Assessment Instrument Description: aimsweb®Plus

Element	Element Description	
Instrument Name	Name of specific instrument (more than vendor name).	aimsweb®Plus
Vendor	Name of the company or organization that produces the instrument.	NCS Pearson, Inc.
Purpose (Intended Use)	The described purpose and appropriate uses of the instrument.	aimswebPlus™ is an online assessment, data management, and reporting system that provides national and local performance and growth norms for the screening and progress monitoring of math and reading skills for all students in Kindergarten through Grade 8 (Users can roster Pre-K and High school (9-12) and test off level as well). aimswebPlus uses two types of measures: *curriculum-based measures* (CBMs)—brief, timed measures of fluency on essential basic skills—and *standards-based assessments* (SBAs), which are comprehensive measures aligned to current learning standards. By combining these two types of measures, aimswebPlus provides the data that schools need for program planning and evaluation and for tiered assessment (multi-tiered system of supports [MTSS], also known as response to intervention [RTI]). Furthermore, aimswebPlus data provides teachers with the information needed to differentiate instruction and determine who will benefit from intensive intervention. aimswebPlus also provides a Lexile® equivalency for reading and a Quantile® equivalency for math. Reports can be generated at the individual, classroom, school, and district levels in the aimswebPlus online system. aimswebPlus is used for benchmarking, universal screening, diagnosing strengths and weaknesses in Reading and Math, and for progress monitoring.
Types of Instruments	Interim, Summative, Diagnostic	Interim. Universal Screening/Benchmarking and Progress Monitoring

Population	Who (which students) could be assessed using the instrument.	aimswebPlus assesses students in Kindergarten through Grade 8. Pre-K and High school (9-12) can be rostered and tested off-grade level. Tier 1 - Assess all students three times per year for universal screening (early identification), general education progress monitoring, and AYP accountability. Tier 2 - Assess and monitor at-risk students and the effectiveness of instructional changes. Tier 3 - Write individualized annual goals and monitor progress more frequently for those who need intensive instructional services.
When? How frequently?	How frequently the instrument can be administered in a school year, and recommended or required administration windows.	Benchmarking is designed to inform instruction to improve achievement. Benchmarks are established three times per year for all students, based on established school and district windows. The screening periods for each season and recommended testing windows include: • Fall: July 15 through November 30 (September 1–October 15) • Winter: December 1 through March 15 (January 1– January 30) • Spring: March 16 through July 14 (May 1–May 31) Although administration is permitted at any time during a given season, administrations within the recommended testing window maximizes the accuracy of the national norms. Strategic Monitor provides schools with the option to increase assessment frequency for students who have been identified as "struggling," or minimally at-risk in the Benchmark process – or for all students, if desired. Increasing assessment frequency provides more opportunity to evaluate the effectiveness of instructional changes and to verify struggling student achievement levels or to confirm there is no degeneration of progress in minimally at-risk students. Progress Monitor is a continuous assessment and improvement system designed specifically for frequent assessment and monitoring of at-risk students, including those receiving Title I services, or those identified with a learning disability or other special needs. Progress monitoring assessments can be administered as frequently as weekly.
Content Area (s)	Content area or areas being assessed.	Early Literacy and Reading
Learning Objectives	Specific learning objectives being assessed, at as detailed a level as is provided. This	**Early Literacy (Grades K-1)** Letter Naming Fluency — Student says the names of visually presented letters; number of corrects are scored. Letter Word Sound Fluency — Student says the sounds of visually presented letter, syllables, and words; number of corrects are scored.

