

**An Analysis of the Architectural Integrity of
Reading Comprehension Assessment
and Instructional Materials for Struggling Readers:**

**Ode to Faulkner, Twain, and Kerr
(with apologies to Hemingway)**

**Edward J. Kame'enui, Ph.D.
Professor and Director
Center for Teaching and Learning
College of Education
University of Oregon**

Ode to Faulkner, Twain, and Kerr (with apologies to Hemingway)

- He has never been known to use a word that might send a reader to the dictionary.

William Faulkner (about Ernest Hemingway)

- I didn't attend the funeral, but I sent a nice letter saying I approved of it.

Mark Twain

- He had delusions of adequacy.

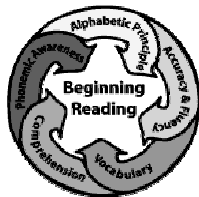
Walter Kerr

Websites



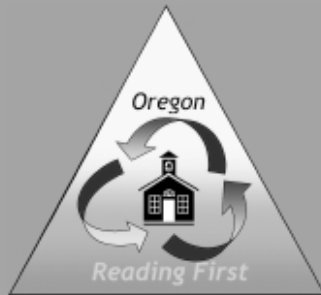
***Institute for the
Development of Educational Achievement (IDEA)***

<http://idea.uoregon.edu/>



IDEA Beginning Reading

<http://reading.uoregon.edu/>



Oregon Reading First Supplemental and Intervention Programs Review

General Information

Oregon Reading First Participants

Teachers

Coaches

Administrators

Pathfinder

[Announcements and News](#)

[Frequently Asked Questions](#)

[Links](#)

Oregon Reading First Supplemental and Intervention Programs Review

Last update: 8/10/04

The following report, the Review of Supplemental and Intervention Reading Programs, describes the work of the Oregon Curriculum Review Panel and the Oregon Reading First Center, who had responsibility for conducting a critical review of supplemental and intervention curriculum programs in beginning reading for the first cohort of 34 schools participating in Oregon Reading First.

The Panel has identified [106 High Priority Supplemental and Intervention programs](#) for review.

Reviews of specific programs will be released in one of two ways:

Programs Reviewed Before June 15, 2004:

Eighty-five programs were reviewed prior to June 15, 2004 and are included in the updated Supplemental/Intervention Curriculum Review.

View a list of the [85 programs included in the updated review](#).

Download the [Table of Contents](#) for the Updated Supplemental and Intervention Curriculum Review.

Full Report

Download the full [Updated Supplemental and Intervention Curriculum Review Report](#).

Programs reviewed prior to June 15, 2004 are also included in the following summary tables:

Program Item-by-Item Summary by Big Idea:

[Comprehension Item Analysis](#)

Reading First Assessment Committee*

Team Leader Edward J. Kame'enui, University of Oregon

- David J. Francis, University of Houston
- Lynn Fuchs, Vanderbilt University
- Roland Good III, University of Oregon
- Rollanda O'Connor, University of Pittsburgh
- Deborah C. Simmons, Texas A&M University
- Gerald Tindal, University of Oregon
- Joseph Torgesen, Florida State University

* Partnership for Reading, a project of the National Institute for Literacy (NIFL)

Focus of Presentation*

- RFAC's activities, procedures, findings, and documents.
- Define scope and purpose of assessment.
- Explain guidelines and process RFAC used to judge technical merit of reading measures.
- Summarize findings of the RFAC review of assessments.

*Based on *Determining the sufficiency of evidence in the assessment of reading competence in the primary grades* (RFAC, manuscript in preparation) and Members of the Assessment Committee (2002) *Final report: The analysis of reading assessment instruments for K-3*. National Institute for Literacy.

Reading First assessment framework aligned with five essential components of reading instruction specified in 2002 NCLB legislation (see Part B--Student Reading Skills Improvement Grants, Subpart I--Reading First, Sec. 1208, pp. 203-205; see purposes of assessment, pp. 205-206)

Four Purposes of Beginning Reading Assessment

1. Screening

- Brief assessments conducted with all children at the beginning of the school year.
- It targets skills that are strongly predictive of important future reading outcomes.
- The goal is to identify children who are at-risk of reading failure, and, therefore, likely to need additional or alternative forms of reading instruction to supplement the conventional general education approach.

