Technical Report # 1110

Progress Monitoring Instrument Development: Silent Reading Fluency, Vocabulary, and Reading Comprehension

Joseph F. T. Nese

Daniel Anderson

Kyle Hoelscher

Gerald Tindal

Julie Alonzo

University of Oregon

Published by

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

Note: Funds for this data set used to generate this report come from a federal grant awarded to the UO from the Office of Special Education Programs, U.S. Department of Education: Steppingstones of Technology Innovation for Children with Disabilities (PR/Award # H327A090005 funded from August 2009 – August 2011).

Copyright © 2011. Behavioral Research and Teaching. All rights reserved. This publication, or parts thereof, may not be used or reproduced in any manner without written permission.

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

Abstract

Curriculum-based measurement (CBM) is designed to measure students' academic status and growth so the effectiveness of instruction may be evaluated. In the most popular forms of reading CBM, the student's oral reading fluency is assessed. This behavior is difficult to sample in a computer-based format, a limitation that may be a function of the lack of available measures for silent reading fluency, vocabulary and comprehension. In this technical report, we describe the development of three specific CBM reading measures designed for a computer format: silent reading fluency, vocabulary, and reading comprehension.

Progress Monitoring Instrument Development:

Silent Reading Fluency, Vocabulary, and Reading Comprehension

Curriculum-based measurement (CBM) is designed to measure students' academic status and growth so the effectiveness of instruction may be evaluated (Deno, Marston, & Tindal, 1985; Fuchs, 2004; Good & Jefferson, 1998; Tindal et al., 1985). In practice, alternate CBM forms representative of grade-level outcomes are developed, administered and scored in a standardized manner, and the results then used to document performance and progress.

CBM has established reliability and validity for decision-making (Deno, 1985). Numerous research studies, dating back nearly 30 years, have demonstrated the usefulness of CBM for monitoring the academic progress of students in the basic skill area of oral reading (ORF) (Fuchs, Deno, & Mirkin, 1984; Marston, Deno, & Tindal, 1983; Marston & Magnusson, 1985). Research on ORF has often appeared in the professional literature over the past three decades. As Foegen, Espin, Allinder, and Markell (2001) write: "the number of words read correctly has been shown repeatedly to be a reliable measure (with test-retest reliability ranging from .93 and .99 and interjudge reliability between .96 and .99) and a valid measure (with validity coefficients between words read and criterion measure ranging from .54 and .92)" (p. 227). This statement summarizes the work of Deno, Marston, Mirkin, Lowry, Sindelar, and Jenkins (1982); Fuchs, Deno, and Marston (1983); Jenkins and Jewell (1993); and Tindal and Marston (1996).

In special education, vocabulary measures have been studied only recently. For example, Espin and Deno, (1994-1995) successfully used vocabulary measures to predict content study task performance in a generalized way that was not limited to specific content areas. In an extended replication of this study, Espin and Foegen (1996) investigated vocabulary measures along with maze tasks and oral reading fluency measures and found that vocabulary explained

most of the variance on all three of these outcomes on content tasks. Nese, Park, Alonzo, and Tindal (in press) likewise found that the easyCBM vocabulary measure accounted for more unique variance in state reading scores than did ORF or comprehension measures. The authors also found the vocabulary and comprehension measures were better predictors of state reading test scores than ORF, indicating that perhaps other reading measures may be better indicators of reading proficiency in the upper elementary grades. As has been suggested by previous research (Cain & Oakhill, 1999; Yovanoff, Duesbery, Alonzo, & Tindal, 2005), beyond third grade learning to read fluently and accurately becomes less important than reading to learn, which may depend more on students' vocabulary and comprehension skills.

Much of the potential of technology has been missed in the development of curriculum-based measures, particularly in the field of computer-based testing (CBT). In particular, most curriculum-based measurements (CBMs) have not yet been computerized (in administration). In part, this limitation may be a function of the behavior being sampled. In the most popular forms of reading CBM, the student's oral reading fluency is assessed. This behavior is difficult to sample in a computer-based format. This limitation also may be a function of the lack of available measures of vocabulary and comprehension. In this technical report, we describe the development of three specific CBM reading measures designed for a computer format: silent reading fluency, vocabulary, and reading comprehension.