may be "topics" or categories or may be actual learning objective statements.	Phoneme Segmentation	Student hears orally-presented words and says their component phonemes. The score is the number of segments said correctly in one minute.
	Word Reading Fluency	Student says a word list aloud for one minute; number of corrects are scored.
	Print Concepts	Student shows understanding of purpose, use, and contents (letters, pictures) of a book.
	Initial Sounds	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.
	Nonsense Word Fluency (NWF)	The Nonsense Word measure requires the student to say each sound (or the whole words) in nonreal words.
	Written Expression (WE)	Student manually writes a story from a Story Starter sentence or prompt for 3 minutes.
	Auditory Vocabulary	Student points to one of four pictures that match an orally presented word.
	Oral Reading Fluency	Student reads two stores aloud, each for one minute.
	Reading (Grades 2-8)	
	Written Expression (WE)	Student manually writes a story from a Story Starter sentence or prompt for 3 minutes.
	Vocabulary	Student identifies the meanings of target words by selecting from multiple-choice options; number of corrects are scored.
	Reading Comprehension	Student reads six passages of text and answers multiple-choice questions about each passage.
	Silent Reading Fluency	Student reads three stories divided into brief sections and answers multiple-choice questions about each story.
	Oral Reading Fluency	Students read two stories aloud, each for one minute.
A		

<table>
<tr>
<td>Individual Metrics</td>
<td>The scores provided at the individual (student) level.</td>
<td>Student results are provided in a score for each measure by screening period (F, W, S). Information provided at the individual student level for the measures includes

Raw scores
National Percentile
Lexile (ORF only)
Composite percentile for early literacy, early numeracy, Reading and Math
Performance level
Risk Status
Rate of Improvement

Student Rate of Improvement
National Rate of Improvement
Student Growth percentile

For progress monitoring on individual measures the following are reported:

Raw score
Errors
Goal rate of improvement
Trend rate of improvement
Aimline (a line connecting the baseline score to the goal score)
Trendline
Student’s likelihood of meeting the performance goal by the goal date

For more detailed information, please refer to the aimswebPlus Development manual and Introductory Guide (appendix?)
Vocabulary and Reading Comprehension report a raw number correct score, which is converted to a vertical scale called the Growth Scale Value. The vertical scale spans the full performance continuum for Grades 2-8. Reading Composite Scores combine scores for Vocabulary, Reading Comprehension, and Oral Reading Fluency or Silent Reading Fluency.
Here is a table showing the Grade, season, and measures included in the Composite score.
<table>
<tr><th>Grade</th><th>Season</th><th>Composite Score Measures</th></tr>
<tr><td>2-3</td><td>Fall</td><td>Vocabulary, Reading Comprehension, Oral Reading Fluency</td></tr>
</table>
</td>
</tr>
</table>

2-3	Winter	Vocabulary, Reading Comprehension, Oral Reading Fluency
2-3	Spring	Vocabulary, Reading Comprehension, Oral Reading Fluency
4-8	Fall	Vocabulary, Reading Comprehension, Silent Reading Fluency
4-8	Winter	Vocabulary, Reading Comprehension, Silent Reading Fluency
4-8	Spring	Vocabulary, Reading Comprehension, Silent Reading Fluency

Individual Comparison Points (cut scores)	Information provided regarding how good is good enough performance on the instrument. Comparison information should be available for every individual metric. This may be performance level ratings with specific cut scores.	

<table>
<tr>
<td>Aggregate Metrics</td>
<td>Scores provided at the group level. The group could be a grade level, school, district, or disaggregated groups (e.g. race/ethnicity, gender, IEP status, FRL status) Specify the group(s) and the score(s) provided.</td>
<td>The table below lists the reporting screens available in the aimswebPlus Student Module, the reporting levels, and a brief description of the type of data displayed. The results summary also indicates which data are reported across seasons and years. Note that some screens include a blend of single season results and longitudinal results. Most screens are filterable and sortable. The only exceptions are the Student Snapshot screen, the Skills Plan screen, and the Scores and Skills Plan screen. Also, using filters provided in the aimswebPlus system, a teacher or administrator can generate reports with disaggregated data, as well. Data fields include the following:

