School Psychology Review
Volume 21, No. 3, 1992, pp. 459-479

CURRICULUM-BASED MEASUREMENT OF ORAL READING FLUENCY: A CONFIRMATORY ANALYSIS OF ITS RELATION TO READING

Mark R. Shinn
University of Oregon

Roland H. Good III
University of Oregon

Nancy Knutson
Oregon Social Learning Center

W. David Tilly III
Heartland Area Agency, Des Moines, IA

Vicki L. Collins
University of Oregon

Abstract: Reading traditionally is characterized as having two major components, decoding and comprehension. Published reading tests are created using these two components. Reading fluency, a combination of reading speed and accuracy, typically is not measured. Attention to reading fluency has increased through the emerging literature on Curriculum-Based Measurement (CBM), which employs standardized oral reading tests derived from basal readers to make decisions about students' general reading skills. Despite a series of published validation studies, questions about what CBM oral reading fluency measures persist. This study examined the relation of CBM oral reading fluency to the reading process from a theoretical perspective. Reading models were tested using confirmatory factor analysis procedures with 114 third- and 124 fifth-grade students. Subjects were tested on tasks requiring decoding of phonetically regular words and regular nonsense words, literal comprehension, inferential comprehension, cloze items, written retell, and CBM oral reading fluency. For third graders, a unitary model of reading was validated with all measures contributing significantly. For fifth graders, a two-factor model was validated paralleling current conceptions of reading measurement. Regardless of the factor model employed, CBM oral reading fluency provided a good index of reading proficiency, including comprehension.

Although not without controversy (Goodman, 1986; Smith, 1973, 1975), the process of reading typically is described as being composed of two separate, but interdependent elements, decoding and comprehension (e.g., Heilman, 1967). Decoding is defined as "translating printed words into a representation similar to oral language" (Carnine & Silbert, 1979, p. 30) or as the "ability to pronounce the printed word" (Golinkoff, 1975, p. 633). Comprehension is defined as the understanding of the words represented (Carnine & Silbert, 1979) or as the extraction of information from text (Gibson & Levin, 1975).

It is not surprising, therefore, to observe that assessment devices and reviews of reading tests have employed this conception of reading. The *Woodcock Reading Mastery Tests-Revised* (Woodcock, 1987), for example, includes a series of decoding tasks as part of its basic skills cluster. The decoding tasks require respondents to read words in isolation (Word Identification) and say "nonsense words" aloud (Word Attack). Comprehension is assessed through the use of tasks

This research was supported in part by Grant No. 8029D80051-90 from the U.S. Department of Education, Special Education Programs to provide leadership training in curriculum-based assessment. The views expressed within this article are not necessarily those of the U.S. Department of Education.

The authors with to thank Mary Goodwin, Victor Nolet, and Todd Robicheaux for their assistance in data collection.

Address all correspondence and questions about this article to Dr. Mark R. Shinn, School Psychology Program, Division of Special Education and Rehabilitation, University of Oregon, Eugene, OR 97403, 503-346-2144.

that require students to provide antonyms, synonyms, and complete analogies (Word Comprehension) or complete a modified cloze task (Passage Comprehension) where students read a passage and supply missing words.

Similarly, the *Kaufman Test of Educational Achievement* (KTEA; Kaufman & Kaufman, 1985) includes decoding and comprehension subtests. In the former, respondents identify letters and read phonetically regular and irregular words. In the latter, respondents read sentences and respond with gestures or orally; respondents also read material and answer literal and inferential comprehension questions. Reviews of reading tests also have been structured around the constructs of decoding and comprehension. For example, Ysseldyke and Marston (1982) dichotomized the skills contained in 12 published reading tests into these two components. Similar descriptions of reading tests have been made by Salvia and Ysseldyke (1988) and Wallace and Larsen (1979).

The Relationship between Decoding and Comprehension

The relationship between decoding and comprehension in the professional literature appears to have settled into one in which adequate decoding skills are viewed as a necessary prerequisite, though not sufficient, condition for comprehension to occur (Golinkoff, 1975; Perfetti & Hogaboam, 1975). Explicit in this conception of the relationship between decoding and comprehension is a directional, causal model. Decoding affects comprehension; comprehension does not affect decoding.

This simple model of reading has been elaborated upon, however, through the inclusion of the concept of reading automaticity or *fluency*. Reading fluency has been defined as the speed and accuracy with which a student reads words. Based upon a review of the reading research, Samuels (1976, 1981) and Laberge and Samuels (1974) expounded a model of reading emphasizing fluent, automatic decoding as a necessary mediator for comprehension to occur. In this model, an individual reader has only a limited amount of attention to devote to the reading process. If the reader allocates that attention to the decoding process (i.e., breaking down the word for pronunciation), then limited or no attention can be devoted to determining the meaning of the word or the textual material. For Samuels and others (e.g., Jager-Adams, 1990), the competent reader is able to decode automatically without the services of attention and thus is able to attend to processing meaning. The importance of rapid, automatic decoding also is emphasized by researchers with different philosophical orientations. For example, Carnine and Silbert (1979), proponents of a behaviorally based direct instruction approach, also identify rapid decoding as a fundamental skill in reading comprehension.

Is Fluency Just Rapid Decoding?

Contemporary models of reading, whether from an information processing model (Samuels, 1981) or from a direct instruction model (Carnine & Silbert, 1979), impart a significant role for reading fluency. However, reading fluency appears to be treated synonymously with rapid decoding, or at best, as a subcomponent of decoding. An examination of reading textbooks by this study's authors failed to find one citation that distinguished between decoding and fluency. Reading fluency and rapid decoding also typically are treated synonymously in textbooks on assessment. For example, Howell and Kaplan (1980) in their text on basic skills assessment recommended that students' decoding skills be examined by having them read passages and word lists aloud from the books at their grade level. They also described numerous methods of assessing comprehension. As might be expected, reading fluency is not listed as a method for assessing comprehension.

One outcome of the conception of reading fluency as rapid decoding is that the importance of assessing it would be useful only with respect to the inferences one wished to make regarding a student's decoding skills. Presumably, one could use such assessment information to decide: (a) whether a student had sufficient decoding skills, and relatedly, (b) what decoding skills the student possessed. This information would be useful primarily for instructional planning. Given the primary importance of comprehension in the reading process, caution would be indicated when making other decisions about reading skills (e.g., problem identification, problem certification, monitoring of student reading achievement) based only on decoding skills. In contrast, if oral reading fluency provided an adequate measure of comprehension skills, its robustness in educational decision making would be supported.

Curriculum-based measurement (CBM; Deno, 1986, 1986; Fuchs & Fuchs, 1984, 1986; Shinn, 1989) is a set of testing strategies that places a high reliance on the measurement of a student's oral reading fluency to make a variety of decisions about students' reading skills. Using standardized procedures, students typically read aloud from a level of their basal reading series for a set of repeated, 1-minute timings (Deno, Mirkin, & Wesson, 1984). Although a rich qualitative sample of reading is obtained, primary decision making is made quantitatively by counting the number of words read correctly per minute. This metric has been used to make special education problem identification (screening) and problem certification (eligibility) decisions (Germann & Tindal, 1985; Marston & Magnusson, 1988; Shinn, Tindal, & Spira, 1987; Shinn, Tindal, & Stein, 1988), write IEP objectives and monitor the effectiveness of reading instructional interventions (Fuchs, 1989; Fuchs, Deno, & Mirkin, 1984; Fuchs, Fuchs, & Hamlett, 1989; Fuchs & Shinn, 1989), and determine when reading problems have been resolved (Allen, 1989; Rodden-Nord, Shinn, & Good, 1992).

Given CBM's high reliance on reading fluency for making these important decisions, it is critical that the aspects of reading to which reading fluency is related be determined. For more than 12 years, studies examining the validity of CBM reading fluency have been conducted. The findings have provided strong support for oral reading fluency as a reliable and valid measure of a student's general reading skill, including reading comprehension (Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988; Fuchs et al., 1983; Marston, 1989; Marston & Deno, 1982; Tindal et al., 1983). Numerous studies have examined CBM's relations with published norm-referenced reading achievement tests such as the *Stanford Achievement Test*, the *Stanford Diagnostic Reading Test*, the *Metropolitan Achievement Test*, the *PIAT*, and the *Woodcock Reading Mastery Tests*. Other evidence of criterion-related validity includes investigations of the relations between CBM reading fluency and criterion-referenced basal reading mastery tests (Fuchs et al., 1983; Tindal et al., 1985), and teacher judgment (Fuchs & Deno, 1981; Tilly, 1989). Typical correlation coefficients in these studies have ranged from .60 to .90 with most correlations around .80.

