School Psychology Review,
2002, Volume 31, No. 4, pp. 540-553

Oral Reading Fluency and Prediction of Reading Comprehension in African American and Caucasian Elementary School Children

John M. Hintze, James E. Callahan III, William J. Matthews, and Stacy A. S. Williams
University of Massachusetts

Kevin G. Tobin
Pittsfield Public Schools, Pittsfield, Massachusetts

Abstract. This study examined the differential predictive bias of CBM in reading across African American and Caucasian students in Grades 2 through 5. Participants included 136 students who were administered CBM oral reading fluency passages and the Reading Comprehension subtest of the Woodcock-Johnson Psychoeducational Battery—Revised. A series of hierarchical multiple regression analyses indicated that CBM neither over- or underpredicted reading comprehension skills controlling for age, sex, and socioeconomic status. Results of this study suggest that CBM continues to appear to be a sensitive form of direct reading assessment in the local curriculum for both African American and Caucasian elementary-age students.

Curriculum-based measurement (CBM) is a set of standardized and specific measurement procedures that can be used to quantify student performance in the basic academic skill areas of reading, spelling, mathematics computation, and written expression (Deno, 1985; Deno & Fuchs, 1987; Fuchs & Deno, 1991; Shinn, 1989). CBM uses alternate-form test items of relatively equal difficulty selected from the general education curriculum for the purpose of making educational decisions (Deno, 1989). For example, in reading, connected text passages of similar difficulty are drawn from graded curricular materials and are administered individually to students to index their proficiency summatively (i.e., comparing an individual student's performance to grade level peers) and formatively (i.e., examination of an individual student's growth over time). A substantive body of literature exists demonstrating the efficacy of CBM and its use in screening, progress monitoring, program planning, and program evaluation (Shinn & Bamonto, 1998).

CBM has been offered as a direct form of academic assessment for the determination of educational need and resource allocation. With a focus on socially important variables that can be implemented on a frequent and continuous basis, CBM represents a dynamic form of assessment whereby the interaction between instruction and student skill development is monitored on a continual basis (Mercer & Ysseldyke, 1977; Ysseldyke & Regan, 1980). Moreover, because performance is assessed in the curriculum content of the local

Address correspondence to John M. Hintze, Ph.D., University of Massachusetts at Amherst, School Psychology Program, 362 Hills South, Amherst, MA, 01003. E-mail: hintze@educ.umass.edu.

education community, CBM avoids some of the often-cited criticisms of more commonly used, commercially available norm-referenced assessment techniques. Because assessment activities are developed directly from the school curriculum and student performance is referenced to locally derived standards, CBM has demonstrated high levels of diagnostic and treatment utility—contributing greatly to the enhancement of classroom instruction (Fuchs, 1993; Fuchs, Deno, & Mirkin, 1984; Fuchs, Fuchs, Hamlett, & Ferguson, 1992).

Surprisingly, although the research literature is replete documenting the psychometric properties of the CBM metric (cf. Marston, 1989), very few have examined possible sources of bias as a function of race or ethnic status. As such, the extent to which CBM is related to current or future performance on measures other than itself as a function of ethnicity is unknown. If CBM differentially predicted word recognition skills, reading comprehension performance, or any other related reading skill as a function of ethnicity, then the measure would be biased for one group over another. Testing for such bias involves the use of regression analysis as a means of predicting performance on a secondary measure (e.g., test of reading comprehension) on the basis of a person's CBM score. For example, if two groups of people differing only in ethnicity with exactly the same CBM reading scores are predicted to have systematically different reading comprehension scores (as predicted from the CBM scores) the conclusion would be that CBM scores are biased in their prediction of reading comprehension as a function of ethnicity.