Instrument Development Process

Measures were developed by a team of three researchers and two teachers. The three researchers included two with master's degrees in education and one doctoral student in education. The teachers were both elementary school general education teachers working in a large suburban district in Oregon. The team wrote five types of measures designed to target three areas of reading – silent reading fluency (sentences and maze measures), vocabulary (context-

embedded vocabulary maze and sentence measures), and comprehension. Each measure was written with varying numbers of items and forms. Two additional researchers – one doctoral student in education and one post-doctoral research fellow – joined the item writing team to review all items and forms for errors (e.g., format and grammatical) and bias (e.g., cultural, religious, and geographical). A computer programmer developed the online delivery system and user interfaces for each of the measures.

Item Review

The item review team (consisting of five researchers) conducted group reviews of all measures. The team met in groups of three to ensure all items had proper item mechanics, contained unbiased language, and met the technical specifications described above. For vocabulary maze and vocabulary sentence forms, all distracters were written during the review process.

Silent Reading Fluency Sentences

The item writing team developed 20 silent reading fluency sentences (SRF-S) forms with five items per form in each of grades 3, 4, and 5. Items consisted of a sentence and question pair. The instructions read: "A sentence will be presented for you to read. When you are done reading it, click on the *Done* button. A new screen will be presented with a question about the sentence. Select the correct option. The sentences and questions will continue. Keep going until you see the cartoon mouse with a balloon. When you are ready to read click *Start*. When you are finished reading click *Done*."

During administration, the sentence first appeared on the student's screen without the question (e.g., "The boys liked to eat ice cream for dessert."). After reading the sentence, the student clicked a button to indicate having finished reading. The question about the sentence then immediately appeared (e.g., "What did the boys like to eat for dessert?"). Each question was

presented in a multiple-choice format with three options – a correct response and two distracters. All distracters were purposely selected to be distant, so that the item itself would be quite easy. The items were intended to be easy because the purpose of the questions following the sentences was not to assess comprehension, but to verify that the sentence was read or read correctly. In instances where the student did not read the sentence and simply clicked the "done" button, the question would capture the student's guessing either by recording an unreasonably fast time, and/or an incorrect response to the question.

The team used high frequency, grade-appropriate words and simple grammar for the sentences and questions. The sentences contained between 4 and 14 words. All questions were strictly literal with response options ranging from one to three words. Among the five items per form, three or four of the questions addressed the first half of the sentence, while one or two of the questions addressed the second half of the sentence. All response options were of the same word type (noun, verb, adjective, etc.) and had parallel grammar structure.

The computer captured the time elapsed from when the question *appeared on the screen* to when the student finished reading and clicked *Done*. A word reading fluency estimate was computed by dividing the number of words in the sentence by the elapsed time it took the student to read the sentence. The resulting value was then converted to a "words read per minute" scale by being multiplied by 60. The computer interface automatically recorded data on all student responses and whether the responses were correct or incorrect.

Silent Reading Fluency Maze

The item writing team developed 20 silent reading fluency maze (SRF-M) forms in each of Grades 3, 4, and 5. A form consisted of a word reading passage (approximately 100-120 words) with 7 words chosen for omission. The omission of words created an "option point," where the student was required to select the most appropriate word to complete the sentence and

the story. The directions read: “A short story will be presented. It will have missing words. Read the sentences up to the missing word and then click to select a word that correctly finishes *the sentence and the story*. Continue reading through the story and select words to correctly finish *the sentence and the story* until you come to the end. When you are ready to read click *Start*. When you are finished reading click *Done*.”

After reading the passage, the student would click a button to indicate having completed reading the passage. The interface allowed the student to select only the next answer choice, not any subsequent choices, to prevent skipping ahead. The computer captured the time elapsed from when the student clicked *Start* to when the first answer choice was selected, and the time elapsed from when the student selected an answer choice to when the student selected the next answer choice. The computer recorded the number of words in the passage for the elapsed time event, including the word being selected and the distracter option. A fluency estimate was computed by dividing the number of words within each elapsed time event by the elapsed time, and multiplying by 60 to convert it to words read per minute. The computer interface automatically recorded data on all student responses and whether the responses were correct or incorrect.