Student ID
Student First, Middle, and Last Name
Student Grade
Student DOB
Student Gender
Student ServiceCode (G = General Ed, T = Title 1, S = Special Ed)
Student MealStatus
Student RaceEthnicity
State Testing Identifier (STI)
ESL
IEP
Disability Codes

aimswebPlus reporting screens available in the Students Module
<table>
<tr><th>Screen Name & Report Name</th><th>Reporting Levels</th><th>Results</th></tr>
<tr><td>Benchmark Comparison</td><td>Class (or roster), Grade, School, District</td><td>All screening/benchmark scores by subject within a grade and season</td></tr>
<tr><td>Student Profile</td><td>Individual</td><td>Fall, Winter, and Spring scores on all measures completed by the student within a school year</td></tr>
</table>
</td>
</tr>
</table>

Individual Benchmark	Individual Student	Fall, Winter, and Spring scores on any single measure completed by the student within a school year
Monitor	Roster/class, Grade, School, District	The most recent PM scores for all students and all PM measures
Scores Snapshot	Individual Student	Brief static interpretive report and graph of all measure scores by subject (e.g., math) in a single season; And a graph of Fall, Winter, and Spring composite scores. Available in Spanish
Skills Plan	Individual Student	Static diagnostic report providing a graph of performance by Math CA domain, a score summary table, and item scores for a single season. For Reading comprehension, the report provides a performance and score summary, a profile of the students strengths and weaknesses, and an interpretive table that examines several factors that may contribute to poor comprehension including vocabulary knowledge, oral and silent reading rate, and the ability to derive meaning from very brief passages.

Aggregate Comparison Points (cut scores) Vendor	Information provided regarding how good is good enough	aimswebPlus recommends using the 15th and 45th national percentiles (defaults) as follows: • Not On Track: ≤15th national percentile • Further assessment may be needed: 16th–45th national percentile • On Track: >45th national percentile

	performance at the group level.	Schools and districts may set their own cut scores for local purposes.
Comparison Points (CDE)	CDE cut scores for requests to reconsider.	As students complete aimswebPlus assessments, their results are automatically scored and available immediately in the system. Results are reported as total scores by measure and composite using either national or local percentiles; these results can be interpreted using both *norm-referenced* and *criterion-referenced* methods. **Norm-referencing and Percentiles** A *norm-referenced interpretation* involves comparing a student's score with the scores from a local or national reference group of students in the same grade who were tested on the same content during the same timeframe of the school year. aimswebPlus provides norm-referenced information in the form of *percentiles*, which represent the percentage of students in the nationally representative sample who scored at or below a given score. For example, a score at the 35th percentile means that 35% of the norm sample achieved a score equal to or lower than this level. Ranging from 1 to 99, percentiles provide a common reference point for interpreting student performance and for comparing groups. aimswebPlus identifies the following performance levels using this scale: • Well-Below Average: 1st–10th percentiles • Below Average: 11th–25th percentiles • Average: 26th–74th percentiles • Above Average: 75th–89th percentiles • Well-Above Average: 90th–99th percentiles These performance levels enable an at-a-glance evaluation of the instructional needs of students, classrooms, and schools. Note that both national and local (i.e., at the school and/or district level) percentile norms are available in the aimswebPlus system. **Criterion-referencing and Performance Targets: Who Is or Is Not On Track** A *criterion-referenced interpretation* involves comparing a student's score with performance targets/benchmarks that designate proficiency or academic success. The performance target may be based on expert judgment, historical data, or percentiles and typically references end-of-grade expectations. Because universal screening

occurs in the Fall and Winter (in addition to the Spring), it is important to have targets in each of those seasons to help indicate who is or is not on track to meet the end-of-year target.

What is considered proficient varies across grades and states. For example, the percentage of students achieving the proficient level on state tests has historically ranged from about 30% to as high as 85%. Benchmarks have been defined for oral reading rates that indicate an independent or instructional level by grade. While not explicitly tied to norms, the reading rates that define these levels take normative results into consideration. Typically, independent levels approximate the 50th percentile.

Schools need to consider the available resources and current performance levels when defining performance targets for their students. Setting a target too low can lead to under-identifying students needing additional support. Conversely, setting a target too high will result in over-identifying students as at risk, which may overwhelm the resource capacity of the school.

