix

## aimswebPlus Measures

### *Early Numeracy (Kindergarten and Grade 1)*

aimswebPlus Early Numeracy comprises the individually administered math measures developed for students in Kindergarten and Grade 1. Note that these measures are also available in Spanish. Table A1 presents the grades and seasons available, tasks, and scoring criteria for these measures, followed by brief descriptions of each measure.

Table A1 Early Numeracy Measure Descriptions

| Measure | Grade | Season | What students do | Score |
|---|---|---|---|---|
| Number Naming Fluency (NNF) | K | F, W, S | Verbally name numbers up to 20 for 1 minute. | Number of items correctly answered |
| Quantity Total Fluency (QTF) | K | F, W, S | Boxes containing blue dots are presented. Students state the total number of dots within each box or each pair of boxes for 1 minute. | Number of items correctly answered |
| Quantity Difference Fluency (QDF) | K | W, S | Pairs of boxes containing dots (one with blue dots, one with red dots) are presented. Students state how many more blue dots are needed to match the number of red dots for 1 minute. | Number of items correctly answered |
| Concepts & Applications (CA) | K, 1 | F, W, S | Mentally solve various types of math problems and state the correct answers. | Number of items correctly answered |
| Number Comparison Fluency–Pairs (NCF–P) | 1 | F, W, S | Pairs of numbers are presented. Students identify which of two numbers is larger for each pair for 1 minute. | Number of items correctly answered |
| Math Facts Fluency–1 Digit (MFF–1D) | 1 | F, W, S | Mentally solve simple addition and subtraction problems involving numbers 0 through 10 and state the correct answers for 1 minute. | Number of items correctly answered |
| Math Facts Fluency–Tens (MFF–T) | 1 | W, S | Mentally add or subtract 10 to/from given numbers and state the correct answers for 1 minute. | Number of items correctly answered |

### Number Naming Fluency (NNF)

- Grade: Kindergarten
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), timed
- Test Content: The student points to and names visually presented numbers for 1 minute. Each form contains 80 items.
- Scoring: 1 point for each correctly named number
- Time Limit: 1 minute

### Quantity Total Fluency (QTF)

- Grade: Kindergarten
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), timed
- Test Content: The student states the total number of dots in each box or pair of boxes for 1 minute. Each form contains 38 items.
- Scoring: 1 point for each correctly answered item
- Time Limit: 1 minute

### Quantity Difference Fluency (QDF)

- Grade: Kindergarten
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), timed
- Test Content: The states how many more blue dots are needed to match the number of red dots for each box pair for 1 minute. Each form contains 24 items.
- Scoring: 1 point for each correctly answered item
- Time Limit: 1 minute

### Concepts & Applications (CA)

- Grades: Kindergarten and Grade 1
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), untimed
- Test Content: The student solves one- and two-step math word problems, each addressing an aspect of grade-appropriate CCSS domains. The examiner reads each item to the student and the student states the correct answer, using the corresponding visual stimulus to solve the problem. The student attempts all 25 items in a given form.
- Scoring: 1 point for each correctly answered item
- Administration time: 7–12 minutes (approximate)

### Number Comparison Fluency–Pairs (NCF–P)

- Grade: 1
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), timed
- Test Content: The student points to and names the larger number in each pair for 1 minute. Each form contains 50 items.
- Scoring: 1 point for each correctly answered item
- Time Limit: 1 minute

### Math Facts Fluency–1 Digit (MFF–1D)

- Grade: 1
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), timed
- Test Content: The student solves addition and subtraction problems involving numbers 0 through 10 for 1 minute. Each form contains 40 items.
- Scoring: 1 point for each correctly answered item
- Time Limit: 1 minute

### Math Facts Fluency–Tens (MFF–T)

- Grade: 1
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), timed
- Test Content: The student solves problems involving the addition and subtraction of 10 for 1 minute. Each form contains 32 items.
- Scoring: 1 point for each correctly answered item
- Time Limit: 1 minute

## *Early Literacy (Kindergarten and Grade 1)*

aimswebPlus Early Literacy comprises the individually administered reading measures developed for students in Kindergarten and Grade 1. Table A2 presents the grades and seasons available, tasks, and scoring criteria for these measures, followed by brief descriptions of each measure.