Four Purposes of Beginning Reading Assessment

2. Diagnosis

- More in-depth assessment of reading strengths and weaknesses.
- Not to be confused with assessment for the purpose of labeling children with disabilities.
- The goal is to provide teachers with a profile of reading skills and deficits sufficient to inform instruction so that students make adequate progress toward reading outcomes.

Four Purposes of Beginning Reading Assessment

3. Progress Monitoring

- Students are assessed a minimum of three times a year but typically more frequently (e.g., weekly, monthly, or quarterly).
- Employs alternate forms of a test when children demonstrate reading difficulties.
- The purpose is to (a) estimate rates of reading improvement, (b) identify children who are not demonstrating adequate progress and will need supplementary instruction, or (c) compare the efficacy of different forms of instruction for an individual student, thereby informing the design of more effective reading programs for at-risk readers.

Four Purposes of Beginning Reading Assessment

4. Outcome measurement

- Conducted to address the evaluation requirement of NCLB.
- Students are classified as having achieved or not achieved grade-level performance or as having improved or not improved.

Reading First Initiative: *Levels of Reading Outcomes*

States need to know, annually, whether they

- a) are making progress towards reducing the number of students who are reading below grade level,
- b) have significantly increased the number of students reading *on grade level* or above; and
- c) have significantly increased the percentages of students in ethnic, racial, and low-income populations who are reading *on grade level* or above.

Defining On Grade Level

- Continuum with three critical points on a scale anchored to level of instructional support necessary to maintain or accelerate learning at sufficient rates to reach grade level proficiency.
- Two approaches: Normative strategy and absolute standard.
- Within either framework, student considered ***on grade level*** when child obtains a score on a test that yields a reasonably confident prediction the he/she will meet grade-level expectations for reading proficiency on a later state accountability assessment.

Defining On Grade Level

- Needs Substantial Intervention: When child's obtained score yields a prediction that indicates the the student is clearly not likely to achieve grade-level expectations on a later state accountability assessment without an intervention protocol.
- Needs Additional Intervention: When no confident prediction can be formulated that the student will either meet or not meet grade level expectations.

Normative Approach

- Norms must be established within past seven years, reasonably representative of U.S. population (i.e., within 10 percentage points of current census with respect to socio-economic status, race, gender, and geographic location).
- On Grade Level: Students scoring at or above the 40th percentile (.25 standard deviations below the mean in a normal distribution which is equivalent to a standard score of 96 with a mean of 100 and standard deviation of 15).
- Needs Additional Intervention: Students scoring above the 20th percentile and at or below the 39th percentile (.84 standard deviations below the mean or a standard score of 87 with a mean of 100 and standard deviation of 15).
- Needs Substantial Intervention: Students scoring at or below the 20th percentile.

Absolute Standard Approach

1. If state has defined standards of performance tied to a state outcome assessment at a given grade level, then standards define the *de facto* absolute standard for that state and grade.
2. Develop an absolute standard for grade level performance on an assessment provide evidence of predictive validity exists for that test to some later criterion of reading (e.g., a state high stakes test) or some functional outcome of interest (e.g., ability to successfully perform grade level work in some specified future grade).
3. Use predictive validity of earlier administered test to define the absolute standard on the earlier administered test.
 - Students for whom the conditional probability of scoring **at or above** the defined state standard on the critical outcome test is at least 67% would be considered on grade level on the earlier test.
 - Students for whom the conditional probability of scoring **below** the defined state standard on the critical outcome test is at least 67% are deemed as Needs Substantial Intervention.
 - Students for whom the conditional probability of scoring **at or above** the defined state standard on the critical outcome test is between 34% and 66% are deemed as Needs Additional Intervention.

Reporting Outcome Assessment

- Outcomes in the important beginning reading skill areas *should be* reported at the beginning and end of the academic year.
 - Phonemic Awareness
 - Phonics
 - Fluency
 - Comprehension
 - Vocabulary
- The following assessment framework specifies the important beginning reading skill areas to be assessed at the beginning and end of each grade level.