Criterion-related validity is but one way to judge the validity of a measure (Messick, 1989). Other aspects of construct validity are important. CBM oral reading fluency has demonstrated discriminant validity (Deno, Marston, Shinn, & Tindal, 1983), longitudinal change (Marston et al., 1981), sensitivity to changes in reading programs (Deno, 1985, 1986), and treatment validity (Fuchs & Fuchs, 1986; Fuchs et al., 1989; Fuchs, Fuchs, & Tindal, 1986).

These investigations raise important questions about prevailing conceptions of reading fluency as just rapid decoding. Given the validated research outcomes, it may be that fluency is more than just "rapid decoding." To test this hypothesis directly, Fuchs et al. (1988) compared reading fluency with commonly used measures of comprehension, including oral and written retells, oral and written cloze, question-answering strategies, and the reading subtests of the *Stanford Achievement Test* (SAT) a published norm-referenced achievement test. Results indicated that CBM reading fluency was correlated at a significantly higher level (r = .91) with the SAT than were any of the other, more accepted, comprehension methods. Results also were examined within a convergent and divergent validity approach. One of the subtests of the SAT was a decoding task (Word Study). The other subtest was Reading Comprehension. CBM reading fluency was significantly more highly correlated to the Comprehension subtest (r = .92) than to the Word Study subtest (r = .81), $t(32) = 4.26, p < .01$.

Investigating Fluency in Relation to Reading Models

Taken together, the research on CBM reading fluency provides sufficient evidence for validating oral reading as an adequate measure of reading skill. However, issues of its acceptance as a measure of reading remain (Fuchs et al., 1988; Potter & Wamre, 1990). Potter and Wamre (1990) hypothesized that acceptance has been hindered by researchers' focusing on proving that the measures work rather than looking at "why they work." They suggested the need to examine the role of reading fluency from a more theoretical perspective. Of particular interest was the construct represented by oral reading fluency measures: general reading proficiency, decoding skill, comprehension, or a distinct fluency component of reading proficiency.

Very few studies have examined explicitly the factor structure of reading. In a factor analysis of reading and verbal processing skills, De Soto and De Soto (1983) found that fluency in coding meaningful and nonmeaningful words was a distinct factor of primary importance for reading comprehension. However, this study did not consider the possibility of decoding or comprehension as distinct factors.

Two investigations of the theoretical contribution of oral reading fluency to a model of reading have been undertaken. In the first, Lomax (1983) hypothesized a causal model of the component processes of reading comprehension. Reading comprehension was thought to be affected by reading rate and word recognition, reading rate was thought to be affected by word recognition, and word recognition was thought to be affected by a phonological component. Lomax (1983) obtained a final model with three reading constructs: comprehension, rate, and word recognition. A reciprocal relation was found between word recognition and reading rate, but "no empirical evidence was found for the existence of a direct causal relationship between reading rate and comprehension" (p. 37).

However, serious methodological concerns limit the interpretation of these results. First, the study did not include specific tests of the numbers of factors involved. In particular, word recognition and comprehension were so highly related (.98) that it may not be reasonable to consider them as distinct constructs. Second, subjects in the study ranged in age from 6 to 11 years, without providing the possibility of different factor structures at different age levels. Thus, if there were developmental differences in the component process of reading as suggested by some (e.g., Chall, 1983; Stanovich, 1986), the actual factor structure would be obscured. Third, the effect of age was not controlled in the modeling process. Thus, although word recognition and comprehension were highly related, shoe size or height also may have functioned well in the model.

In the second study, Collins (1989) investigated the contribution of oral reading fluency to reading proficiency using confirmatory factor analysis to examine the constructs of decoding, comprehension, and fluency. The three constructs were operationalized using CBM reading tests, a basal mastery end-of-unit criterion-referenced test, and a published, norm-referenced reading test. Subjects were 58 second-grade, general education students assigned to two groups by the level of the basal reading series in which they received instruction. In the Collins (1989) study, a three-factor model of reading proficiency with reading fluency as a distinct component was superior to the alternative models. However, the results of this study also must be interpreted cautiously. In particular, the sample size examined by Collins (1989) was small, resulting in unstable factor solutions, and the measures useo to operationalize the decoding and comprehension constructs were limited psychometrically.

The purpose of this study was to examine the role of CBM reading fluency from a more theoretical perspective. Building on the exploratory work of Collins, a variety of psychometrically sound measures of decoding and comprehension were investigated using a larger sample of elementary-aged students. In addition, potential developmental differences in reading models were examined by including students at two different grades.

METHOD

Subjects

A total of 238 third-grade ($N = 114$) and fifth-grade ($N = 124$) students served as subjects. Of the sample, 49% were female (57 third-graders and 70 fifth-graders) and 51% were male (57 third-graders and 64 fifth-graders). Subjects were selected from 13 elementary schools located in a predominantly white public school district in a mid-size northwestern city. The third-grade students attended 9 of these schools while the fifth-grade students attended 10. A total of 95% of the subjects received instruction in general education; the remaining 5% received special education services for less than 50% of the school day.

Subjects were recruited via a four-step process. First, permission to conduct this study was obtained at the district administrative level. Second, principals of elementary schools were provided with a written description of the study's goals, reading measures, and testing procedures. Third, the third- and fifth-grade teachers in the buildings were contacted. Fourth, if teachers gave permission for students in their class to serve as subjects, letters requesting consent for participation in the study were sent to the students' parents.

A total of 714 letters (322 and 392 for parents of third- and fifth-grade students, respectively) were mailed and 45% of these letters were returned. Of the returned letters, 6% were past the deadline or indicated the child would be absent on the testing date. Of the usable returned letters (39% of the letters mailed), almost all parents (86%) granted permission. Complete data were obtained on 238 students. Thirty students were absent from one or both testing sessions.

MATERIALS

All subjects were administered eight reading measures: two conceptualized traditionally as measuring decoding skills, four conceptualized traditionally as measuring comprehension skills, and two measuring oral reading fluency. Measures were administered individually and in a group of same-grade peers.

Individual Measures

Materials for the individual testing were organized into two packets, one for the examiner and one for the student. The examiner materials consisted of two grade-level CBM reading passages with the cumulative number of words written in the margin, a list of the phonetically regular words, and a list and pronunciation guide for the phonetically regular nonsense words. The student packet consisted of unnumbered copies of the CBM passages and the word lists.

CBM reading passages. Students were given two passages of approximately 250 words from their grade-level basal reader to read aloud for 1 minute each. To develop the CBM passages, four reading passages (two per grade) were selected from the Harcourt-Brace-Jovanovich basal reading series, the district's most frequently used text. Passages were chosen by: (a) randomly selecting eight stories from the two basal texts at each grade level; (b) examining each passage's readability using the Spache, Dale-Chall, and Fry readability formulas; and (c) selecting one passage with difficulty approximating beginning grade level and one passage with approximately ending grade level difficulty.

Standardized directions instructed students to read aloud from the beginning of each passage using their "best reading skills" (Deno et al., 1984). At the end of 1 minute, students were told to stop reading. The measures of oral reading fluency were the number of words read correctly in 1 minute for each passage. Words pronounced correctly or mispronounced initially, but self-corrected within 3 seconds were scored as correctly read. Repetition or insertion of a word was not counted as an error. Omissions, substitutions, and mispronunciations of words were scored as errors. If students hesitated or struggled with a word for 3 seconds, they were told the word and it was counted as incorrect; other errors were not corrected by the examiner. Oral reading fluency measured using these procedures has demonstrated very acceptable technical characteristics (Deno, 1985; Deno et al., 1982; Fuchs et al., 1988; Marston, 1989).

Phonetically regular words. Students were presented with a typed list of 45 phonetically regular words taken in total from the *Test of Written Spelling* (TWS; Larsen & Hammill, 1976). The TWS was not used as a measure of spelling, but rather as a source of familiar, phonetically regular words. The words were arranged hierarchically in order of difficulty and retyped five across the page. Students were instructed to read aloud as many words as they could in 1 minute. If students finished the entire list with time remaining, they were instructed to begin reading again from the top of the page. Student performance was scored using the methods described previously, with one difference. If the student hesitated or struggled with a word for 3 seconds, the examiner pointed to the next word and instructed the student to go on without pronouncing the missed word.

Phonetically regular nonsense words. Students were administered the Word Attack subtest of the *Woodcock Reading Mastery Tests* (WMRT; Woodcock, 1973) consisting of 45 phonetically regular "nonsense" words (e.g., "bim," "bafmotbem"). Reliability estimates for this subtest generally range from .80 to .95 (Luftig, 1989). This decoding task was intended to measure a student's skill at applying phonetic rules to unfamiliar, phonetically regular words.