Table A2 Early Literacy Measure Descriptions

| Measure | Grade | Season | What students do | Score |
|---|---|---|---|---|
| Print Concepts (PC) | K | F | Show understanding of purpose, use, and contents (letters, pictures) of a book. | Number of questions answered correctly |
| Letter Naming Fluency (LNF) | K | F, W, S | Say the names of visually presented letters for 1 minute. | Number of letters named correctly |
| Initial Sounds (IS) | K | F, W | Look at four pictures and either point to the one that begins with a given letter sound or make the sound that begins the word. | Number of correct letter sounds and picture names |
| Auditory Vocabulary (AV) | K, 1 | F, W, S | Point to the one of four pictures that matches an orally presented word. | Number of pictures chosen correctly |
| Letter Word Sounds Fluency (LWSF) | K<br>1 | W, S<br>F | Say the sounds of visually presented letters, syllables, and words for 1 minute. | Number of sounds or words said correctly |
| Phoneme Segmentation (PS) | K<br>1 | W, S<br>F | Say the phonemes in orally presented words. | Number of phonemes said correctly |
| Word Reading Fluency (WRF) | K<br>1 | S<br>F, W, S | Read a word list aloud for 1 minute. | Number of words read correctly |
| Oral Reading Fluency* (ORF) | 1 | F, W, S | Read two stories aloud, each for 1 minute. | Average number of words read correctly |

**Note.* The ORF information in this table applies to the screening seasons of Fall, Winter, and Spring. When using ORF to progress monitor, students read ***one*** story aloud for 1 minute per testing session and the reported score is the number of words read correctly for that single story.

### Print Concepts (PC)

- Grade: Kindergarten
- Test Format: Individual, student storybook and examiner digital record form (online), untimed
- Test Content: The student shows understanding of the purpose, use, and contents (letters, pictures) of a book (specific criteria for selecting appropriate books are provided in the *aimswebPlus Early Literacy Administration and Scoring Guide*). The student attempts all 9 items.
- Scoring: 1 point for each correctly answered item
- Administration time: 2–3 minutes (approximate)

### Letter Naming Fluency (LNF)

- Grade: Kindergarten
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), timed
- Test Content: The student says the names of visually presented letters for 1 minute. Each form contains 100 letters (mix of upper- and lower-case) presented in a student-friendly font.
- Scoring: 1 point for each correctly named letter
- Time limit: 1 minute

### Initial Sounds (IS)

- Grade: Kindergarten
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), untimed
- Test Content: The student looks at four pictures and either points to the one that begins with a given letter sound or makes the sound that begins the word. The student attempts all 12 items in a given form.
- Scoring: 1 point for each correctly answered item
- Administration time: 2–3 minutes (approximate)

### Auditory Vocabulary (AV)

- Grades: Kindergarten and Grade 1
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), untimed; all items have four response options
- Test Content: The student looks at four pictures and points to the picture that matches an orally presented word. The student attempts all 25 items in a given form.
- Scoring: 1 point for each correctly answered item
- Administration time: 2–4 minutes (approximate)

### Letter Word Sounds Fluency (LWSF)

- Grades: Kindergarten and Grade 1
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), timed
- Test Content: The student says the sounds of visually presented letters, syllables, and words for 1 minute. Each form contains 45 letters and 10 three-letter words.
- Scoring: 1 point for each letter or word sound correctly made
- Time limit: 1 minute

### Phoneme Segmentation (PS)

- Grades: Kindergarten and Grade 1
- Test Format: Individual, examiner digital record form (online), untimed
- Test Content: The student says the phonemes of orally presented words that are made up of up to four phonemes. The student attempts all 15 items in a given form.
- Scoring: 1 point for each phoneme correctly made
- Administration time: 2–3 minutes (approximate)

### Word Reading Fluency (WRF)

- Grades: Kindergarten and Grade 1
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), timed
- Test Content: The student reads read words aloud for 1 minute. Each form contains two pages of word lists, totaling 99 words.
- Scoring: 1 point for each word correctly read
- Time limit: 1 minute

### Oral Reading Fluency (ORF)

- Grade: 1
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), timed
  Test Content: The student reads one or two stories aloud, each for 1 minute. Each screening form contains two stories, while each progress monitoring forms contains one story.
- Scoring: Mean number of words read correctly in the two stories (screening) or words read correctly in one story (progress monitoring)
- Time Limit: 1 minute per story

## *Math (Grades 2–8)*

aimswebPlus Math comprises the measures developed for students in Grades 2 through 8. Note that these measures are also available in Spanish.Table A3 presents the grades and seasons available, tasks, and scoring criteria for these measures, followed by brief descriptions of each measure.

**Table A3** Math Measure Descriptions

| Measure | Grade | Season | What students do | Score |
|---|---|---|---|---|
| Number Comparison Fluency–Triads (NCF–T) | 2–8 | F, W, S | Compare three numbers within and across number systems to determine the relative distance between each number for 3 minutes. | Number of items correctly answered, corrected for guessing* |
| Mental Computation Fluency (MCF) | 2–8 | F, W, S | Solve multiple-choice math computation problems for 4 minutes. | Number of items correctly answered, corrected for guessing* |
| Concepts & Applications (CA) | 2–8 | F, W, S | Solve multiple-choice math word problems. | Number of items correctly answered |

**Note*. NCF–T and MCF employ a correction for guessing when calculating the total score. Items not attempted (skipped) and items not reached are ignored in the calculation of the corrected total score. Together, NCF–T and MCF combine into the Number Sense Fluency (NSF) score, which is the simple sum of the NCF–T and MCF corrected scores. This NSF score is the basis for progress monitoring decisions.