Kindergarten Outcome Assessment

Suggested
Beginning of
year

Optional
Middle of
Year

Suggested
End of
Year

Phonemic Awareness	X	X	X
Phonics		X	X
Fluency			
Comprehension			
Vocabulary	X	X	X

First Grade Outcome Assessment

	<i>Suggested</i> Beginning of year	<i>Optional</i> Middle of Year	<i>Suggested</i> End of Year
Phonemic Awareness	X	X	X
Phonics	X	X	X
Fluency		X	X
Comprehension			X
Vocabulary	X	X	X

Selecting & Obtaining Reading Measures

Hundreds of measures & limited timeframe (8 months)

1. Not possible to review every assessment tool.
2. Goal to develop criteria and procedures for evaluating technical adequacy and sufficiency of evidence for K-3 reading assessment measures that meet RF requirements.
3. Provide recommendations for an initial set of measures; a process for judging merit of other tools not reviewed; process for reconsideration of tools that had been revised or further researched.
4. Began with Southwest Educational Development Laboratory's *Reading Assessment Database for Grades K-2: Statewide Assessments (SEDL, 2001)*.
5. Five criteria applied to selection of tools: (a) widely used in schools, (b) represent one of four assessment purposes, (c) assess at least one of five essential components of RF, (d) available and accessible for wide-scale use, and (e) readily available or quickly attainable for review.

Analysis of Reading Assessment Measures (ARAM) Coding Form

Level I Review: Basic Information

A. MEASURE

Name: _____

Developer(s): _____

Publisher: _____ Reviewer: _____

Publication Date: _____

DESCRIPTIVE INFORMATION OF INSTRUMENT

1. The instrument is intended for use in grade(s) K, 1, 2, and /or 3. *(Check all that apply)*

<input type="checkbox"/> Kindergarten	<input type="checkbox"/> Third Grade
<input type="checkbox"/> First Grade	<input type="checkbox"/> Beyond Third Grade
<input type="checkbox"/> Second Grade	

The instrument is intended for use with age(s) 4-9. *(Check all that apply)*

<input type="checkbox"/> 4	<input type="checkbox"/> 6	<input type="checkbox"/> 8	<input type="checkbox"/> Beyond 9
<input type="checkbox"/> 5	<input type="checkbox"/> 7	<input type="checkbox"/> 9	

B. DESCRIPTIVE INFORMATION OF INSTRUMENT continued

2. The instrument assesses one or more of the following dimensions of reading. *(Check all that apply)*

- | | |
|------------------------|--|
| q Phonemic Awareness : | q Listening Comprehension |
| q Initial Sound | q Reading Comprehension |
| q Blending | q Oral q Cloze q Maze |
| q Segmenting | q Silent q Retell q Question/Answer |
| q Invented Spelling | q Other |
| q Letter Names | _____ |
| q Timed | _____ |
| q Untimed | |
| q Letter Sounds | q Vocabulary <i>List Specific Skills or Subtests</i> |
| q Timed | q Oral _____ |
| q Untimed | q Reading _____ |
| q Word Attack | |
| q Nonsense Words | q Other <i>List Specific Skills or Subtests</i> |
| q Timed | _____ |
| q Untimed | _____ |
| q Word Identification | |
| q Regular | |
| q Irregular | |
| q Reading Fluency | |
| q Words | |
| q Sentences | |
| q Connected Text | |

B. DESCRIPTIVE INFORMATION OF INSTRUMENT continued

3. The instrument provides the following type(s) of information: *(Check all that apply)*

- | | | |
|--|-------------------------------------|-----------------------------------|
| <input type="checkbox"/> Screening | <input type="checkbox"/> Definitely | <input type="checkbox"/> Possibly |
| <input type="checkbox"/> Diagnostic | <input type="checkbox"/> Definitely | <input type="checkbox"/> Possibly |
| <input type="checkbox"/> Progress Monitoring | <input type="checkbox"/> Definitely | <input type="checkbox"/> Possibly |
| <input type="checkbox"/> Outcome | <input type="checkbox"/> Definitely | <input type="checkbox"/> Possibly |

4. The instrument provides information on student performance in English and/or Spanish. English Spanish Other

FORMATIVE SAMPLE

1a. Check type of sample: Normative Criterion-referenced

1b. Characteristics of the normative sample. National representation? Yes No

Number of States: _____

Date: _____

Regions: _____

Size: _____

Level II Review: Development and Administration

A. TIME, ADMINISTRATION, AND FREQUENCY

Assessment format:

individual only group or individually
 both explain: _____
 _____ minutes

		Individual
Test Name	Admin.	Scoring
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

Discontinue Rules:

yes no
 basals ceilings other
 no yes (# of forms: _____)

D. DESCRIPTION (Narrative Overview of Test):

TEST / SUBTEST NAME:

This test assesses:

Assessment administration looks like:

What is the response format?

TEST / SUBTEST NAME:

This test assesses:

Assessment administration looks like:

What is the response format?