Group Measures

Each data collector used one examiner packet and two student materials packets for each grade. The examiner packet consisted of the standardized directions for the written retell and cloze tasks and the Comprehension subtest of the *Stanford Diagnostic Reading Test* (SDRT) (Karlsen, Madden, & Gardner, 1975). Student Packet 1 consisted of: (a) a 400-word grade-level folktale; (b) a blank sheet of lined paper; and (c) a second 400-word grade-level folktale. Student Packet 2 consisted of a duplicate of the second 400-word grade-level folktale, but with every seventh word replaced with a blank, and the Comprehension subtest of the SDRT.

The folktales used for the written retell and the cloze tasks were chosen because of their use in previous studies of reading (Fuchs et al., 1988; Jenkins, Heliotis, Haynes, & Beck, 1986). Of the original pool of 14 folktales, 4 were used in this study, 2 for each grade. These folktales were selected based on estimates of their readability as determined using the Spache, Dale-Chall, and Flesch readability formulas. Readabilities were determined by entering the first 100 words of each folktale into a commercially available computer program (MECC, 1983). The two third-grade and two fifth-grade passages used in this study were selected by identifying all folktales from the original 14 that approximated third- and fifth-grade ratings and then by choosing two passages from each grade that had the least variability in readability ratings.

Written retell. Students were presented with the first 400-word folktale and were given 5 minutes to read it silently. They then were given an additional 5 minutes to retell the story in writing using their own words. Students were given four verbal prompts at 1-minute intervals (e.g., "Is there anything else you can remember about [folktale title], write it down") to encourage the students' best performance. No further assistance was given. The written retell task was scored by counting the total number of recognizable words written, regardless of whether the words pertained to the passage (Fuchs et al., 1988). Recognizable words included: (a) incorrectly spelled words, (b) numbers, (c) isolated letters functioning as words (e.g., I, a), (d) abbreviations, and (e) incorrectly capitalized words. Compound, hyphenated, and contracted words were scored as one word. Punctuation marks and nonsense letter clusters were not counted as recognizable words. Fuchs et al. (1988) reported correlations of .76 between total recognizable words and the Comprehension subtest of the *Stanford Achievement Test* (SAT). The correlations ranged from .60 to .79 with other informal measures of comprehension.

Cloze. Students were given 5 minutes to read the second 400-word folktale silently. At the end of 5 minutes, Packet 1 was collected and Packet 2 was distributed. Students then were given an additional 5 minutes to fill in the missing words on the cloze worksheet. The cloze worksheet was a duplicate copy of the folktale, but with every seventh word replaced with a blank. The score obtained on this measure was the number of exact matches written. An exact match was defined as a written response that was phonetically the same as the word used in the story, even if the word was spelled incorrectly. The number of blanks per worksheet differed slightly with 55 and 59 blanks for the third- and fifth-grade stories, respectively. Fuchs et al. (1988) reported a correlation of .69 between the number of cloze exact matches and the Comprehension subtest of the SAT and correlations ranging from .47 to .73 with other informal measures of comprehension.

SDRT Comprehension subtest. The SDRT is a group-administered test that is designed to identify specific reading strengths and weaknesses. The Reading Comprehension subtest is divided into two types of comprehension skills, literal and inferential. Third-grade students were administered the Green Level consisting of 24 literal questions and 24 inferential questions. Fifth-grade students were administered the Brown Level containing 30 literal questions and 30 inferential questions. Students were required to read silently and answer questions pertaining to several short passages. The questions were presented in a 4-item multiple choice format. The number of inferential and literal questions answered correctly were recorded. Internal consistency reliability estimates exceed .80 for this subtest at both levels (Salvia & Ysseldyke, 1988).

PROCEDURE

Training of Data Collectors

Eight graduate students and one faculty member from the school psychology program at the University of Oregon served as data collectors. They attended three, 1-hour training sessions. During the first session, they were trained to give the three group measures (i.e., written retell, cloze, and SDRT subtest). This training consisted of having the data collectors read a set of scripted directions in a prescribed sequence after modeling was provided by the trainer. Instructions on how to set up the test environment also were reviewed. The second training session covered administration and scoring of the three individual measures (CBM passages, phonetically regular words, phonetically regular nonsense words). The training format was similar to that of the first session. Additionally, trainees practiced scoring recorded CBM readings. During the third session, data collectors were instructed on how to score the group measures. They received practice on scoring two written retell samples and two cloze protocols.

Interrater Agreement

All data collectors scored five identical reading tasks: (a) two audio taped CBM reading passages, (b) an audio taped phonetically regular word list, (c) an audio taped phonetically regular nonsense word list, (d) a cloze worksheet, and (e) a written retell sample, for reliability of scoring purposes. Interrater agreement coefficients of .99, .98, .93, .91, and .99 were obtained for reading passages, phonetically regular word list, nonsense word list, cloze, and written retell tasks, respectively, using the formula:

$$\frac{\text{number of agreements}}{(\text{number of agreements} + \text{number of disagreements})} * 100$$

Data Collection

Testing was completed in two sessions. One testing session consisted of the individually administered measures. The other session consisted of the group-administered test. For both individual and group testing sessions, measures were given in a prescribed order. Typically, both testing sessions occurred on the same day. Total testing lasted approximately 1 hour and 10 minutes. During group testing, at least one data collector for every 10 students was present to answer individual questions, make sure students were on the correct page, and maintain a quiet testing environment.

Data Analysis

The reading factor(s) accounting for the relationships among the measured reading variables were examined using confirmatory factor analysis. Confirmatory factor analysis offers important advantages over exploratory factor analysis for the examination of reading factor structures because of the researcher's ability to impose substantively motivated constraints on the model (Long, 1983). That is, the researcher is able to specify a factor model a priori, on the basis of theory, and examine the extent to which the data are inconsistent with the model (Fassinger, 1987). Although exploratory factor analysis determines the most likely (best fitting) model given the obtained data (with the imposition of certain fairly arbitrary constraints), other models might also provide an adequate fit to the data. In contrast, confirmatory factor analysis tests whether the theoretically derived model is one of the models that would fit the data adequately. Thus, instead of relying on subjective judgment that the theoretical model is adequately reflected by the empirical model as in exploratory factor analysis, the researcher can test explicitly the hypothesis that the theoretical model adequately fits the data.

Confirmatory factor analysis is particularly useful when contrasting the fit of competing models that can be supported theoretically. In this study, the role of reading fluency was investigated by comparing the fit of four models to the data: (a) a unitary model where decoding, fluency, and comprehension were not distinct; (b) a two-factor model of decoding and comprehension where reading fluency was defined as part of the decoding construct; (c) a decoding and comprehension two-factor model where reading fluency was defined as comprehension; and (d) a three-factor model of reading proficiency that specified reading fluency as a distinct construct. Data were analyzed with the LISREL computer program (Joreskog & Sorbom, 1984).

Because there is no single accepted criteria for evaluating the adequacy of a model, seven types of indices were used to evaluate model fit: (a) *t*-values for estimated parameters; (b) modification indices for constrained parameters; (c) the goodness-of-fit Chi-Square statistic; (d) various goodness-of-fit indices including the goodness-of-fit index (GRI), the adjusted goodness-of-fit index (AGFI), and the Tucker-Lewis Index (TLI); (e) the root-mean-square residual (RMSR); (f) normalized residuals; and (g) the Chi-Square difference test (for hierarchical model comparisons only).

Each model proposed to account for the relationships between measured variables specifies parameters that are estimated and constrained. Estimated parameters include factor loadings, residual variances, and correlations that were expected to be nonzero for that model. All model parameters are reported as standardized coefficients (i.e., factor loadings were correlations with the latent variable). For each estimated parameter, a *t*-value provided a significance test of the null hypothesis that the estimated parameter was not necessary (i.e., that the factor loading or correlation was 0.0) in specifying the model. Parameter *t*-values greater than 2.0 indicate a significant contribution to the model (Fassinger, 1987). Constrained parameters were those factor loadings and correlations that were specified to be 0.0 for that model. Modification indices are estimates of the expected improvement in fit that would be obtained from relaxing a single model constraint (i.e., allowing a variable to load on a factor or residual variances to be correlated). Modification indices larger than 9.0 indicate that a substantial improvement in fit would result from relaxing the associated constraint (Fassinger, 1987).

The goodness-of-fit Chi-Square statistic "allows a test of the null hypothesis H_0 that a given model provides an acceptable fit of the observed data" (Long, 1983, p. 63). A nonsignificant goodness-of-fit Chi-Square statistic fails to reject the hypothesis that the specified model accounts for the observed data. A significant goodness-of-fit Chi-Square statistic indicates that the model does not adequately fit the data.