Each passage was approximately 100 words of fictional narrative text within one grade-level of the targeted grade, as measured by the Flesch-Kincaid readability calculator. The first option point was placed at least 10 words from the beginning of the story, and subsequent option points were spaced evenly apart over the remainder of the story. Omitted words were of varied word types (noun, verb, adjective, particle, etc.). Each option point had two response options. All distracters were purposely selected to be distant, as we were more interested in measuring the time between responses than the accuracy of responses. For this reason, the distracter was not necessarily the same word type as the correct response option. Distracters were made to be easy so that they would require as little thinking time as possible. The answer choices were meant to

serve only to capture students' reading time and to keep students focused, with minimal disruption of fluent reading.

Vocabulary Maze

The item writing team developed 20 vocabulary maze (VM) forms in each of Grades 3, 4, and 5. Similar to the SRF-M forms, the VM forms had option points where students were prompted to select the most appropriate word given the context of the passage. Each form contained 12 choice points with three response options. The directions read, "A short story will be presented. It will have missing words. Read the sentences up to the missing word and then click to select a word that correctly finishes *the sentence and the story*. Continue reading through the story and select words to correctly finish *the sentence and the story* until you come to the end. When you are ready to read click *Start*. When you are finished reading click *Done*." During administration, the passage would appear for students to read, select the most appropriate word at each option point, and then click a button to indicate having finished reading. The computer captured only students' responses and whether those responses were correct; the computer did not record the time it took for the student to read the passage.

The VM forms were longer than the SRF-M forms, with each passage being approximately 200 words. All forms contained fictional narrative text with grade-level vocabulary words embedded in the text. The first option point was placed at least 10 words from the beginning of the story, and subsequent option points were spaced apart by at least 4 words and distributed evenly throughout the story. Vocabulary words were of varying word types (noun, verb, adjective, adverb, etc.) and were drawn from the Marzano, Kendall, and Paynter (2008) list. This list was further used to select appropriate distracters. Where possible, distracters were of the same word type as the correct option, with one distracter closer in meaning and one

distracter more distant. Words used for the correct option were not repeated across any of the forms, though some were used as distracters on other forms.

Vocabulary Sentences

The item writing team developed 20 vocabulary sentence (VS) forms in each of Grades 3, 4, and 5. Each item contained one sentence with one omitted word. The student was prompted to select the most appropriate word given the context of the sentence. Each form contained 13 items with three response options for each item. The directions read: "A series of sentences will be presented with missing words. Read the sentences up to the missing word and then click to select a word that correctly finishes the sentence. Continue reading each sentence and select words to correctly finish them until you come to the last sentence. When you are ready to read click *Start*. When you are finished reading click *Done*." The computer captured only students' responses and whether those responses were correct; the computer did not record the time it took for students to read the passage.

All sentences were written with varied sentence lengths on each form, ranging from approximately 6-18 words per sentence. All targeted vocabulary words and distracters were drawn from the Marzano, Kendall, and Paynter (2008) list at the appropriate grade level. Where possible, distracters were of the same word type as the correct option, with one distracter closer in meaning and one distracter more distant. Words used for the correct option were not repeated across any of the vocabulary sentence or VM forms, though some were used as distracters on other forms.

Reading Comprehension

The item writing team developed three reading comprehension (RC) forms in each of Grades 3, 4, and 5. Each form consisted of a fictional narrative passage of approximately 500 words with seven multiple-choice questions. During administration, students read the passage

from the form administered, then responded to each multiple-choice question. The computer captured only students' responses and whether those responses were correct; the computer did not record the time it took for students to read the passage.

Passages were written to be within one grade level of the target grade for each form, as measured by the Flesch-Kincaid readability calculator. Among the seven questions were four literal and three inferential questions. Literal questions addressed events, details, and sequence of events. Inferential questions addressed prediction, main idea, character motivations, symbolism, and ideas not explicitly stated. Each question had three response options: the correct answer, one close distracter, and one distant distracter. As a rule all distracters were written to have parallel grammar structure to the correct response option.