To date, only one study has assessed the degree to which CBM reading fluency measures predict differential outcomes for Caucasian and African American students (Kranzler, Miller, & Jordan, 1999). In this study, 326 randomly selected students (225 Caucasians, 79 African Americans) in Grades 2 through 5 were administered grade-appropriate CBM reading passages and the Reading Comprehension section of the California Achievement Test. Using a series of simultaneous multiple regression analyses, results suggested that CBM in reading failed as an unbiased indicator of reading comprehension. Specifically, in Grades 4 and 5 CBM reading evidenced both intercept and slope bias for Caucasian and African American students. That is, CBM reading performance overestimated the reading comprehension abilities of African American students and underestimated the reading comprehension abilities of Caucasian students. As suggested by the authors, in practice such predictive bias could potentially lead to the systematic underidentification of compensatory programs for African American students due to the fact that CBM reading overpredicts their actual reading comprehension skills. Conversely, for Caucasian students CBM reading could potentially overidentify the need for compensatory remediation in reading comprehension because their CBM reading performance might underestimate their true reading comprehension abilities. Consequently, in order to safeguard against such bias, the authors suggest that users of CBM reading might consider adopting different cutoff scores or criteria for screening and identification decisions based on different groups of students.

Although the Kranzler et al. (1999) study served to highlight the importance of considering predictive bias with respect to CBM in reading, the design and subsequent data analytic plan present a number of limitations that limit the overall generalizability of the findings. In particular, the use of different CBM reading passage probes at each grade inhibited combining the sample into one regression or prediction model. Although in practice the use of grade-appropriate reading passages for screening and identification decisions is quite common, from a research perspective the use of different reading passages at each grade precludes any comparisons across grades. Furthermore, doing so necessitated conducting a separate multiple regression analysis at each grade level. As such, age as a critical developmental indicator was omitted from the analysis. Indeed, previous research has clearly indicated that developmental level (as measured by age or grade) is an extremely strong predictor of CBM oral reading fluency (Fuchs & Deno, 1992; Hintze, Shapiro, Conte, & Basile, 1997), exhib-

iting a linear relationship in Grades 1 through 5. Without a developmental indicator such as age in the multiple regression analysis competing for variability, any and all other variables have an increased chance of accounting for significant portions of variability due to chance. Second, the disparate sample sizes across the two groups may have likely led to unequal variances and possibly to a Type I error with respect to the influence of racial and ethnic factors.

The purpose of this study was to replicate and extend the work of Kranzler et al. (1999) in examining the differential predictiveness of CBM oral reading fluency on the reading comprehension skills of African American and Caucasian students. In particular, the study focused on an evaluation of: (a) the effectiveness of a theoretical model that combined the influence of developmental level, oral reading fluency, and socioeconomic status (SES) on the prediction of reading comprehension; and (b) whether differential prediction of reading comprehension would vary as a function of ethnicity. To answer each of these questions, a multiple regression model was first proposed that began with the combined influence of developmental level (i.e., age) and oral reading fluency (i.e., CBM performance) at the first step. These two variables were considered first on the basis of previous research that demonstrated that these two variables combined account for a sizable portion of observed variability in childrens' reading performance (Fuchs & Deno, 1992; Hintze, Owen, Shapiro, & Daly, 2000; Hintze et al., 1997). Next, the influence of SES was considered. The rationale for this variable in the second step of the model was based on the literature that suggests that children from families with limited resources or children from minority backgrounds are at greater risk for poorer reading outcomes than middle-class majority children (National Research Council, 1998; Spear-Swerling & Sternberg, 1998). Finally, ethnicity was last to enter the model, as it was hypothesized a priori that once developmental, reading, and SES variables were accounted for, ethnicity would explain very little of the variability in reading comprehension abilities. Once this first set of analyses was conducted, the second question was answered by building two separate but equivalent models, for each ethnic grouping variable (i.e., African American and Caucasian). In doing so, observed and predicted reading comprehension performance was compared as a function of ethnicity. As with the first research question, it was hypothesized a priori that neither slope nor intercept would be significantly different as a function of ethnicity.