Various goodness-of-fit indices have been proposed, with little consensus on the most appropriate index to consider. The goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), and Tucker-Lewis index (TLI) generally range between 0.0 and 1.0, with 1.0 representing a perfect fit, and values above .90 indicating an adequate fit (Bentler & Bonett, 1980; Fassinger, 1987; Marsh, Balla, & McDonald, 1988).

The root-mean-square residual (RMSR) is obtained from the residual correlation matrix and represents the middle residual correlation between measured variables. Small values of RMSR (e.g., below .10) typically indicate an adequate fit (Fassinger, 1987). The normalized residual correlations provide a significance test of the hypothesis that there was no relationship between measured variables that was unexplained by the model. Normalized residuals below 2.0 are considered nonsignificant and an indication of adequate fit (Fassinger, 1987).

The Chi-Square difference test was used to evaluate the null hypothesis that there was no improvement in model fit as a result of estimating additional parameters when hierarchically nested models were compared (Bentler & Bonett, 1980). A model is said to be hierarchically nested within another model when it can be obtained by placing one or more constraints on the parameters estimated in the other model (e.g., specifying that latent variables are identical by constraining the correlation between factors as 1.00). In this study, the one-factor model was nested within the two-factor models Because the one- and two-factor models were not, strictly speaking, nested within the three-factor model, the Chi-Square difference test was reported but not evaluated for significance. The two-factor, fluency as decoding model and the two-factor, fluency as comprehension model were not hierarchically related and consequently could not be compared directly with the Chi-Square difference test.

RESULTS

Means and standard deviations for all measured variables by grade are reported in Table 1. While means for all variables except Cloze Exact Matches increased from third to fifth grade, the increase is illusory since different, more difficult materials were used for each grade level. The correlations among all measured variables are reported by grade in Table 2. It is interesting to note that 24 of the 28 correlation coefficients decreased in magnitude from the third to the fifth grade, with the decoding of nonsense words and the written retell variables displaying the most substantial decreases. The correlation matrices for third and fifth grade were significantly different, $\chi^2(36) = 62.14$, $p < .01$. Consequently, separate analyses were conducted for each grade level.

The results of each model tested are represented in Figures 1 to 4, which display the relations between measured

TABLE 1
Descriptive Statistics for Third- and Fifth-Grade Students' Performance on Reading Tests

Variable	Grade 3[a] Mean	Grade 3[a] SD	Grade 5[b] Mean	Grade 5[b] SD
CBM Oral Reading Fluency 1	110.2	36.1	134.3	33.8
CBM Oral Reading Fluency 2	119.1	42.2	162.8	35.9
Phonetic Regular Words	33.2	13.1	47.6	16.6
Nonsense Words	29.4	8.5	32.0	6.7
Written Recall	46.1	20.5	62.1	18.2
Cloze Exact Matches	20.5	8.1	17.5	7.3
SDRT Literal Comprehension	19.8	4.5	23.2	6.4
SDRT Inferential Comprehension	18.6	4.6	23.1	6.5

Note. CBM = Curriculum-Based Measurement. SDRT = Stanford Diagnostic Reading Test. All scores are reported in raw score units as the number correct with the exception of Written Retell which is the raw score number of words written.

[a]$n = 114$. [b]$n = 124$.

variables. In each figure, values before the slash are for the third-grade sample and values after the slash are for the fifth-grade sample. The results for a unitary model with decoding, fluency, and comprehension not functioning as distinct constructs are presented in Figure 1. Illustrated in Figure 2 are the results for a two-factor model with oral reading fluency considered a measure of Decoding. A two-factor model of reading with oral reading fluency treated as a measure of Comprehension produced the results in Figure 3. Finally, the results of specifying a three-factor model of reading proficiency with Oral Reading Fluency as a distinct construct are displayed in Figure 4.

For each measured variable, the proportion of residual variance is portrayed in a circle with an arrow toward the variable. Larger residual variances indicate poorer functioning of the variable as a measure of a reading construct. Correlations between measured-variable residuals (double-headed arrows) were allowed for the SDRT measures and for the CBM measures. Because both SDRT literal comprehension and SDRT inferential comprehension were assessed simultaneously with similar formats (i.e., read a passage and darken a circle on the answer sheet indicating the correct answer to a question), it was reasonable to anticipate method or response effects between the two variables (e.g., Newcomb & Bentler, 1988). Thus, the correlated residuals were considered measurement variance rather than representative of a meaningful reading construct. This interpretation was supported by the stable and consistent residual correlations across all reading models for the literal and inferential comprehension subtests. Regardless of whether the SDRT was considered a measure of overall Reading Competence (one-factor model) or of Reading Comprehension (two- or three-factor models), the SDRT residuals were correlated about .76. Similar reasoning was employed in allowing correlated residuals for the CBM measures as well. However, it was not possible computationally to separate measurement variance (correlated residuals) from construct variance (factor loadings) for the three-factor fluency distinct model.

Third-Grade Sample

For third-grade students, the one-factor model of reading (Figure 1) was the most parsimonious model that provided

TABLE 2
Correlations between Reading Variables for Third- and Fifth-Grade Samples

Variable	1	2	3	4	5	6	7	8
1. CBM Oral Reading Fluency 1		.94	.71	.69	.59	.75	.57	.58
2. CBM Oral Reading Fluency 2	.91		.70	.68	.60	.79	.58	.60
3. Phonetic Regular Words	.70	.68		.59	.49	.60	.52	.55
4. Nonsense Words	.49	.48	.47		.55	.63	.63	.67
5. Written Recall	.38	.40	.31	.27		.59	.56	.50
6. Cloze Exact Matches	.62	.63	.53	.45	.46		.61	.61
7. SDRT Literal Comprehension	.60	.62	.44	.34	.33	.64		.89
8. SDRT Inferential Comprehension	.55	.54	.40	.30	.36	.63	.84	

Note. Correlations for the third-grade sample (n = 114) are above the main diagonal, correlations for the fifth-grade sample (n = 124) are below the main diagonal. CBM = Curriculum-Based Measurement, SDRT = Stanford Diagnostic Reading Test.

an adequate fit to the data. The one-factor model hypothesized that all measured variables represented one latent variable: Reading Competence. Factor loadings on Reading Competence ranged from .69 for written retell to .90 for CBM Oral Reading Fluency 2. Residual variances ranged from a low of 20% of the variance for CBM Oral Reading Fluency 2 to a high of 53% of the variance of the written retell measure. All parameter t-values were greater than 2.0, indicating that all parameters were necessary for the model. None of the modification indices were larger than 9.0, indicating that little further improvement in fit would be obtained from relaxing parameter constraints.

The goodness-of-fit Chi-Square statistic for the one-factor mode of reading was nonsignificant, $p > .05$, indicating that the departure of the observed covariance matrix from the covariance matrix specified by the model could be due reasonably to chance sampling variability. For the one-factor model, the goodness-of-fit index (GFI) and Tucker-Lewis Index (TLI) were both greater than .90, again indicating an adequate fit to the data. Although the adjusted goodness-of-fit index (AGFI) was somewhat lower at .88, none of the alternative models provided an improvement in the AGFI. The root mean square residual was small (.03), and all normalized residuals were smaller than 1.0, indicating that all relationships between measured variables were explained adequately by the model.

In addition, using the Chi-Square difference test, a significant improvement in fit was not obtained by specifying either: (a) a two-factor model of reading with fluency as decoding (Figure 2), $\chi^2(1) = 0.62$, $p > .50$; or (b) a two-factor model with fluency as comprehension (Figure 3), $\chi^2(1) = 0.00$, $p > .50$. A three-factor model of reading with fluency distinct (Figure 4) also did not offer a substantial improvement in fit, $\chi^2(2) = 0.62$. Thus, the one-factor model of reading competence could not be rejected statistically as a plausible explanation of the data. The one-factor model was supported further by the very strong correlations between factors for the alternative two- and three-factor models (r = .98 to .990). When latent variables are so highly correlated, it is difficult to make the case that they are measuring distinct constructs.

