



Institute for the Development of Educational Achievement

EXECUTIVE SUMMARY

Final Report

Analysis of Reading Assessment Instrument for K-3

Introduction

The Reading First legislation has five clearly stated purposes, two of which are:

“(1) To provide assistance to State educational agencies and local educational agencies and local educational agencies in establishing reading programs for students in kindergarten through grade 3 that are based on scientifically based reading research, to ensure that every student can read at grade level or above not later than the end of grade 3...”

“(3) To provide assistance to State educational agencies and local educational agencies and local educational agencies in selecting or administering screening, diagnostic, and classroom-based instructional reading assessments” (See Part B--Student Reading Skills Improvement Grants, Subpart 1--Reading First, Sec. 1201. Purposes, p.178).

The work and results represented in the Final Report are designed to provide State educational agencies and local educational agencies assistance on the selection and use of reading assessment instruments for kindergarten through grade 3.

Purpose of Report

The two purposes of the Final Report are:

1. To document the process, criteria, decision rules, guidelines, rationale, and procedures the Assessment Committee, initially identified as the “Screening, Diagnosis, and Progress Monitoring Group” (SDPM), employed in the identification and selection of a set of reading assessment instruments for possible use in K - 3;
2. To describe a “field-tested” process and set of criteria for the adjudication of evidence that meet the intent of the Reading First legislation for the selection of reading assessment instruments for use in K - 3.

The purposes of the Executive Summary for the Final Report are more modest than the Final Report. To avoid redundancy and for purposes of brevity, the Executive Summary has two purposes:

- (a) to provide a brief description of the process used to identify, select, and analyze K-3 reading assessment instruments, and
- (b) to provide a brief summary of the Assessment Committee’s results and recommendations.

Process and Analysis

The Assessment Committee developed and implemented a multifaceted process that involved the:

- (a) development of criteria for evaluating the technical adequacy of reading assessment instruments designed for use in kindergarten through grade 3;
 - (b) selection and acquisition of a set of K-3 reading assessment instruments designed to assess five essential reading components that include phonemic awareness, phonics, fluency, vocabulary, and reading comprehension;
-

- (c) analysis of the technical adequacy of four different types of reading assessment instruments designed for the purpose of screening, diagnosis, progress monitoring, and outcome;
- (d) determination of the sufficiency of the evidence that support the reading assessment instruments; and
- (e) documentation of the process the Assessment Committee employed in the evaluation and analysis of reading assessment instruments designed for kindergarten through grade 3.

Context for Report

It should be noted that the Final Report was written by one member of the committee, Edward J. Kame'enui, who served as "Team Leader" of the Assessment Committee. Therefore, it represents one committee member's narrative representation of the process in carrying out the Committee's work. This does not mean that other Committee members did not contribute ideas, words, sentences, or extensive narrative descriptions to this report, because they did in substantive, insightful, and conspicuous ways. It does mean that because of time constraints in writing and developing the final report, it was not possible to gain the full Committee's critical review, input, and collective approval on the report.

It is also important to note the Committee was convened on relatively short notice to initiate and conduct an ambitious, and at the outset, unknown and unprecedented scope of work originally scheduled to be completed in four months (September-December, 2001). What punctuated the complexity, ambitiousness, and timeline for the scope of work is that the Assessment Committee for the most part could not turn to a previous effort as a model or protocol. The Committee was required to basically create the content, process, procedures, and actual documentation (e.g., coding forms, summary forms for the coding forms) necessary to conduct the analyses of reading assessment instruments. This unprecedented scope of work in the context of an unprecedented federal legislation was enormously challenging.

It is difficult to understate the technical and conceptual complexity of the Committee's work, which makes its charge, scope of work, and timeframe unprecedented, and like all human endeavors, subject to human error. In short, the Committee was asked to make judgments about the "technical adequacy" of assessment instruments. These judgments required the Committee to: (a) develop technical criteria for evaluating existing reading assessment instruments aligned with the content focus of the Reading First legislation, (b) develop a process for applying the criteria to evaluate a set of reading assessment instruments, and (c) conduct a process for identifying, selecting, and adjudicating the "final selection" of reading assessment instruments that met the technical criteria. Accomplishing this scope of work in a concentrated timeframe (6-8 months) was an enormous challenge, and every effort was made to recognize the complexity of the work and minimize error.

Ultimately, however, the decisions in the Final Report represent human judgments about complex technical psychometric matters. In recognition of this complexity, the Committee chose to characterize its decisions using language that addresses the "sufficiency" of evidence for specific assessment instruments. For example, a favorable judgment is reported as "Sufficient" (coded as S), which means the Committee found "sufficient evidence" in the technical manual and materials of a particular assessment instrument to judge it acceptable for use in Reading First (i.e., the instrument met all or most of the essential criteria the Committee established). An unfavorable judgment is reported as "Not Sufficient" or NS, which means the Committee did not find "sufficient evidence" in the technical manual and materials of an assessment instrument to judge it acceptable for use as a screening, diagnostic, progress monitoring, or outcome instrument. It is important to note this does not mean the assessment instrument itself is "not sufficient," inadequate, or invalid. It simply means the Committee did not find evidence from the materials it reviewed "sufficient" to judge the instrument acceptable. Evidence to judge the instrument sufficient may indeed exist, but the committee did not have materials available to use in this determination. In many cases, evidence was not explicitly provided or available (NE = No Evidence) in the technical manuals of the instrument.