Conclusion

The development of reading measures designed for computer administration is complex. Information obtained from traditional paper-pencil based reading CBMs is valuable, but administration can be time consuming, as students must be administered each measure individually. The measures reported here are experimental in nature, but they represent a potential step forward in the efficiency of reading CBM. Moving the measures to a computer interface means a whole class of students could potentially take the measures simultaneously in a computer-lab setting. However, moving the measures to a computer interface also required modifications to the paper-pencil based assessments. It remains to be seen if data obtained from the measures reported here will be comparable to data obtained from more traditional oral reading CBM measures. It is likely that the new measures will require a shift in the way we think of the data from CBMs (i.e., silent versus oral reading fluency). But the potential time that computer administration could save educators may make the shift in thinking well worth the effort.

References

Cain, K. & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing, 11,* 489-503.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219-232.

Deno, S. L., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., & Jenkins, J. (1982). *The use of standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study*. Minneapolis, MN, Institute for Research on Learning Disabilities-University of Minnesota.

Deno, S., Marston, D., & Tindal, G. (1985). Direct and frequent curriculum-based measurement: An alternative for educational decision making. *Special Services in the Schools, 2,* 5-27.

Espin, C. A., & Deno, S. L. (1994-1995). Curriculum-based measures for secondary students: Utility and task specificity of text-based reading and vocabulary measures for predicting performance on content-area tasks. *Diagnostique, 20*(1-4), 121-142.

Espin, C. A., & Foegen, A. (1996). Validity of general outcome measures for predicting secondary students' performance on content area tasks. *Exceptional Children, 62,* 497-514.

Foegen, A., Espin, C. A., Allinder, R. M., & Markell, M. A. (2001). Translating research into practice: Preservice teachers' beliefs about curriculum-based measurement. *The Journal of Special Education, 34,* 226-236.

Fuchs, L. S., Deno, S. L., & Marston, D. (1983). Improving the reliability of curriculum based measures of academic skills for psycho-educational decision-making. *Diagnostique, 8,* 135-149.

Fuchs, L.S., Deno, S.L., & Mirkin, P.K. (1984). The effects of frequent curriculum based measurement and evaluation on pedagogy, student achievement and student awareness of learning. *American Educational Research Journal, 21,* 449-460.

Fuchs, L. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33,* 188-192.

Good, R., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61-88). New York: Guilford Press.

Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: reading aloud and maze. *Exceptional Children,* 59, 421-432.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children, 52,* 266-276.

Marston, D., Deno, S., & Tindal, G. (1983). *A comparison of standardized achievement tests and direct measurement techniques in measuring pupil progress* (Research Rep. No. 126). Minneapolis, MN: University of Minnesota, Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 236 198).

Marzano, R. J., Kendall, J. S., Paynter, D. E. (2008). Appendix: A list of essential words by grade level. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about vocabulary instruction* (pp. 127-202). Newark, DE: International Reading Association.

Nese, J., Anderson, D., Lai, C. F., & Tindal, G. (2009). *Within-year reading growth by student group: An application of hierarchical linear modeling with curriculum-based measurement*. Unpublished manuscript.

Nese, J. F. T., Park, B. J., Alonzo, J., & Tindal, G. (in press). Applied curriculum-based measurement as a predictor of high-stakes assessment: Implications for researchers and teachers. *Elementary School Journal.*

Tindal, G., Fuchs, L., Fuchs, D., Shinn, M., Deno, S., & Germann, G. (1985). Empirical validation of criterion-referenced tests. *Educational Researcher, 78,* 203-209.

Tindal, G., & Marston, D (1996). Technical adequacy of alternative reading measures as performance assessments. *Exceptionality, 6,* 201-230.

Tindal, G., Nese, J., & Alonzo, J. (2009). *Criterion-related evidence using easyCBM® reading measures and student demographics to predict state test performance in grades 3 – 8.* (Technical Report No. 0910). Eugene, OR: Behavioral Research and Teaching: University of Oregon.

Yovanoff, P., Duesbery, L., Alonzo, J. Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24*, 4-12.