Fifth-Grade Sample

For the fifth-grade sample, a two-factor model of reading with fluency representing decoding was the most parsimonious model that provided an adequate fit to the data. Factor loadings on Reading Comprehension ranged from .51 for written retell to .86 for cloze exact

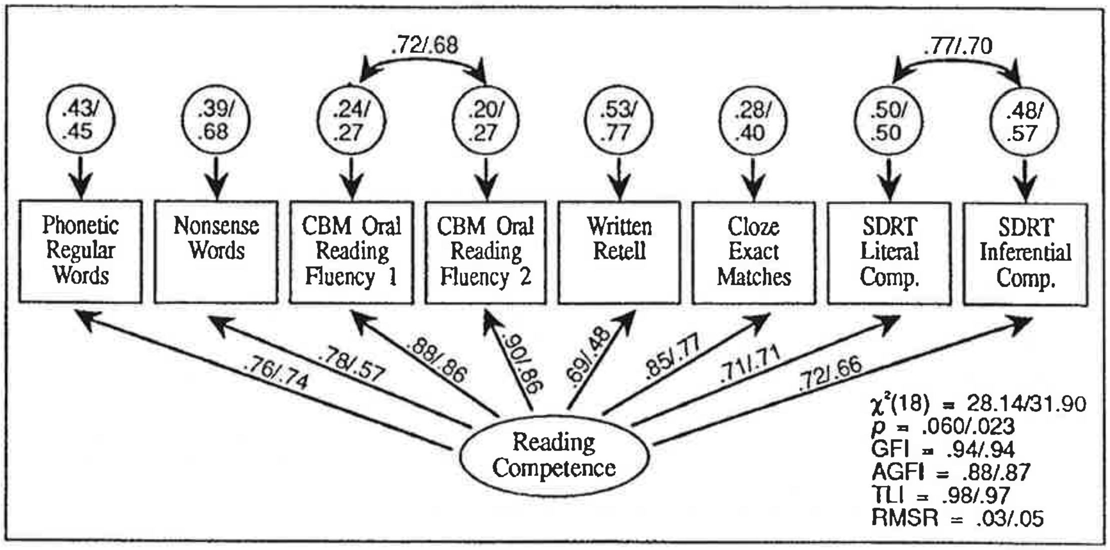

FIGURE 1. One-factor model of reading for Grade 3/Grade 5. Ellipses represent latent constructs, rectangles are measured variables, and circles with numbers are residual variances. Factor loadings are standardized. Chi-Square and fit indices were obtained from unstandardized coefficients. Root mean square residual (RMSR) was based on standardized coefficients. Two-headed arrows represent correlations.

matches. Factor loadings on Reading Decoding ranged from .56 for nonsense words to .90 for Oral Reading Fluency 1. Residual variances ranged from a low of 19% of the variance for CBM Oral Reading Fluency 1 to a high of 74% of the variance of the written retell measure. In particular, the nonsense words and written retell measures demonstrated a very high residual variance indicating that they did not function well as measures of reading constructs for fifth-grade students. All parameter t-values were greater than 2.0 except for the residual correlation between CBM Oral Reading Fluency measures (t = 1.95), indicating that all parameters were necessary for the model. The residual correlation between DBM Oral Reading Fluency measures was maintained in the model because of its substantive meaning and t value approaching 2.0. None of the modification indices were larger than 9.0, indicating that little further improvement in fit would be obtained from relaxing parameter constraints.

For fifth graders, the goodness-of-fit Chi-Square statistic was nonsignificant for the two-factor, fluency as decoding model of reading; the hypothesis that the specified model generated the observed data could not be rejected. Conversely, the goodness-of-fit Chi-Square statistic was significant for the one-factor and the two-factor, fluency as comprehension models of reading. Therefore, the hypothesis that those models accounted for the data was rejected.

The GFI, AGFI, and TLI for the two-factor, fluency as decoding model were all greater than .90, indicating an adequate fit of the data to the model. No other model provided an improvement in fit. For Model 2, the root-mean-square residual correlation was small (.03), and all normalized residuals were smaller than 1.0, indicating that all relationships between measured variables were explained adequately.

Using the Chi-Square difference test, a significant improvement in fit over the one-factor model was obtained from a two-factor model of reading with fluency considered decoding, $\chi^2(1)$ = 17.09, $p <$ .05; the corresponding improvement in fit

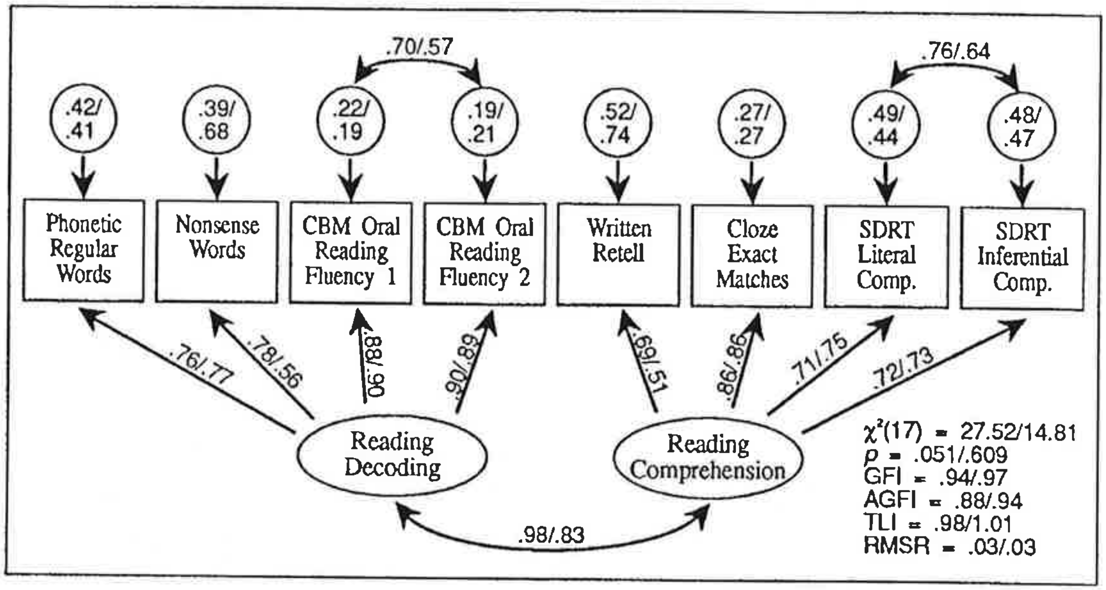

FIGURE 2. Two-factor model of reading with fluency as decoding for Grade 3/Grade 5. Ellipses represent latent constructs, rectangles are measured variables, and circles with numbers are residual variances. Factor loadings are standardized. Chi-Square and fit indices were obtained from unstandardized coefficients. Root mean square residual (RMSR) was based on standardized coefficients. Two-headed arrows represent correlations.

was not significant for a two-factor model of reading with fluency considered comprehension, $\chi^2(1) = 1.44$, $p > .05$. Further improvement in fit was not substantial for a three-factor model of reading with fluency distinct, $\chi^2(1) = 1.12$. Although the fluency as decoding and fluency as comprehensions models (Figures 2 and 3) could not be compared directly with the Chi-Square difference test, this pattern of results also supports the fluency as decoding, two-factor model as superior to the fluency as comprehension model.

In the two-factor, fluency as decoding model, the decoding and CBM measures were constrained to have a zero loading on the Reading Comprehension factor, and the comprehension subtests were constrained to a zero loading on the Reading Decoding factor (i.e., no direct relationship existed between the CBM measures and the comprehension factor). However, the Reading Decoding and Reading Comprehension factors were highly correlated, $r = .83$, $p < .05$. Consequently, the actual correlations (representing both direct and indirect relationships) between measured and latent variables still were substantial. The correlations between measured and latent variables (i.e., the factor structure matrix) are reported in Table 3. Correlations that are not underlined represent the direct relationship between the measured and latent variables while underlined correlation represent an indirect relationship. Although there was no direct relationship between the CBM measures and the Reading Comprehension factor, the indirect relationship (i.e., from CBM to Reading Decoding to Reading Comprehension) was comparable in magnitude to the direct relationship between the SDRT measures and the Reading Comprehension factor. Thus, even for the fifth-grade sample where the fluency as decoding model proved the best fit to the data, the CBM measures provide a very good (albeit indirect) estimate of the Reading Comprehension factor.