Method

Participants and Setting

A total of 136 (66 male and 70 female; 65 African American and 71 Caucasian) students enrolled in Grades 2 through 5 located in a small urban school in the Northeastern United States served as participants in the study. A power analysis (Cohen, 1988; Hintze, 2000) was conducted indicating that a sample size of 130 would provide adequate power (.80) for main and interaction effects assuming a medium effect size (.15) and an alpha level of .05.

Overall, the school served approximately 560 students of which 73% were Caucasian, 23% African American, 3% Latino, and 1% Other (e.g., Asian, Native American). The median income of the participating students as measured by the Hollingshead Index of Socioeconomic Status (Hollingshead, 1975) was 40, thus reflecting a low- to middle-income level. The Hollingshead Index is a rank ordering of 480 occupations, each with an associated index score ranging from 10 (e.g., unemployed, bellhop) to 90 (e.g., aeronautical engineer, physician). Respondents are asked to indicate the profession that matches or most closely approximates their current line of employment. More specifically, approximately 47% of the participants' families fell within the low-income range, 33% within the middle-income range, and 20% within the high-income range. Participants were solicited by sending parental/guardian participation permission forms home with each child outlining the study and offering a $5 McDonald's™ restaurant gift certificate to those students who agreed to participate. Table

Table 1
Participant Sample's Sex and Ethnicity by Grade

		Grade			
		2	3	4	5
Sex					
Girls	*N*	18	19	18	15
	%	54.5	61.3	52.9	39.5
Boys	*N*	15	12	16	23
	%	45.5	38.7	47.1	60.5
Total	*N*	33	31	34	38
Racial/Ethnic Group					
African American	*N*	14	12	18	21
	%	42.4	38.7	52.9	55.3
Caucasian	*N*	19	19	16	17
	%	57.6	61.3	47.1	44.7
Total	*N*	33	31	34	38

1 presents the sample's sex, ethnicity, and SES by grade. Inferential statistics indicated no significant relationship between grade and sex, P^2 (3) = 3.54, p = .32; grade and ethnicity, P^2 (3) = 2.62, p = .45; grade and SES, P^2 (36) = 18.81, p = .99; or ethnicity and SES, P^2 (12) = 14.57, p = .27.

Measures

CBM reading probes. Using the methods outlined by Hintze and Shapiro (1997), material representative of third grade was selected from the Silver, Burdett, & Ginn (1991) reading series. Across the entire grade level, three passages of approximately 250 words were selected using a random number generator that indicated the page from which a potential passage was to be selected. For each potential reading probe, the first 100 words were evaluated for readability using the Spache (1953) formula. For a passage to be considered for selection, the readability score had to be within 1 month of the middle of third grade (i.e., 3.4 to 3.6). If a reading passage was beyond plus or minus 1 month from the middle of the third grade, then it was discarded. This sampling procedure continued until three reading passage probes falling within the prescribed readability ranges at the third-grade level were constructed.

Woodcock Johnson Psychoeducational Battery—Revised (WJ-R). The Woodcock-Johnson Psychoeducational Battery—Revised (WJ-R; Woodcock & Johnson, 1989) is a standardized comprehensive test of cognitive and academic achievement abilities. The Passage Comprehension subtest uses both a multiple-choice format requiring the subject to point to a picture represented by a phase and a modified cloze technique whereby the subject is asked to read a short passage and identify missing key words. For the primary and intermediate school population, internal consistency reliability coefficients for the Passage Comprehension subtest range from .883 at age 9

to .948 at age 6. Concurrent validity of the Broad Reading factor (i.e., combination of Letter-Word Identification, Passage Comprehension, Word Attack, and Reading Vocabulary) has been demonstrated with validity coefficients ranging from .633 to .857 with the Basic Achievement Skills Individual Screener (BASIS), Kaufman Assessment Battery for children (K-ABC), Kaufman Test of Educational Achievement (K-TEA), and Peabody Individual Achievement Test (PIAT). The Passage Comprehension subtest of the WJ-R was chosen because it is a widely used index of global reading comprehension and is commonly used by general and special educational support staff to assess generalized achievement.