E. DEVELOPMENT

Development:

Date:

Renormed:

Date:

Level III Review: Technical Adequacy

Complete one technical adequacy sheet for each subtest of the measure.

SKILL AREA/SUBTEST: _____

VALIDITY

Type of Validity	Age or Grade	Test or Criterion	n (range)	Coefficient	
				range	median

Other forms of validity: _____

Manual cites other published validity studies: yes no

SKILL AREA/SUBTEST: _____

RELIABILITY

Type of Reliability	Age or Grade	n (range)	Coefficient		SEM
			range	median	

Manual cites other published reliability studies: yes no

Outcome Measures

Complete this section for outcome measures. Complete one copy of this sheet for each subtest, if process for operationalizing improvement differs by subtest.

SKILL AREA/SUBTEST: _____

INDEXING SIGNIFICANT GAINS

Improvement is operationalized as: _____

q norm referenced

q absolute criterion referenced (*specify*)

q not specified

INDEXING GRADE-LEVEL PERFORMANCE

Grade-level performance is operationalized as: _____

q norm referenced

q absolute criterion referenced (*specify*)

q not specified

Screening Measures

Complete this section for screening measures. Complete one sheet for each subtest.

DECISION-MAKING UTILITY

EVIDENCE OF SPECIFICITY

How are false negatives and false positives assessed? (Criterion and grade/age)

	Negative	Positive
Absent	TN	FP
Present	FN	TP

EVIDENCE OF SENSITIVITY

Odds ratios and conditional probabilities if given:

Reliability of decisions:

Calculations:

Specificity: $TN / (TN + FP) =$

Hit rate: $(TP + TN) / N =$

Sensitivity: $TP / (TP + FN) =$

Diagnostic Measures

Complete this section for diagnostic measures.

1. Check a reas f or which diagn ostic informat ion is pr ovided.

	Some Informat ion ab out Doma in	No Informat ion abo ut Doma in
Phon emic Aware ness		
Letter Names		
Letter Sounds		
Word Attack		
Word Identificat ion		
Rea ding Fluency		
Listening Compre hens ion		
Rea ding Compre hens ion		
Vocab ulary		
Other		

If information is provided at a skill or strategy level, please specify below:

Example s:

For decoding: CVC, CVCe, r-controlled For comprehension: literal, inferential, strategy use

Diagnostic Measures

Complete items 2, 3, and 4 for each skill area or subtest.

2. Evidence f or co rresp onde nce with c riter ion me a sures:
3. Evidence f or reliab ility:
4. Evidence f or improve d instruc tion or lea rning:

Progress Monitoring Measures

Complete this section for progress monitoring measures.

Recommended Administration Period:

Test / Subtest Name:

Recommended Frequency of Data Collection

For At-Risk Students: _____ For Others: _____

Does the measurement construct remain constant over the school year? yes no

Criterion for Adequate Growth:

Test / Subtest Name:

_____ Specified Not specified

Criterion for Adequate Performance:

Test / Subtest Name:

_____ Specified Not specified

Number of Data Points Needed To Make Reliable Decision: _____

Evidence for Utility (Describe research on improved instruction or learning):

Evidence for Sensitivity to Growth (Describe research showing relation between growth on this measure with growth on other measures):

Evidence for Sensitivity to Treatment Effects (Describe research showing that treatment effects are revealed with this repeated measurement):

Acquisition Information

ACQUISITION INFORMATION

Where to Obtain: _____

Phone #: _____

Address: _____

Web Site: _____

Cost:

\$ _____ Complete Kit (describe contents)

q can purchase replacement components individually
q information not available
q materials not consumable, so NA

\$ _____ Manuals and Test Materials

\$ _____ Directions for Administration

\$ _____ Test Forms

\$ _____ Technical Manuals

\$ _____ Protocol per Student

Other (*Describe below*)

Level IV Review: Summary

A. PURPOSES

1. STRENGTHS

2. WEAKNESSES

B. RECOMMENDATIONS

Assessment Committee

Criteria to Evaluate Reading Measure Evidence

Reading Measure/Composite/Subtest: _____

[If subtests form a composite, analyze at the composite level]

Domain: _____

[Phonological Awareness, Phonics, Fluency, Vocabulary, Reading Comprehension]