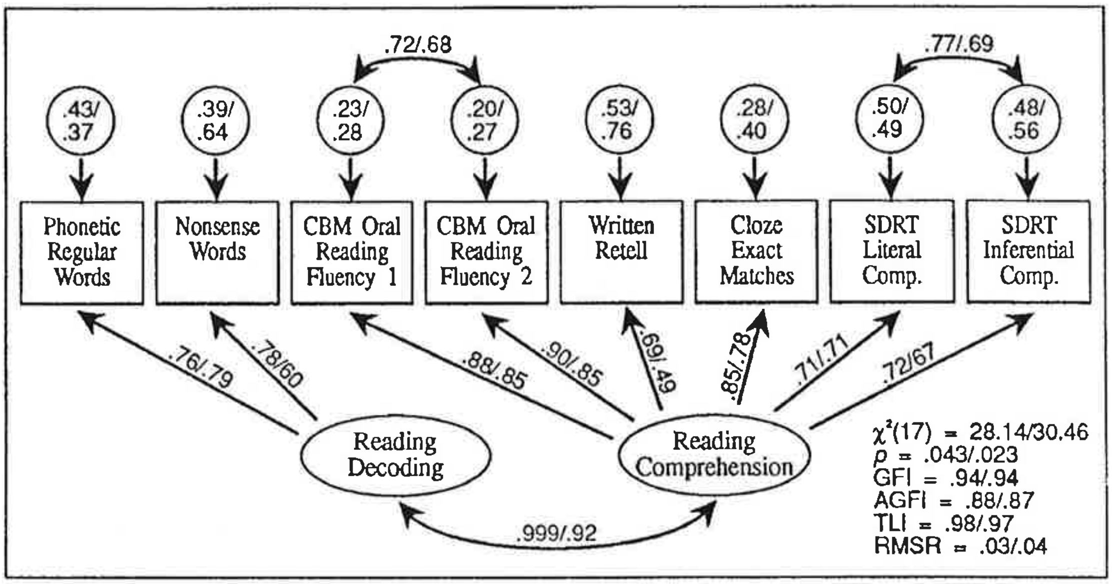

FIGURE 3. Two-factor model of reading with fluency as comprehension for Grade 3/Grade 5. Ellipses represent latent constructs, rectangles are measured variables, and circles with numbers are residual variances. Factor loadings are standardized. Chi-Square and fit indices were obtained from unstandardized coefficients. Root mean square residual (RMSR) was based on standardized coefficients. Two-headed arrows represent correlations.

DISCUSSION

In spite of a number of empirical studies supporting the use of oral reading fluency as an index of reading proficiency including comprehension, test producers, and a number of reading authors (e.g., Duffelmeyer, 1983; Johnston & Allington, 1983) have been less than enthusiastic about the measurement of oral reading fluency as an index of reading proficiency. Practitioners also frequently express concern that oral reading fluency measures only decoding. It has been argued that, in part, the lack of acceptance of oral reading measurement is a face validity issue (Fuchs et al., 1988; Potter & Wamre, 1990).

As a potential solution to the face validity problem, Potter and Wamre (1990) suggested that a validated tie between oral reading fluency and reading *theory* be demonstrated. They argued that no new theories of reading must be developed to test this tie; it need only be validated. This study was conducted primarily to investigate CBM oral reading fluency's contribution to theoretical reading process models. Confirmatory factor analysis was used to examine whether oral reading constituted a significant role in a single-factor model of reading or whether it should be defined as a decoding construct, a comprehension construct, or as a separate construct. Additionally, the study attempted to discern whether oral reading functioned differently from a developmental perspective.

As would be expected from previous research, the efficacy of oral reading fluency as a measure of reading proficiency and comprehension was supported strongly. Preliminary work by Collins (1989) suggested that a three-factor model of reading with oral reading fluency as a distinct construct best explained the observed relations between reading measures in her study of second graders. However, her sample size was small and some of the measures used lacked sufficient technical adequacy. Replication

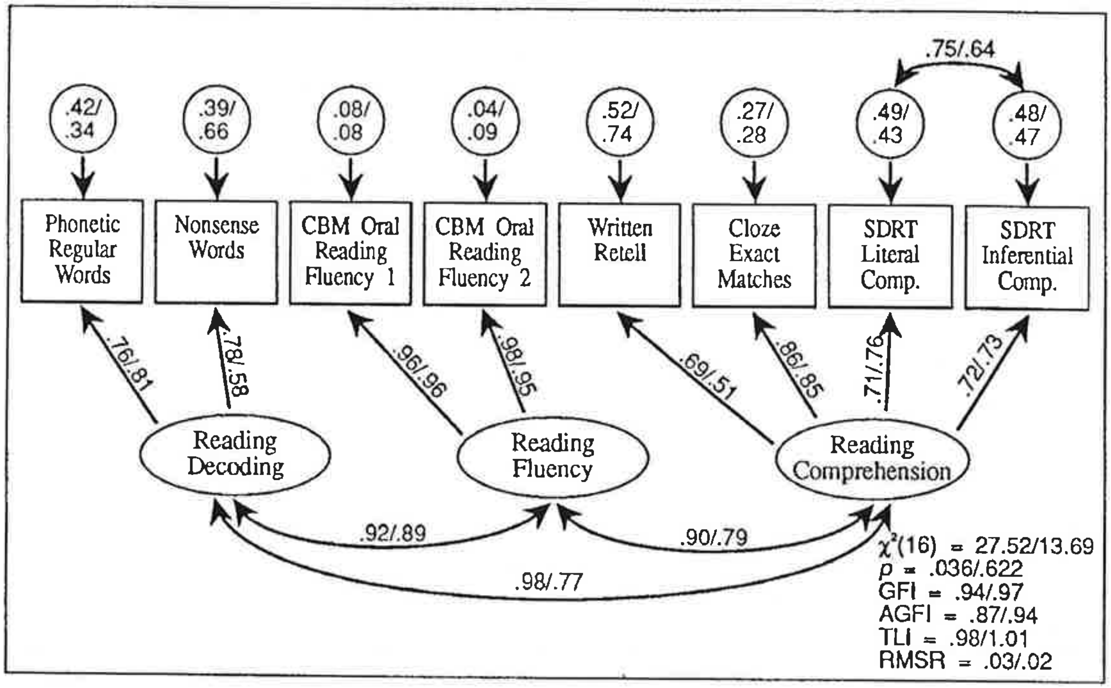

FIGURE 4. Three-factor model of reading with fluency distinct for Grade 3/Grade 5. Ellipses represent latent constructs, rectangles are measured variables, and circles with numbers are residual variances. Factor loadings are standardized. Chi-Square and fit indices were obtained from unstandardized coefficients. Root mean square residual (RMSR) was based on standardized coefficients. Two-headed arrows represent correlations.

therefore was required. The three-factor model explained the obtained relationships adequately for both grades in this study. However, it was not the most parsimonious explanation.

For third-graders, a single factor model of reading, labeled Reading Competence could not be rejected. Each of the specific measures tested in the study contributed significantly to the Reading Competence model. Two CBM reading measures where students read aloud from third-grade basal readers correlated the highest with the model (r = .88 and .90). They correlated higher with the model than either of the two more conventional comprehension measures, SDRT inferential and literal comprehension (r = .71 and .72, respectively).

For fifth graders, a unitary model was rejected. Reading proficiency was best portrayed as being composed of two constructs fitting the common conception of reading, decoding, and comprehension. In the most defensible model, CBM oral reading fluency best fit with the construct of decoding. However, even though decoding and comprehension could be differentiated as constructs in the fifth grade, they still were correlated highly (r = .83). Additionally, the CBM oral reading measures correlated as high or higher with Reading Comprehension (r = .74 and .75) as did the SDRT measures (r = .73 and .75). Only cloze exact matches was more highly correlated with Reading Comprehension for fifth graders.

Three notes of caution must always be raised when interpreting any theoretical work such as that conducted in this investigation. As noted by Potter and Wamre (1990), theories can be proven false, but they cannot be proven true. The very nature of confirmatory factor analysis is oriented to *rejecting* theories, not proving them "true" (Cliff, 1983). The

TABLE 3
Correlations between Measured Variables and Latent Factors for Fifth-Grade Students from the Two-Factor Model of Reading Competence with Oral Reading Fluency as Decoding

Variable	Decoding	Comprehension
CBM Oral Reading Fluency 1	.90	.75
CBM Oral Reading Fluency 2	.89	.74
Phonetic Regular Words	.77	.64
Nonsense Woods	.56	.47
Written Retell	.43	.51
Cloze Exact Matches	.71	.86
SDRT Literal Comprehension	.62	.75
SDRT Inferential Comprehension	.60	.73

Note. Underlined correlations were obtained from the factor-structure matrix. CBM = Curriculum-Based Measurement. SDRT = Stanford Diagnostic Reading Test.

conclusions drawn from model testing also are particularly susceptible to the influence of sample sizes. Finally, the external validity of the study must be considered. The students in this sample represented primarily general education students who came from a community where reading scores typically are above the norm on published group achievement tests. These cautions require that the generalizations from this study's model testing be cross-validated. Important variables to consider should be an increased sample size, different populations of readers, and different operationalizations of the reading constructs. In the absence of cross-validation, generalizations of the reading model outcomes to more unique readers (e.g., students in special education) should be made with caution.