### Number Comparison Fluency–Triads (NCF–T)

- Grades: 2–8
- Test Format: Group, online, timed
- Test Content: The student answers multiple-choice math items, comparing numbers within and across number systems, for 3 minutes. Each item is presented as a triad of numbers, with the student determining whether the top number in the triad is closer in value to the bottom left number, the bottom right number, or exactly between the two numbers. Each form contains 40 items.
- Scoring: 1 point for each correctly answered item, total score then adjusted for guessing
- Time Limit: 3 minutes

### Mental Computation Fluency (MCF)

- Grades: 2–8
- Test Format: Group, online, timed
- Test Content: The student answers multiple-choice math items, each requiring one- or two-step mental computation of a math expression, for 4 minutes. The use of friendly (e.g., round) numbers facilitates the mental computation of answers. Each form contains 42 items.
- Scoring: 1 point for each correctly answered item, total score then adjusted for guessing
- Time Limit: 4 minutes

### Concepts & Applications (CA)

- Grades: 2–8
- Test Format: Group, online, untimed; audio is available for all students at all grade levels
- Test Content: The student answers multiple-choice math word problems, each addressing an aspect of grade-appropriate CCSS domains. Each form contains between 29 and 31 items, depending on grade and season. The student attempts all items in a given form.
- Scoring: 1 point for each correctly answered item, total score then converted to a developmental scale score
- Administration time: 15–25 minutes (approximate)

### *Reading (Grades 2–8)*

aimswebPlus Reading comprises the measures developed for students in Grades 2 through 8. Table A4 presents the grades and seasons available, tasks, and scoring criteria for these measures, followed by brief descriptions of each measure.

**Table A4** Reading Measure Descriptions

| Measure | Grade | Season | What students do | Score |
|---|---|---|---|---|
| Vocabulary (VO) | 2–8 | F, W, S | Identify the meanings of target words by selecting from multiple-choice options. | Number of items correctly answered |
| Reading Comprehension (RC) | 2–8 | F, W, S | Read six passages of text and answer multiple-choice questions about each passage. | Number of items correctly answered |
| Silent Reading Fluency (SRF) | 4–8 | F, W, S | Read three stories divided into brief sections and answer multiple-choice questions about each story. | Median reading rate of three stories |
| Oral Reading Fluency* (ORF) | 2–8 | F, W, S | Read two stories aloud, each for 1 minute. | Average number of words read correctly |

**Note*. The ORF information in this table applies to the screening seasons of Fall, Winter, and Spring. When using ORF to progress monitor, students read *one* story aloud for 1 minute per testing session and the reported score is the number of words read correctly for that single story.

## Vocabulary

- Grades: 2–8
- Test Format: Group, online, untimed; audio is available for all students at all grade levels
- Test Content: The student answers multiple-choice vocabulary items, choosing the response that best matches the meaning of a target word. Each form contains 16 (Grade 2) or 22 items (Grades 3–8), presented one per screen. The student attempts all items in a given form.
- Scoring: 1 point for each correctly answered item
- Administration time: 4–7 minutes (approximate)

## Reading Comprehension (RC)

- Grades: 2–8
- Test Format: Group, online, untimed
- Test Content: The student reads passages (three literary and three informational) and answers multiple-choice questions about each passage to demonstrate comprehension of the text. The student attempts all 24 items in a given form.
- Scoring: 1 point for each correctly answered item
- Administration time: 15–25 minutes (approximate)

### Silent Reading Fluency (SRF)

- Grades: 4–8
- Test Format: Group, online, untimed
- Test Content: The student reads story segments and answers multiple-choice questions about each segment, receiving immediate correct/incorrect feedback after each question before moving on to the next segment and question. The time spent reading each passage is captured to compute the student's reading rate for each story. Each form contains three stories broken into four segment/question pairs, resulting in 12 questions per form. The student attempts all items in a given form.
- Scoring: Median reading rate of three stories, if sufficient comprehension demonstrated (i.e., at least three of four questions correctly answered on at least two stories)
- Administration time: 4–6 minutes (approximate)

### Oral Reading Fluency (ORF)

- Grades: 2–8
- Test Format: Individual, student stimulus book (print) and examiner digital record form (online), timed
- Test Content: The student reads one or two stories aloud, each for 1 minute. Each screening form contains two stories, while each progress monitoring forms contains one story.
- Scoring: Mean number of words read correctly in the two stories (screening) or words read correctly in one story (progress monitoring)
- Time limit: 1 minute per story

**Pearson Executive Office 5601 Green Valley Drive Bloomington, MN 55437**

**800.627.7271 www.aimsweb.com**



Produced in the United States of America.

3 4 5 6 7 8 9 10 11 12 13 14 15 16