Directions

1. Circle an **S** for **SUFFICIENT EVIDENCE**, an **NS** for **NOT SUFFICIENT EVIDENCE**, and an **NE** for **NO EVIDENCE** in each cell for each appropriate feature on the attached pages. These codes are used to index the nature of the evidence that a measure/composite/subtest is valid, reliable, and based on scientifically based reading research as specified in the Reading First legislation.
2. Complete Sections I and II for the entire test (not composite or subtest).
3. For Sections III, IV, and V, complete a separate form for each domain of reading (i.e., phonemic awareness, phonics, fluency, vocabulary, and comprehension) within each measure.
4. In each cell of Sections III, IV, and V, circle the grade (s) (K, 1, 2, and 3), and the code (SE, NS, NE) for each grade, if appropriate.
5. NOTE: *For items #6, 7, and 8, Reviewers are encouraged to use their expert judgment based on the body of evidence for each measure to determine if the reliability and validity evidence are sufficient.

Feature	Screening			Diagnosis			Prog. Monitoring			Outcome						
<i>I. Development and Administration</i>																
1. Test description provided.	S	NS	NE	S	NS	NE	S	NS	NE	S	NS	NE				
2. Test development described.	S	NS	NE	S	NS	NE	S	NS	NE	S	NS	NE				
3. Individual or group administration specified.	S	NS	NE	S	NS	NE	S	NS	NE	S	NS	NE				
4. Training requirements specified.	S	NS	NE	S	NS	NE	S	NS	NE	S	NS	NE				
<i>II. Acquisition</i>																
5. Acquisition information provided.	S	NS	NE	S	NS	NE	S	NS	NE	S	NS	NE				
<i>III. Reliability*</i>																
6. Reliability information specified for each subtest.	K	1	2	3	K	1	2	3	K	1	2	3	K	1	2	3
	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
<i>IV. Validity*</i>																
7. Concurrent validity information specified for each subtest.	K	1	2	3	K	1	2	3	K	1	2	3	K	1	2	3
	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
8. Predictive validity information specified for each subtest.	K	1	2	3	K	1	2	3	K	1	2	3	K	1	2	3
	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
<i>V. Use</i>																
9. Sample provides an adequate reference point for interpretation.	K	1	2	3	K	1	2	3	K	1	2	3	K	1	2	3
	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
10. Scoring structure specified.	K	1	2	3	K	1	2	3	K	1	2	3	K	1	2	3
	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE

One-Page Form

Feature	Screening	Diagnosis	Progress Monitoring	Outcome																
<i>V. Use</i>																				
11. Method to document significant gain or improvement specified.				<table border="1"> <tr><td>K</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>S</td><td>S</td><td>S</td><td>S</td></tr> <tr><td>NS</td><td>NS</td><td>NS</td><td>NS</td></tr> <tr><td>NE</td><td>NE</td><td>NE</td><td>NE</td></tr> </table>	K	1	2	3	S	S	S	S	NS	NS	NS	NS	NE	NE	NE	NE
K	1	2	3																	
S	S	S	S																	
NS	NS	NS	NS																	
NE	NE	NE	NE																	
12. Method to document grade-level performance specified.				<table border="1"> <tr><td>K</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>S</td><td>S</td><td>S</td><td>S</td></tr> <tr><td>NS</td><td>NS</td><td>NS</td><td>NS</td></tr> <tr><td>NE</td><td>NE</td><td>NE</td><td>NE</td></tr> </table>	K	1	2	3	S	S	S	S	NS	NS	NS	NS	NE	NE	NE	NE
K	1	2	3																	
S	S	S	S																	
NS	NS	NS	NS																	
NE	NE	NE	NE																	
13. Diagnostic information specified.		<table border="1"> <tr><td>K</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>S</td><td>S</td><td>S</td><td>S</td></tr> <tr><td>NS</td><td>NS</td><td>NS</td><td>NS</td></tr> <tr><td>NE</td><td>NE</td><td>NE</td><td>NE</td></tr> </table>	K	1	2	3	S	S	S	S	NS	NS	NS	NS	NE	NE	NE	NE		
K	1	2	3																	
S	S	S	S																	
NS	NS	NS	NS																	
NE	NE	NE	NE																	
14. Instructional level specified.		<table border="1"> <tr><td>K</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>S</td><td>S</td><td>S</td><td>S</td></tr> <tr><td>NS</td><td>NS</td><td>NS</td><td>NS</td></tr> <tr><td>NE</td><td>NE</td><td>NE</td><td>NE</td></tr> </table>	K	1	2	3	S	S	S	S	NS	NS	NS	NS	NE	NE	NE	NE		
K	1	2	3																	
S	S	S	S																	
NS	NS	NS	NS																	
NE	NE	NE	NE																	
15. Alternate forms are used to monitor progress toward a long term goal.			<table border="1"> <tr><td>K</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>S</td><td>S</td><td>S</td><td>S</td></tr> <tr><td>NS</td><td>NS</td><td>NS</td><td>NS</td></tr> <tr><td>NE</td><td>NE</td><td>NE</td><td>NE</td></tr> </table>	K	1	2	3	S	S	S	S	NS	NS	NS	NS	NE	NE	NE	NE	
K	1	2	3																	
S	S	S	S																	
NS	NS	NS	NS																	
NE	NE	NE	NE																	
16. Evidence of validity for positive and negative screening decisions provided.	<table border="1"> <tr><td>K</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>S</td><td>S</td><td>S</td><td>S</td></tr> <tr><td>NS</td><td>NS</td><td>NS</td><td>NS</td></tr> <tr><td>NE</td><td>NE</td><td>NE</td><td>NE</td></tr> </table>	K	1	2	3	S	S	S	S	NS	NS	NS	NS	NE	NE	NE	NE			
K	1	2	3																	
S	S	S	S																	
NS	NS	NS	NS																	
NE	NE	NE	NE																	
17. Criteria for adequate performance and/or measuring change in performance within the academic year can be specified.			<table border="1"> <tr><td>K</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>S</td><td>S</td><td>S</td><td>S</td></tr> <tr><td>NS</td><td>NS</td><td>NS</td><td>NS</td></tr> <tr><td>NE</td><td>NE</td><td>NE</td><td>NE</td></tr> </table>	K	1	2	3	S	S	S	S	NS	NS	NS	NS	NE	NE	NE	NE	
K	1	2	3																	
S	S	S	S																	
NS	NS	NS	NS																	
NE	NE	NE	NE																	
18. Procedures for decision making are specified and demonstrated to be reliable.			<table border="1"> <tr><td>K</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>S</td><td>S</td><td>S</td><td>S</td></tr> <tr><td>NS</td><td>NS</td><td>NS</td><td>NS</td></tr> <tr><td>NE</td><td>NE</td><td>NE</td><td>NE</td></tr> </table>	K	1	2	3	S	S	S	S	NS	NS	NS	NS	NE	NE	NE	NE	
K	1	2	3																	
S	S	S	S																	
NS	NS	NS	NS																	
NE	NE	NE	NE																	

One-Page Form

SUMMARY of EVIDENCE

Test _____

Rating: S = Sufficient Evidence NS = Not Sufficient Evidence NE = No Evidence

Key: S = Screening, D = Diagnostic, PM = Progress Monitoring, O = Outcome

Directions: For each test:

- Circle the appropriate code by grade, “Big Idea,” and purpose of the measure.
- Be sure to rate each subtest of a Big Idea.
- Write the test, composite, and test title below the Big Idea.

Kindergarten	Grade 1	Grade 2	Grade 3
Phonemic Awareness	Phonemic Awareness	Phonemic Awareness	Phonemic Awareness
<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE
Phonics (Winter/Spring)	Phonics	Phonics (Fall only)	Phonics
<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE
Fluency	Fluency (Winter/Spring)	Fluency	Fluency
<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE
Vocabulary	Vocabulary	Vocabulary	Vocabulary
<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE
Reading Comprehension	Reading Comprehension (Spring only)	Reading Comprehension	Reading Comprehension
<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE	<i>S</i> S NS NE <i>D</i> S NS NE <i>PM</i> S NS NE <i>OU</i> S NS NE

* Shaded box means that Big Idea is not a priority for that particular grade level