The aimswebPlus system provides a range of scores for defining Spring performance targets. Users can select from 12 targets ranging from the 15th to the 70th national percentiles, provided in increments of five percentiles. National percentiles can be used to approximate proficiency on state tests by aligning the Spring percentile with the percentage of students below proficient because percentiles represent the percentage of students at or below a give score. For example, if 40% of students are proficient, the corresponding percentile is 60.

aimswebPlus defines seasonal *cut scores* that indicate who is unlikely to meet the Spring target. In each season, two cut scores are defined that represent the break points between the tiers. The lower cut score is associated with a high probability of failing to meet the Spring target, while the other is associated with a moderate probability of failing to meet the Spring target.

Students with scores below the lower cut score are considered at *high risk* because they have a low probability of meeting the Spring performance target without intensive intervention. Students with scores between the two cut scores are considered at *moderate risk* and will likely need additional instruction to meet the target. Students with scores above the upper cut score are considered at *low risk* and will likely remain on track with the core instructional program (i.e., without additional instruction). Once the Spring target percentile has been selected, the Fall and Winter cut scores are automatically computed by the aimswebPlus system and each student’s risk

		level is immediately updated. **aimswebPlus uses this same procedure for defining tiers.** In a tiered assessment and instruction system, tiers are used to define the level of instruction needed for the students within each tier. aimswebPlus follows the traditional model of three instructional tiers, defined as follows: • Tier 1 = Low Risk (about 75%–95% of students in this group will meet the target): Students are on track to meet the end-of-year target and are least likely to need intervention. These students should continue to receive the general instructional program. Typically, the majority of students fall into this category. • Tier 2 = Moderate Risk (about 25%–65% of students in this group will not meet the target): Students are not on track and have a moderate risk of not meeting the end-of-year target. These students require some type of intervention, often taking the form of supplemental small- group instruction. • Tier 3 = High Risk (about 50%–90% of students in this group will not meet the target): Students are not on track and are typically well-below grade level. These students have a high risk of not meeting the end-of-year target without intensive, individualized instructional intervention. Local norms can also be used to guide the selection of Spring targets. The rationale for using local norms is that instruction in a given school or district is geared to the average level of performance specific to that school or district, so students who are within the average range relative to their classmates (as indicated by local norms) should be well served by the general instructional program. A practical rationale is that the use of local norms leads to a reasonable and consistent allocation of resources across the tiers.
Alignment	Information provided by the vendor about alignment of this instrument to other instruments, standards, etc.	aimswebPlus measures are aligned to the Common Core State Standards (CCSS) and to the learning standards for many individual states, including Colorado. Additional information on the development of aimswebPlus, which included criterion validity studies with other assessments, can be found in the aimswebPlus Technical Manual (Appendix xxx).

Data Reports	Description of data reports that are provided/available at the individual and aggregate level(s).	Please see the information provided for Aggregate Metrics.
Technical Quality		**Technical Research** The aimswebPlus team regularly carries out studies to collect validity, reliability, and fairness evidence in accordance with the Joint Committee Standards (AERA, APA, NCME, 2014). This evidence has been consolidated and published in a set of technical and development manuals, which are updated with each new revision of the test. For that reason, much of the aimswebPlus research we summarize in the following section has been completed internally. We encourage you to consult the aimswebPlus Technical Manual (Pearson, 2017). Also see the latest efficacy report here: https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/efficacy-assessment-reports/aimsweb-Plus-research-report.pdf **Research studies** Each aimswebPlus measure, revised or new, was put through multiple rounds of field testing, with refinements made as needed based on the results of this testing. aimswebPlus field testing comprised the following research studies, with each study type spanning the Kindergarten through Grade 8 range: • Pilot studies: multiple studies, 1,000+ students tested • National tryout study: 14,000+ students tested • National norms study: 16,000+ students tested • Progress monitoring form equivalency studies: multiple studies, 15,000+ students tested These new normative, reliability, and validity data were collected based on a representative sample of US students. Additionally, the psychometric properties of all the aimswebPlus measures were evaluated to meet

Pearson’s and industry standards during the field testing process.