Theoretical Contributions

The study's demonstration of the potential role that oral reading fluency plays in the reading process fits into a number of theoretical reading models (e.g., Chall, 1983; Jager-Adams, 1990; Laberge & Samuels, 1974; Stanovich, 1986). Each of these authors propose a pivotal role for oral reading fluency in the reading process. For example, Jager-Adams (1990) stated that "the ability to read words, quickly, accurately, and effortlessly, is critical to skillful reading comprehension" (p. 3). Further, she wrote that "reading proficiency is strictly limited by the speed, accuracy, and effortlessness with which readers can respond to print as coherent, orthographic, phonological, and semantic (meaning bearing) patterns" (p. 8). The different reading models observed in this study also match Chall's (1983) developmental model of reading acquisition where readers progress through the stages of reading proficiency development. Three of them have direct relevance to this study. Stage 1 readers (kindergarten through second grade) acquire phonemic awareness and decoding skills by learning the alphabet, letter sounds, and letter groups. Stage 2 readers (second and third grade) practice decoding skills and develop fluency and speed. Stage 3 readers (Grades 4 through 8) concentrate less on the printed words and more on the ideas and content of the reading material (i.e., comprehension). From Chall's perspective, reading comprehension would not emerge as a distinct construct before Stage 3. For the third graders in this study, no evidence of comprehension as a unique, clearly separable factor was observed. The emergence of reading comprehension as a construct distinct from reading decoding in a model of overall reading proficiency was observed for fifth graders, as predicted in Chall's Stage 3.

Other observed differences between third and fifth graders fit Chall's hypothesis of developmental changes in the structure of reading activities. For example, there was a substantial decrease in the quality of nonsense words and written retell as measures of Decoding and Comprehension for the older students. Factor loadings for nonsense words and written retells decreased markedly from third to fifth grade and residual variances increased. In contrast, the quality of oral reading fluency, cloze, and the SDRT subtests as measures of reading constructs was relatively consistent across the two grade levels.

The developmentally different models of reading may correspond to observations made recently by Jenkins and Jewell (1990). They examined the construct measured by CBM oral reading fluency with respect to published, norm-referenced measures of reading proficiency (Metropolitan Achievement Test) and a maze reading comprehension task for students in Grades 1 to 6. Jenkins and Jewell (1990) noted a decrease in the relationship between CBM oral reading fluency and the reading comprehension subtests with advancing grade level from 4 to 6. These data also could indicate a shift in the role of oral reading fluency from comprehension to decoding with advancing grade. The Jenkins and Jewell (1990) conclusions, however, should be interpreted cautiously since very easy reading passages (i.e., first-grade level) were used to measure the reading fluency of subjects in second through sixth grades. Thus, at upper grade levels, student CBM oral reading fluency may have reached asymptote, resulting in a nonnormal distribution. This ceiling effect with easy materials was observed in other work (e.g., Rodden-Nord & Shinn, 1991) and limits the interpretation of relations between variables. The study reported here did not document a change in the relationship between CBM oral reading and either the published or cloze comprehension measures. In addition, Jenkins and Jewell (1990) did not demonstrate a corresponding increase in the relationship between CBM oral reading fluency and measures of decoding.

Practical Implications

Earlier in this article, a question was proposed as to whether CBM oral reading fluency measures decoding or comprehension. Implications for the utility of CBM for educational decision making were identified. If CBM oral reading fluency were indicative of only decoding skills, inferences about broad reading skills (e.g., level of reading proficiency, rates of progress, instructional intervention efficacy) would be limited and inappropriate. Should oral reading fluency prove to be as valid a measure of reading comprehension as other measures, then its use as a general indicant of reading would be justified. For third graders, the question of "what CBM oral reading measures" appears to be moot; there did not appear to be a distinction between decoding and comprehension constructs and CBM correlated highly (.88 and .90) with the observed Reading Competence construct.

With respect to what CBM measures for fifth graders, a different answer(s) is required. Oral reading fluency clearly loaded on what traditionally would be defined as a Reading Decoding factor. However, Reading Decoding was strongly related to Reading Comprehension (r = .83) and the *indirect* relationship between CBM oral reading fluency and comprehension (i.e., Reading Comprehension to Reading Decoding to oral reading fluency) was high. Indeed, the indirect relationship between CBM oral reading fluency and comprehension still was as strong or stronger than the direct relationship between the SDRT comprehension subtests and Reading Comprehension. Only the Cloze procedure loaded more highly on Reading Comprehension than the indirect relationship of oral reading fluency (r = .86).

The best fitting model for fifth graders also may be open to a different interpretation. A serious limitation to the interpretation of these results regards the labeling and interpretation of the latent variables. Cliff (1983) cautions strongly against the "nominalistic fallacy." The nominalistic fallacy refers to the belief that the naming of a factor accurately identifies what the factor represents.

> Even in what is called "confirmatory" factor analysis, it is not the nature of the factors which is confirmed; the only thing which is confirmed is that the observed covariance matrix is not *inconsistent* with a certain pattern of parameters. It does not tell us what those parameters mean, and experience has shown that our belief that we do know what they mean is often ill-founded. (p. 122)

In the most parsimonious model fitting the fifth-grade data, the two factors were labeled Decoding and Comprehension. However, alternative interpretations also are possible. It should be noted in particular that all measured variables loading on the "Decoding" factor involve *oral* reading and *oral* responding. All measured variables loading on the "Comprehension" factor involve *silent* reading and *written* responding. A case could be made that the two factors could be explained by common method variance and should be labeled more appropriately "Oral Reading/Verbal Response" and "Silent Reading/Written Response." Future research should attempt to parse response formats from the constructs measured by including an oral reading/oral responding measure of comprehension.

Regardless of which two-factor model is considered or what the latent constructs are labeled, it may not matter which construct CBM oral reading fluency loads most on theoretically. From a pragmatic perspective, CBM oral reading works as a general index of reading proficiency including comprehension. Similar conclusions have been reached about other measured variables and their relationships to theoretical models (Macmann & Barnett, in press). Regardless of the most parsimonious reading model, the best measures of overall reading competence were CBM oral reading fluency and cloze for *both* grades.

Summary

Potter and Wamre (1990) maintained that a merger of applied and basic theoretical research on CBM and oral reading fluency has much to contribute to the knowledge base and acceptance of this emerging measurement technology. This study confirms that CBM oral reading fluency fits current theoretical models of reading well and can be validated as a measure of general reading achievement, including comprehension. In conjunction with previous research providing validity data from a variety of perspectives, the results of this study suggest that the face validity arguments regarding CBM oral reading measures should be put to rest. Instead efforts should focus on further increasing the utility of using short duration, content valid measures on a repeated basis such as CBM to make a variety of problem-solving decisions. CBM was designed to meet traditionally accepted standards of reliability and validity, but to *exceed* the utility of published tests with respect to the purposes of assessment and problem-solving decision making.

REFERENCES

Allen, D. (1989). Periodic and annual reviews and decisions to terminate special education services. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 184-203). New York: Guilford.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588-606.

Carnine, D., & Silbert, J. (1979). *Direct instruction reading.* Columbus, OH: Merrill.

Chall, J. S. (1983). *Stages of reading development.* New York: McGraw-Hill.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research, 18,* 115-126.

Collins, V. L. (1989). *Reading fluency: An examination of its role in the reading process.* Unpublished master's thesis. University of Oregon, Eugene, OR.

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

Deno, S. L. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review, 15,* 358-374.

Deno, S. L., Marston, D., Shinn, M. R., & Tindal. G. (1983). Oral reading fluency: A simple datum for scaling reading disability. *Topics in Learning and Learning Disability, 2,* 53-59.

Deno, S. L., Mirkin, P., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49,* 36-45.

Deno, S. L., Mirkin, P., & Wesson, C. (1984). How to write effective data-based IEPs. *Teaching Exceptional Children, 16,* 99-104.

De Soto, J. L., & De Soto, C. B. (1983). Relationship of reading achievement to verbal processing abilities. *Journal of Educational Psychology, 75,* 116-127.

Duffelmeyer, F. A. (1983). Response to Fuchs, Fuchs, and Deno. *Reading Research Quarterly, 18,* 370-371.

Fassinger, R. E. (1987). Use of structural equation modeling in counseling psychology research. *Journal of Counseling Psychology, 34,* 425-436.

Fuchs, L. S. (1989). Evaluating solutions: Monitoring progress and revising intervention plans. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 155-183). New York: Guilford.