Results and Recommendations

The Assessment Committee reviewed and analyzed the technical information and materials for 29 reading assessment instruments designed for use in kindergarten through grade 3.

The following 24 measures out of the 29 reviewed were found to have “sufficient evidence” for use as screening, diagnosis, progress monitoring, and/or outcome instruments to assess one or more essential reading components at one or more grade levels K - 3:

Clinical Evaluation of Language Fundamentals - 3rd Edition (CELF-3)

Comprehensive Test of Phonological Processing (CTOPP)

CBM Oral Reading Fluency

Degrees of Reading Power (DRP)

Dynamic Indicators of Basic Early Literacy Skills - 5th Edition (DIBELS)

Early Reading Diagnostic Assessment (ERDA)

Gray Oral Reading Test - IV (GORT-IV)

Iowa Test of Basic Skills (ITBS)

Letter Sound Fluency

Lindamood Auditory Conceptualization Test, LAC Test

Peabody Picture Vocabulary Test - III (PPVT-3)

Phonological Awareness Test

Slosson Oral Reading Test - Revised (SORT-R)

Stanford Achievement Test - 9th Edition (SAT-9)

Terra Nova - CAT (2nd edition Terra Nova; 6th edition CAT)

Test of Language Development-Primary: 3 (TOLD-P:3)

Test of Phonological Awareness (TOPA)

Test of Word Knowledge (TOWK)

Test of Word Reading Efficiency (TOWRE)

Texas Primary Reading Inventory (TPRI)
Wechsler Individual Achievement Test-II (WIAT-II)
Woodcock Reading Mastery Test - Revised (WRMT-R)
Woodcock-Johnson III Tests of Achievement
Yopp-Singer Test of Phoneme Segmentation

The following 5 measures out of the 29 reviewed were found not to have “sufficient evidence” for use as screening, diagnosis, progress monitoring, and/or outcome instruments to assess one or more essential reading components at one or more grade levels K - 3:

Auditory Analysis Test
An Observation Survey of Early Literacy Achievement
Qualitative Reading Inventory (QRI)
Roswell-Chall Auditory Blending
Woodcock-Johnson III Tests of Cognitive Abilities

The results of the Assessment Committee are summarized in tables in the final section of this report. The tables are organized by grade levels (kindergarten through grade 3) and purpose of reading assessment (e.g., screening, diagnosis, progress monitoring, outcome). Each “Grade Level by Type of Measure” table lists the specific measures that were found to be sufficient for use to assess the five essential components of reading instruction specified in Reading First. A narrative summary of the findings for each grade level is provided below:

Kindergarten. In Kindergarten, the number and type of assessment instruments judged to have sufficient evidence to assess the five essential components as defined in the Reading First legislation were:

- Screening instruments: 5 to assess phonological awareness, 6 to assess phonics, 2 to assess vocabulary, and 1 to assess reading comprehension.
-

- Diagnostic instruments: 7 to assess phonological awareness, 6 to assess phonics, 4 to assess vocabulary, and 1 to assess reading comprehension.
- Progress monitoring instruments: 3 to assess phonological awareness, 3 to assess phonics, and 1 to assess vocabulary.
- Outcome instruments: 10 to assess phonological awareness, 7 to assess phonics, 6 to assess vocabulary, and 1 to assess reading comprehension.

Grade 1. In first grade, the number and type of assessment instruments judged to have sufficient evidence to assess the five essential components were:

- Screening instruments: 5 to assess phonological awareness, 9 to assess phonics, 3 to assess vocabulary, and 4 to assess reading comprehension.
- Diagnostic instruments: 7 to assess phonological awareness, phonics, vocabulary, and reading comprehension, and 1 to assess fluency.
- Progress monitoring instruments: 3 to assess phonological awareness, 5 to assess phonics, 3 to assess fluency, and one to assess vocabulary and reading comprehension.
- Outcome instruments: 9 to assess phonological awareness, 10 to assess phonics, 4 to assess fluency, 10 to assess vocabulary, and 8 to assess reading comprehension.