Analyses confirmed that using a multi-test battery approach provides stronger predictive data for student performance and risk status, as well as additional information about specific skills or knowledge areas that can be useful when interpreting student test scores. The combined information about automaticity of foundational skills and standards-based assessment of skills required for classroom success allow aimswebPlus to provide a more complete picture of what each student knows and can do.

Normative sample

Tables 1, 2, 3 below 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.

Table 1: Demographic characteristics of the normative samples — early numeracy

		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	I	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	I	1000	50	1000	50	265	13	506	25	201	10	1028	51	10	32	32	36

Table 2: Demographic characteristics of the normative samples — Math

			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	1580	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 3: Demographic characteristics of the normative samples — reading

			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	CRF	1158	0.49	1187	0.51	355	0.15	614	0.26	246	0.10	1130	0.48	10	32	32	36
Reading	3	CRF	1180	0.51	1125	0.49	266	0.12	583	0.25	225	0.10	1231	0.53	10	32	32	36
Reading	4	CRF	1251	0.50	1233	0.50	307	0.12	591	0.24	250	0.10	1336	0.54	10	32	32	36
Reading	5	CRF	1138	0.52	1066	0.48	337	0.15	518	0.24	251	0.11	1098	0.50	10	32	32	36
Reading	6	CRF	842	0.50	828	0.50	260	0.16	396	0.24	153	0.09	859	0.51	10	32	32	36
Reading	7	CRF	814	0.51	790	0.49	255	0.16	413	0.26	137	0.09	799	0.50	10	32	32	36
Reading	8	CRF	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

Reliability

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 used; 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. Reliability results are presented in table organized by domain: Early Numeracy, Math, Early Literacy, and Reading. Reliability coefficients are provided for each measure, season, and grade within these domains.

Table 4: Reliability of early numeracy measures and composites, kindergarten and grade

			Cronbach's Alpha		Alternate Form				
Measure	Grade	Season	*n*	Coefficient	*n* range	Coefficient mean	Range	Stratified Alpha	SEM
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	I	F, W, S	—	—	222—239	0.88	0.86—0.89	—	2.47
MFF—ID	I	F, W, S	—	—	217—234	0.86	0.86—0.89	—	2.33
MFF—T	I	F, W, S	—	—	167—175	0.93	0.93	—	1.25
CA	I	F	1459	0.86	—	—	—	—	2.12
CA	I	W	1259	0.87	—	—	—	—	1.88
CA	I	S	1364	0.88	—	—	—	—	1.75
Composite	I	F	—	—	—	—	—	0.96	3.30
Composite	I	W	—	—	—	—	—	0.97	3.29
Composite	I	S	—	—	—	—	—	0.97	3.38

Table 5: Reliability of math measures and composites, Grades 2 through 8

			Cronbach's Alpha			Alternate Form			Stratified Alpha		
Measure	Grade	Season	*n* range	Coefficient mean	Range	*n* Range	Coefficient mean	Range	Coefficient mean	Range	SEM
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

Measure	Grade	Season	Cronbach's Alpha			Alternate Form			Stratified Alpha		
			n range	Coefficient mean	Range	*n* Range	Coefficient mean	Range	Coefficient mean	Range	SEM
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 6: Reliability of early literacy measures and composites, Kindergarten and Grade 1

Measure	Grade	Season	Cronbach's Alpha *n*	Cronbach's Alpha Coefficient	Alternate Form *n* range	Alternate Form Coefficient mean	Alternate Form Range	Stratified Alpha	SEM
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	I	F, W, S	—	—	173—180	0.94	0.93—0.95	—	5.51
ORF	I	F	—	—	1341	0.97	—	—	5.21
ORF	I	W	—	—	1389	0.96	—	—	6.53
ORF	I	S	—	—	1502	0.96	—	—	7.11
PS	I	F	1329	0.83	—	—	—	—	3.84
AV	I	F	1346	0.85	—	—	—	—	0.96
AV	I	W	1390	0.87	—	—	—	—	0.88
AV	I	S	1503	0.87	—	—	—	—	0.6