Scoring Procedures

CBM oral reading fluency. The number of words read correctly per minute served as the datum for each reading probe. Words read correctly were those pronounced correctly within reading context. For example, the word *read* must have been pronounced "reed" not as "red" when presented in the context of "He will read the book." Repetitions and self-corrections were counted as correct.

The following types of errors were counted as words read incorrectly: (a) mispronunciations, (b) substitutions, and (c) omissions. Furthermore, if a student skipped an entire line, he or she was redirected and 1 error was recorded. Also, if a student struggled to pronounce a word or hesitated for 3 sec, he or she was provided the word by the examiner (for a detailed description of the exact scoring rules the reader is referred to the appendix of Shinn, 1989). Research strongly supports the reliability and validity of the CBM oral reading fluency metric as an index of a student's global reading functioning (cf. Marston, 1989).

WJ-R Passage Comprehension subtest. For each student, the raw score (i.e., correctly answered items and credited items per establishment of basal) served as the primary outcome datum for the Passage Comprehension subtest. Raw scores were not converted to standard scores to facilitate cross-age and grade comparisons. The Passage Comprehension subtest of the WJ-R was administered at the same time as the CBM oral reading fluency passages.

Procedures

Training procedures. Three doctoral level school psychology students were trained to administer and score the CBM oral reading fluency passages and the Passage Comprehension section of the WJ-R. After direct instruction and explanation of the written instructions, each data collector practiced administering and scoring the assessment materials. Training in the scoring of CBM reading probes was facilitated by using audiotaped passages read by students from a previous research study and compared to a master-scoring template. Each data collector listened to and scored a minimum of 10 CBM reading probes. Scores of at least 95% agreement with the master template were required for each data collector before initiation of the study.

Assessment administration procedures. Students were asked to read three grade level CBM oral reading fluency passages and complete the Passage Comprehension subtest of the WJ-R in one assessment setting. In addition, the reading passages were counterbalanced across students, within grade. All assessments were conducted in the morning of the school day over a 7-day school period in April. The duration of each assessment session ranged from approximately 15 to 25 min.

Standard CBM administration and scoring procedures were used for the oral reading fluency passages (Shinn, 1989). For each student, the data collector began by saying, "When I say 'Start,' begin reading aloud at the top of the page [data collector demonstrated by pointing]. Try to read each word, even those you do not know. Be sure to do your best reading. Do you have any questions?" The data collector then gave the student a copy of the first reading passage, made sure the stopwatch was ready, and instructed the student to begin reading aloud. The data collector marked errors on the corresponding scoring sheet. Separate scoring sheets were used for each student. At the end of 1 min, the data collector stopped the student. If the student was in the middle of a

Table 2

Hierarchical Regression Analysis of Age, CBM Oral Reading Fluency, SES, and Ethnicity on Reading Comprehension

Variables	WJ	Age	CBM	SES	B	S_{ee}	b
Age	.41				.124	.273	.054
	(.91)						
CBM	.65	.57			.007	.009	.633
	(.76)	(.82)					
SES	.26	.14	.28		.001	.011	.097
	(.96)	(.99)	(.96)				
Ethnicity	.12	.12	.00	.05	.979	.554	.120
	(.99)	(.99)	(1.00)	(.99)			
				Intercept = 17.121			
Mean	23.07	9.68	82.44	43.46			
(*SD*)	(4.20)	(1.24)	(37.17)	(25.22)			
					R^2 = .441		
					Adjusted R^2 = .423		
					R = .664		
					Standard Error of Estimate = 3.19		

Note. Values in parentheses associated with correlations represent standard errors.

sentence, she or he was allowed to finish; however, the data collector marked the last word read by the student at the end of the minute. The data collector then computed the number of words that the student read correctly during the minute. The median number of words read correctly across the three reading passages served as the summary datum for CBM reading performance. The median score controlled for any severe variance that might have occurred due to one extreme score.