Grade 2. In second grade, the number and type of assessment instruments judged to have sufficient evidence to assess the five essential components were:

- Screening instruments: 7 to assess phonics, 4 to assess to fluency, 2 to assess vocabulary, and 4 to assess reading comprehension.
 - Diagnostic instruments: 1 to assess phonological awareness, 6 to assess phonics, 1 to assess fluency, and 8 to assess vocabulary and reading comprehension.
 - Progress monitoring instruments: 4 to assess phonics, 2 to assess fluency, and 1 to assess vocabulary, and 2 to assess reading comprehension.
 - Outcome instruments: 1 to assess phonological awareness, 11 to assess phonics, 4 to assess fluency, and 11 to assess vocabulary and reading comprehension.
-

Grade 3. In third grade, the number and type of assessment instruments judged to have sufficient evidence to assess the five essential components were:

- Screening instruments: 1 to assess phonics, 4 to assess to fluency, and 2 to assess vocabulary.
- Diagnostic instruments: 1 to assess phonological awareness, phonics, and fluency, and 6 to assess vocabulary and reading comprehension.
- Progress monitoring instruments: 1 to assess phonics, 3 to assess fluency, and 1 to assess vocabulary and reading comprehension.
- Outcome instruments: 1 to assess phonological awareness and phonics, 4 to assess fluency, 9 to assess vocabulary and 7 to assess reading comprehension.

Discussion

The Assessment Committee analyzed 29 reading assessment instruments designed to assess one or more of the five essential components identified in the Reading First legislation. Of this set of assessment instruments, a descriptive analysis reveals that in general, there are more assessment instruments judged as sufficient for use as outcome measures than for progress monitoring reading performance in K-3. In addition, more diagnostic and outcome measures were found to have sufficient evidence in vocabulary and reading comprehension in grades 2 and 3 than in kindergarten and first grade. Of the five essential components of reading instruction, the number of assessment instruments found to have sufficient evidence for use to assess reading fluency was noticeably limited (e.g., a maximum of four outcomes measures in kindergarten and grade 1).

Limitations of Final Report

As with any substantial effort that involves breaking new ground, “blank spots” and “blind spots” (Wagner, 1993) are inevitable. This Final Report is no exception. If the Assessment Committee was to repeat this entire effort (God forbid!), the substantive analysis and logic rules

would probably remain the same. There is no doubt, however, that the Committee would insist on more time. The tyranny of time recognized, there are several limitations to this effort that require comment.

Because of time constraints and the ambitious scope of work, the Assessment Committee was not able to determine the sufficiency of evidence for any of Spanish reading assessment instruments. However, the documentation process and coding forms that the Assessment Committee developed (see Appendices) could readily be used to carry out this task. Extending the process of review and analysis to the following Spanish assessment instruments is important:

- Test de Vocabulario en Imagenes Peabody, TVIP
- The Observation Survey (Spanish Equivalent)
- Developmental Reading Assessment (Spanish Equivalent)
- Woodcock-Munoz Language Survey
- Aprenda: La Prueba de Logros en Espanol- 2nd Ed
- Pre-Las 2000
- Spanish Reading Comprehension Test
- La Prueba de Realizacion, Segunda Edicion*
- Spanish Assessment of Basic English- Second Ed
- Texas Primary Reading Inventory – Spanish Version

In addition, the Assessment Committee did not conduct a descriptive analyses of the entire set of 29 assessment instruments around the character of the technical adequacy of the different types of measures (e.g., progress monitoring, diagnostic, outcome, screening). While such an analysis would prove interesting and offer what is likely to be a provocative perspective on the “sufficiency” of evidence for the different types of measures (e.g., the extent to which screening measures provided sufficient evidence on predictive validity, specificity, sensitivity; the extent to which outcome measures addressed features of grade level improvement or significant gain), this analysis was beyond both the Committee’s scope of work and timeline.

Because of the time constraints for completing the scope of work, the Assessment Committee did not explore a process that included establishing a collaborative working relationship with the publishers and developers of assessment instruments. Such an effort is likely to result in a more thorough documentation of the technical adequacy of assessment instruments in general. Moreover, this collaboration is likely to be a recursive and instructive process for all parties involved in the collaboration, which only benefits the improvement of assessment instruments.

The most conspicuous limitation to the Assessment Committee's effort is the limited number of assessment instruments analyzed, coded, review, and evaluated. Clearly, more assessment instruments must be reviewed and evaluated. More importantly, an independent infrastructure and dispassionate process must be established to build a formal "registry" of documentation and codification in both process and product for evidence-based assessment instruments and instructional tools and programs.

Finally, the work captured in the Final Report must be continued in many ways that include, for example:

- (a) verifying the judgments of sufficiency made for each of the assessment instruments;
 - (b) replicating the entire codification and documentation process to determine its integrity, viability, fiscal feasibility, efficiency, and validity;
 - (c) expanding the review and analysis process to other assessment instruments in reading, mathematics, and content areas;
 - (d) exploring a more efficient but equally objective adjudication process;
 - (e) determining the feasibility of building a collaborative coalition between publishers and developers of assessment instruments and professional educational organizations for establishing a similar review process;
 - (f) examining the technical, fiscal, professional, and scientific requirements of establishing an infrastructure for this kind of effort and focus that will inform schools,
-

stakeholders, administrators, policymakers, and others about the trustworthiness of educational tools, technology, and print materials; and
(g) evaluating the utility of the product for consumers.