Table 7: Reliability of reading measures and composites, Grades 2 through 8

			Cronbach's Alpha			Alternate Form			Stratified Alpha		
Measure	Grade	Season	*n* range	Coefficient mean	Range	*n* Range	Coefficient mean	Range	Coefficient mean	Range	SEM
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

Table 7: Reliability of reading measures and composites, Grades 2 through 8

			Cronbach's Alpha			Alternate Form			Stratified Alpha		
Measure	Grade	Season	*n* range	Coefficient mean	Range	*n* Range	Coefficient mean	Range	Coefficient mean	Range	SEM
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

In summary, reliability estimates typically met common benchmarks for adequate consistency for measures used to make decisions about individual students. There were only two cases in which a reliability estimate fell below 0.70 — the internal consistency of Print Concepts scores for students in grades K-1 was 0.63 and average internal consistency of Vocabulary scores for students in grades 2 was 0.67. In particular:

- Internal consistency of untimed early numeracy measures for students in grades K-1 ranged from 0.83 to 0.88, average alternate forms reliability for timed measures ranged from .74 to .93, and stratified alpha for composite scores ranged from 0.88 to 0.97.
- Average internal consistency of untimed math measures for students in grades 2-8 ranged from 0.77 to 0.85, average alternate forms reliability for timed measures ranged from .78 to .93, and average stratified alpha for composite scores ranged from 0.90 to 0.92.
- Internal consistency of untimed early literacy measures for students in grades K-1 ranged from 0.63 to 0.93, average alternate forms reliability for timed measures ranged from .78 to .97, and stratified alpha for composite scores ranged from 0.91 to 0.95.

- Average internal consistency of untimed reading measures for students in grades 2-8 ranged from 0.67 to 0.87, average alternate forms reliability for timed measures ranged from .86 to .96, and average stratified alpha for composite scores ranged from 0.87 to 0.92.

Validity

Validity is the degree to which evidence supports interpretations of test scores for a given purpose. There are several different types of validity evidence that can be provided, depending on the proposed use of the test. Because aimswebPlus is used to identify students at risk of academic failure and track progress toward academic goals in reading and math, one particularly relevant form of validity evidence is the extent to which performance on the tests correlates with performance on other measures, which are called criterion measures. Correlations with criterion measures administered at the same time are called concurrent validity coefficients, and correlations with criterion measures administered at a later time are called predictive validity coefficients. These coefficients can range from -1.0 to 1.0, with positive values closer to 1.0 indicating a stronger positive relationship. To the extent that the coefficients are high, this suggests that the tests are doing a good job measuring targeted reading and math skills and predicting future performance on end-of-year achievement tests.

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, Math, Early Literacy, and Reading benchmark composites are provided.

Early Literacy

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 12 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 13 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.

As can be seen, coefficients varied by criterion measure, with higher coefficients seen for ITBS scores. In particular, when coefficients were adjusted for range restriction, predictive validity ranges from 0.58 (for Word Reading Fluency administered in the fall) to 0.72 (for ITBS administered in the fall) and concurrent validity was 0.57 (for Word Reading Fluency) and 0.74 (for ITBS).

Table 12: Unadjusted and adjusted predictive validity coefficients of aimswebPlus

Criterion	Grade	*n*	Predictive		Sex		Race/Ethnicity			
			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	I (Fall)	61	0.57	0.72	41	59	25	25	17	33

Table 13: Concurrent validity coefficients

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

Reading

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 14 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 15 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.

As can be seen, all average coefficients but two, adjusted for range restriction were at least 0.70. In particular, mean predictive validity coefficients range from 0.69 to 0.83 and mean concurrent validity coefficients range from 0.68 to 0.80.

References

Adams, M. J. (1990). Beginning to read: Thinking and learning about print. Cambridge, MA: The MIT Press.

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

Baglici, S. P., Codding, R., & Tryon, G. (2010). Extending the research on the tests of early numeracy: Longitudinal analyses over two school years. *Assessment for Effective Intervention, 35*(2), 89102.

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. Journal of Learning Disabilities, 38(4), 333–339.