Table 2. Test Features by Test Domain

Feature	Phonemic Awareness (n=16)					Phonics (n=17)					Fluency (n=11)			
	n	(%)	X	(SD)	Ê	n	(%)	X	(SD)	Ê	n	(%)	X	(SD)
Administration														
Adm. Time: Group	2	(13)	41.25	(33.58)		1	(6)	65.00	(--)		0	(0)	--	(--)
Individual	8	(50)	13.44	(11.17)		9	(53)	11.22	(14.19)		4	(36)	9.25	(11.32)
Score Time	5	(31)	5.50	(5.47)		5	(29)	4.90	(3.72)		1	(9)	1.00	(--)
Alternate Forms	15	(94)	2.33	(5.11)		16	(94)	3.31	(4.87)		11	(100)	6.27	(9.65)
Training > 1 Hour ^a	7	(58)				10	(67)				8	(89)		
Professional Examiner ^a	9	(75)				10	(67)				7	(70)		
Technical Features														
Representative Sample	10	(63)				12	(71)				9	(82)		
Validity: Concurrent	6	(38)	.41	(0.08)		11	(65)	.64	(0.20)		7	(64)	.70	(0.24)
Criterion	5	(31)	.52	(0.17)		4	(24)	.62	(0.20)		2	(18)	.42	(0.18)
Construct	0	(0)	----			1	(6)	.48	(--)		1	(9)	.71	(--)
Predictive	4	(25)	.53	(1.15)		5	(29)	.65	(0.15)		2	(18)	.71	(0.04)
Unspecified	1	(6)	.48	(--)		1	(6)	.44	(--)		0	(0)	--	(--)
Reliability:														
Consistency	8	(50)	.82	(0.17)		12	(71)	.91	(0.07)		6	(55)	.88	(0.14)
Stability	4	(25)	.85	(0.10)		9	(53)	.89	(0.07)		7	(64)	.90	(0.08)
Form	1	(6)	.67	(--)		4	(24)	.89	(0.05)		3	(27)	.94	(0.01)
Unspecified	1	(6)	.90	(--)		1	(6)	.92	(--)		1	(9)	.90	(--)
Ê	Ê	ÊÊ	Ê	Ê		Ê	ÊÊ	Ê	Ê		Ê	ÊÊ	Ê	

Table 2. Test Features by Test Domain continued

Feature	Vocabulary (n=14)				Comprehension (n=12)			
	n	(%)	X	(SD)	n	(%)	X	(SD)
Administration								
Adm. Time: Group	1		65.00	(--)	2	(17)	55.00	(14.14)
Individual	5		29.20	(18.86)	4	(33)	39.38	(25.03)
Score Time	1		5.00	(--)	1	(8)	7.50	(--)
Alternate Forms	13	(93)	1.85	(1.57)	44	(92)	2.82	(2.40)
Training > 1 Hour ^a	10	(83)			7	(78)		
Professional Examiner ^a	10	(71)			10	(83)		
Technical Features								
Representative Sample	13	(93)			12	(100)		
Validity: Concurrent	4	(29)	.52	(0.21)	5	(42)	.57	(0.18)
Criterion	5	(36)	.58	(0.26)	5	(42)	.51	(0.22)
Construct	2	(14)	.49	(0.20)	1	(8)	.77	(--)
Predictive	0	(0)	--	(--)	0	(0)	--	(--)
Unspecified	1	(7)	.48	(--)	1	(8)	.20	(--)
Reliability:								
Consistency	10	(71)	.81	(0.11)	9	(75)	.88	(0.11)
Stability	6	(43)	.84	(0.05)	5	(42)	.87	(0.05)
Form	2	(14)	.87	(0.10)	2	(17)	.80	(0.07)
Unspecified	0	(0)	--	(--)	1	(8)	.90	(--)

Not reported for all tests. To determine number of tests reported (i.e., denominator for percentage) divide the n by the percentage (e.g., for Training >/for Phonemic Awareness, 7/.58=12) so, 12 tests reported training requirement for Phonemic Awareness).

Note: If there were more than 1 subtest/test assessing a domain, information was collapsed across the subtests.

Findings

- 29 Measures: Many instruments failed to report essential administration information (e.g., administration & scoring times, number of alternate forms, training time or required administrator credentials).
- Many instruments failed to report essential technical features of tests.
 - only 6 of 16 (38%) phonemic awareness measures reported concurrent validity.
 - only 4 of these 16 measures reported some form of consistency reliability.

Findings

- Although representative samples of norm-referenced decision making available for vocabulary and comprehension, largely inadequate or missing for PA and phonic measures.
- Validity and reliability coefficients tended to be disappointingly low across all assessment purposes and essential reading components.
 - particularly for construct and predictive validity.
 - most publishers failed to address test stability, alternate-form reliability, with estimates provided for only 50% and 75% of measures respectively.
 - only one screening instrument published any analyses of false positives and false negatives.

Grade	Screening	Diagnostic	Progress Monitoring	Outcome
K	1	1	0	1
1	5	11	1	13
2	5	12	2	13
3	0	10	1	12