Fuchs, L. S., & Deno, S. L. (1981). *The relationship between curriculum-based mastery measures and standardized achievement tests in reading* (57). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Fuchs, L. S., Deno, S. L., & Mirkin, P. (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. S., & Fuchs, D. (1984). Criterion-referenced assessment without measurement: How accurate for special education. *Remedial and Special Education, 5,* 29-32.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation on student achievement: A meta-analysis. *Exceptional Children, 53,* 199-208.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989). Monitoring reading growth using student recalls: Effects of two teacher feedback systems. *Journal of Educational Research, 83,* 103-110.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9,* 20-28.

Fuchs, L. S., Fuchs, D., & Tindal. G. (1986). Effects of mastery learning procedures on student achievement. *Journal of Educational Research, 79,* 286-291.

Fuchs, L. S., & Shinn, M. R. (1989). Writing CBM IEP Objectives. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 132-154). New York: Guilford.

Fuchs, L. S., Tindal, G., Fuchs, D., Shinn, M. R., Deno, S. L., & Germann, G. (1983). *Technical adequacy of basal readers' mastery tests: Holt basic series* (130). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Germann, G., & Tindal, G. (1985). An application of curriculum based assessment: The use of direct and repeated measurement. *Exceptional Children, 52,* 244-265.

Gibson, E., & Levin, H. (1975). *The psychology of reading.* Cambridge, MA: MIT Press.

Golinkoff, R. M. (1975). A comparison of reading comprehension processes in good and poor comprehenders. *Reading Research Quarterly, 4,* 624-659.

Goodman, K. S. (1986). *What's whole in whole language.* Portsmouth, NH: Heinemann.

Heilman, A. W. (1967). *Principles and practices of teaching reading.* Columbus, OH: Merrill.

Howell, K. W., & Kaplan, J. S. (1980). *Diagnosing basic skills: A handbook for deciding what to teach.* Columbus, OH: Charles Merrill.

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

Jenkins, J. R., & Jewell, M. (1990). *Examining the validity of two measures for formative teaching: Reading aloud and maze.* Unpublished manuscript, University of Washington, Seattle.

Jenkins, J. R., Heliotis, J., Haynes, M., & Beck, K. (1986). Does passive learning account for disabled readers' comprehension deficits in ordinary reading situations? *Learning Disability Quarterly, 9,* 69-75.

Johnston, P. H., & Allington, R. L. (1983). How sharp is a unicorn's horn? *Reading Research Quarterly, 18,* 498-500.

Joreskog, K. G., & Sorbom, D. (1984). *LISREL: Analysis of linear structural relationships by the method of maximum likelihood.* Mooresville, IN: Scientific Software.

Karlsen, B., Madden, R., & Gardner, E. F. (1975). *Stanford Diagnostic Reading Test* (SDRT). San Antonio, TX: Psychological Corporation.

Kaufman, A., & Kaufman, N. (1985). *Kaufman Test of Educational Achievement.* Circle Pines, MN: American Guidance Service.

Laberge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293-323.

Larsen, S., & Hammill, D. (1976). *Test of Written Spelling* (TWS). Austin, TX: PRO-ED.

Lomax, R. G. (1983). Applying structural modeling to some component processes of reading comprehension development. *Journal of Experimental Education, 52,* 33-40.

Long, J. S. (1983). *Confirmatory factor analysis.* Beverly Hills, CA: Sage.

Luftig, R. L. (1989). *Assessment of learners with special needs.* Boston: Allyn and Bacon.

Macmann, G., & Barnett, D. (in press). Redefining the WISC-R: Implications for professional practice and public policy. *The Journal of Special Education.*

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391-410.

Marston, D. (1989). Curriculum-based measurement: What is it and why do it? In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford Press.

Marston, D., & Deno, S. L. (1982). *Implementation of direct and repeated measurement in the school setting* (106). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Marston, D., Deno, S. L., Mirkin, P. K., Lowry, L., Sindelar, P., & Jenkins, J. (1981). *An analysis of learning trends in simple measures of reading, spelling, and written expression: A longitudinal study.* Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Marston, D., & Magnusson, D. (1988). Curriculum-based assessment: District-level implementation. In J. Graden, J. Zins, & M. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 137-172). Washington, DC: National Association of School Psychologists.

MECC (Minnesota Educational Computing Consortium). (1983). *Utilities III.* Minneapolis, MN: Minnesota Educational Computing Consortium.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Newcomb, M. D., & Bender, P. M. (1988). Impact of adolescent drug use and social support on problems of young adults: A longitudinal study. *Journal of Abnormal Psychology, 97,* 64-75.

Perfetti, C. A., & Hogaboam, T. (1975). Relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology, 67,* 461-469.

Potter, M. L., & Wamre, H. M. (1990). Curriculum-based measurement and developmental reading models: Opportunities for cross-validation. *Exceptional Children, 57,* 16-25.

Rodden-Nord, M. K., & Shinn, M. R. (1991). The range of reading skills within and across general education classrooms. *The Journal of Special Education, 24,* 441-453.

Rodden-Nord, K., Shinn, M. R., & Good, R. H. (1992). Effects on classroom performance data on general education teachers' willingness to reintegrate mildly handicapped students. *School Psychology Review, 21,* 138-154.

Salvia, J., & Ysseldyke, J. E. (1988). *Assessment in special and remedial education* (4th ed.). Boston: Houghton-Mifflin.

Samuels, S. J. (1976). Hierarchical subskills in reading acquisition process. In J. T. Guthrie (Ed.), *Aspects of reading acquisition.* Baltimore: Johns Hopkins Press.

Samuels, S. J. (1981). Some essentials of decoding. *Exceptional Education Quarterly, 1,* 11-25.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children.* New York: Guilford.

Shinn, M. R., Tindal, G., & Spira, D. (1987). Special education referrals as an index of teacher tolerance: Are teachers imperfect tests? *Exceptional Children, 54,* 32-40.

Shinn, M. R., Tindal, G., & Stein, S. (1988). Curriculum-based assessment and the identification of mildly handicapped students: A research review. *Professional School Psychology, 3,* 69-85.

Smith, F. (1973). Decoding: The great fallacy. In F. Smith (Ed.), *Psycholinguistics and reading.* New York: Holt, Rinehart, & Winston.

Smith, F. (1975). *Comprehension and learning: A conceptual framework for teachers.* New York: Holt, Rinehart, & Winston.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360-406

Tilly, W. D. (1989). *Curriculum-based reading measures: Are they related to teacher perception of student skill?* Unpublished master's thesis, University of Oregon, Eugene, OR.

Tindal, G., Fuchs, L., Fuchs, D., Shinn, M. R., Deno, S., & Germann, G. (1983). *The technical adequacy of a basal reading series mastery test: The Scott-Foresman reading program* (128). Minneapolis: Institute for Research on Learning Disabilities.

Tindal, G., Fuchs, L., Fuchs, D., Shinn, M. R., Deno, S., & Germann, G. (1985). Empirical validation of criterion-referenced tests. *Journal of Educational Research, 74*, 203-209.

Wallace, G., & Larsen, S. (1979). *Educational assessment for learning problems: Testing for teaching.* Boston: Allyn & Bacon.

Woodcock, R. M. (1973). *Woodcock Reading Mastery Tests.* Circle Pines, MN: American Guidance Corp.

Woodcock, R. M. (1987). *Woodcock Reading Mastery Tests-Revised.* Circle Pines, MN: American Guidance Corp.

Ysseldyke, J. E., & Marston, D. (1982). A critical analysis of standardized reading tests. *School Psychology Review, 11*, 257-266.

Mark Shinn, PhD, is Associate Professor and Director, School Psychology Program, University of Oregon, Eugene, OR. His research interests include the use of Curriculum-Based Measurement in educational decision making and issues related to the reintegration of special education students into general education classrooms.

Roland Good III, PhD, is an Assistant Professor in the School Psychology Program at the University of Oregon. He received his doctorate from Pennsylvania State University in 1985. His research interests include the prevention of early academic problems and alternative service delivery systems.

Nancy Knutson, PhD, is Project Coordinator of the Oregon Divorce Study at the Oregon Social Learning Center. She received her doctorate in school psychology from the University of Oregon in 1990. Her research interests include use of Curriculum-Based Measurement in problem-solving assessment and the effects of divorce on family interactions.

W. David Tilly III, PhD, received his doctorate in school psychology from the University of Oregon in 1991. He is a school psychologist for the Heartland Area Education Agency. Research interests include Curriculum-Based Measurement and special education reform.

Vicki L. Collins is completing her doctorate in school psychology from the University of Oregon. Her interests include non-discriminatory assessment, functional problem-solving assessment with minority students, and reading instruction.