Burland, A. (2011). Statistical relationship among number sense, computational fluency and Montana comprehensive assessment system (Doctoral dissertation). University of Montana, Missoula, MT.

Carnine, D. W., Silbert, J., Kame'enui, E. J., & Tarver, S. G. (2010). Direct instruction reading (5th ed.), Boston, MA: Merrill.

Clarke, B., Baker, S.K., Smolkowski, K., and Chard, D. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in student outcomes. *Remedial and Special Education*, 29, 46–57.

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.

Clemens, N. H., Shapiro, E. S., & Thoemmes, F. (2011). Improving the efficacy of first grade reading screening: An investigation of word identification fluency with other early literacy indicators. School Psychology Quarterly, 26, 231–244.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. Exceptional Children, 49(1), 3645.

Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, & DHHS. (2000).
Report of the National Reading Panel: Teaching Children to Read: Reports of the Subgroups (00-4754). Washington, DC: U.S. Government Printing Office.

Feldmann, G. (2012). Early numeracy: technical adequacy of select kindergarten and first grade screening measures. (Doctoral Dissertation, University of Iowa, 2012). http://ir.uiowa.edu/etd/2869

Floyd, R. G., Hojnoski, R., & Key, J. (2006). Preliminary evidence of the technical adequacy of the preschool numeracy indicators. School Psychology Review, 35(4), 627–644.

Fry, E. B, & Kress, J. E. (2006). The reading teacher's book of lists (5th ed.). San Francisco, CA: Jossey-Bass.

Fuchs, L. S., Fuchs, D. & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. Exceptional Children, 71, 7–21.

Fuchs, L. S., Fuchs, D., & Deno, S. L. (1982). Reliability and validity of curriculum-based informal reading inventories. *Reading Research Quarterly, 18*(1), 626.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education*, 9(2), 2028.

Fuchs, L.S., & Vaughn, S.R. (2005). Response-to-intervention as a framework for the identification of learning disabilities. *Trainer's Forum: Periodical of the Trainers of School Psychologists, 25*(1), 12–19.

Gersten, R., Clarke, B., Jordan, N. C., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for the primary grades: Beginnings of a research base. Council for Exceptional Children, 78(4), 423–445.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. Journal of Learning Disabilities, 38(4), 293–304.

Goldman, R., & Fristoe, M. (2000). Goldman-Fristoe Test of Articulation 2. Bloomington, MN: Pearson.

Hiebert, E. H., Samuels, S. J., & Rasinski, T. V. (2012). Comprehension-based silent reading rates. What do we know? What do we need to know? Literary Research and Instruction, 51(2), 110–124. http://dx.doi.org/10.1080/19388071.2010.531887

Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in Kindergarten and first grade: Prediction of mathematics proficiency in third grade. School

Psychology Review, 39(2), 181–195.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. Learning Disabilities Research & Practice, 22(1), 36–46.

Jordan, N. C., Kaplan, D., Oláh, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. Child Development, 77(1), 153–175.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. Developmental Psychology, 45(3), 850–867. doi:10.1037/a0014939.

Landauer, T. K. (2011). Pearson's text complexity measure. Iowa City, IA: Pearson White Paper. Retrieved from http://www.pearsonassessments.com/textcomplexity

Landauer, T. K., Kireyev, K., & Panaccione, C. (2011). Word maturity: A new metric for word knowledge. Scientific Studies of Reading, 15(1), 92–108.

Lembke, E., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and first-grade students. Learning Disabilities Research & Practice, 24(1), 12–20.

Lembke, E., Foegen, A., Whittaker, T. A., & Hampton, D. (2008). Establishing technically adequate measures of progress in early numeracy. Assessment for Effective Intervention, 33(4), 206–214.

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. Journal of Learning Disabilities, 41(5), 451–459.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. Developmental Psychology, 36(5), 596–613. http://dx.doi.org/10.1037/0012–1649.36.5.596

Markovitz, Z., & Sowder, J. (1988). Mental computation and number sense. In M. J. Behr, C. B. Lacampagne, & M. M. Wheeler (Eds.), Proceedings of the tenth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 58–64). DeKalb, IL: Northern Illinois University. (ERIC Document Reproduction Service No. ED 411 126).

Martinez, R. S., Missall, K. N., Graney, S. B., Aricak, O. T., & Clarke, B. (2009). Technical adequacy of early numeracy curriculum-based measurement in kindergarten. *Assessment for Effective Intervention, 34*(2), 116125.

Mazzocco, M. M. M., & Thompson, R. E. (2005). Kindergarten predictors of math learning disability. Learning Disabilities Research and Practice, 20(3), 142–155. doi:10.1111/j.1540–5826.2005.00129.x

McGraw-Hill Education. (2008). Number knowledge test. New York, NY: Author.

McIntosh, A., Reys, B. J., & Reys, R. E. (1992). A proposed framework for examining basic number sense. For the Learning of Mathematics, 12(3), 1–7.

Meisinger E., Dickens, R., & Tarar, J. (2015). Oral and silent reading fluency: Assessment to intervention. Paper presented at the annual meeting of the National Association of School Psychologists, Orlando, FL.

Methe, S. A., Begeny, J. C., & Leary, L. L. (2011). Development of conceptually focused early numeracy skill indicators. Assessment for Effective Intervention, 36(4), 230–242. doi: 10.1177/1534508411414150

Nation, K., & Hulme, C. (1997). Phonemic segmentation, not onset-rime segmentation, predicts early reading and spelling skills. Reading Research Quarterly, (32)2, 154–167. doi:10.1598/RRQ.32.2.2

National Council of Teachers of Mathematics. (1989). Principles and standards for school mathematics. Reston, VA: Author.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). Common core state standards. Washington, DC: Authors.

National Institute of Child Health and Human Development See Eunice Kennedy Shriver National Institute of Child Health and Human Development

National Mathematics Advisory Panel. (2008). Final report. Washington, DC: Author.

National Research Council. (2001). Adding it up: Helping children learn mathematics. J. Kilpatrick, J. Swafford, and B. Findell (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Center on Educational Outcomes (NCEO). (2002). Universal design of assessments. Retrieved from https://nceo.info/Assessments/universal_design

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

Partnership for Assessment of Readiness for College and Careers (PARCC). (2014). Mathematics model content frameworks: Kindergarten through grade 2. Washington, DC: Author.

Pearson. (2007). Stanford achievement test series (10th ed.). San Antonio, TX: Author. Pearson. (2017). aimswebPlus Technical Manual. Bloomington, MN: Author.

Purpura, D. J., Reid, E. E., Eiland, M. D., & Baroody, A. J. (2015). Using a brief preschool early numeracy skills screener to identify young children with mathematics difficulties. School Psychology Review, 44, 41–59.

Runge, T. J., & Watkins, M. W. (2006). The structure of phonological awareness among kindergarten students. School Psychology Review, 35, 370–386.

Seethaler, P. M., & Fuchs, L. S. (2011). Using curriculum-based measurement to monitor kindergarteners' mathematics development. Assessment for Effective Intervention, 36. doi:10.1177/1534508411413566

Shinn, M. R. (2012). Progress on early literacy universal screening and progress monitoring: Highly decodable reading passages. Unpublished manuscript.

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. (1992). Curriculum-Based reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*(3), 458–478.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). Preventing reading difficulties in young children. Washington, DC: National Academy Press.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994) (PS) Longitudinal studies of phonological processing and reading. Journal of Learning Disabilities, 27, 276–286.

Vellutino, F. R., & Scanlon, D. M. (1987) Phonological coding, phonological awareness, and reading ability: Evidence from a longitudinal and experimental study. Merrill-Palmer Quarterly (Wayne State University. Press), 33(3), 321–363.

Woodcock, R. W., Shrank, F. A., McGrew, K. S., & Mather, N. (2005). Woodcock–Johnson III. Boston, MA: Houghton Mifflin Harcourt.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. Reading Research Quarterly, 23, 159–178.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). The educator's word frequency guide. Brewster, NY: Touchstone Applied Science Associates.