The Passage Comprehension subtest of the WJ-R was administered in standard fashion as specified by the test manual. All participants were first administered Sample A. Once completed, starting points for Grades 2 and 3 and 4 and 5 were Items 5 and 9, respectively. Basal level performance was established by the six lowest numbered items administered that were correct. All items below the basal level were credited as correct. Moreover, ceiling level performance was established by the six highest numbered items administered that were failed. As such, testing did not proceed beyond the ceiling level. As indicated by the scoring manual and unless otherwise noted, only one-word responses were considered acceptable. If the student provided a two-word response or longer, he or she was asked for a one-word response. Once appropriate basal and ceiling levels were established, the student's raw score was calculated by subtracting the number of failed items from the highest numbered item attempted of the ceiling level. The raw score then served as the summary datum for each student's performance on the Passage Comprehension subtest.

Finally, background and demographic information were collected from parents in conjunction with consent for participation. As such, parents were asked to indicate their child's date of birth, sex, ethnicity, and complete the Hollingshead Index of Socioeconomic Status.

Results

A series of hierarchical multiple regression analyses were performed to determine whether age, CBM oral reading fluency, SES, sex, and/or ethnicity significantly predicted reading comprehension. In the first analysis, data from all 136 participants were analyzed. Descriptive statistics indicated that the data were normally distributed with the exception of SES, which was positively skewed and kurtotic (leptokurtic). Because this is not an uncommon finding with respect to SES information, the data were not transformed. Results of evaluation of assumptions indicated no violations of linearity, homoscedasticity of residuals, multicollinearity, or singularity. Using $p <$ .001 as the criterion for Mahalanobis distance, no outliers among the cases were found. Furthermore, no cases had missing data and no suppressor variables were found.

Table 2 provides the correlations between the variables, the unstandardized regression coefficients (B) and intercept, the standardized regression coefficients (β), the standard errors of estimates (S_{ee}) and R^2, and adjusted R^2 after entry of all four independent variables. At the first step, age and CBM oral reading fluency scores combined accounted for 42% of the observed variation in reading comprehension scores, $R^2 = .42$, F_{inc} (2, 133) = 48.38, $p < .001$. In Steps 2 and 3, SES and ethnicity were entered one at a time. Results at each step indicated that none of these variables contributed significantly to the prediction of reading comprehension scores. Specifically, the R^2 change for both SES and ethnicity was .01 of the overall final solution.

In addition, to test for the interaction of CBM and ethnicity, a separate set of analyses were conducted to test for slope and intercept bias. In order to test for the first, a hierarchical model was developed that regressed WJ-R Reading Comprehension scores on CBM, ethnicity, and the interaction of CBM and ethnicity. Again, at the first step, CBM oral reading fluency scores accounted for 42% of the observed variation in reading comprehension scores, $R^2 = .42$, F_{inc} (1, 134) = 96.70, $p < .001$. Neither ethnicity at Step 2 nor the interaction of CBM and ethnicity at Step 3 was significant and accounted for .01 and .02 of observed variation in the model, respectively. In the second analysis, an ANCOVA was conducted with ethnicity as the grouping variable with age, SES, and CBM oral reading fluency scores serving as covariates. Results of this analysis indicated no significant differences as a function of ethnicity controlling for age, SES, and CBM performance, F (1, 134) = 3.12, $p > .05$. Results of this analysis indicate that for all participants (i.e., both African American and Caucasian) only age and CBM oral reading fluency significantly predicted reading comprehension scores as measured by the WJ-R Reading Comprehension subtest. Neither SES nor ethnicity added significantly to the prediction of reading comprehension.

Next, in an effort to more fully explore the predictive bias of age, CBM, and SES, a series of hierarchical multiple regression analyses were conducted separately for African Americans and Caucasians. In so doing, each analysis was identical to the full model regression analysis with the exception that ethnicity was analyzed separately and did not appear as a step in the reduced models. Table 3 provides the correlations between the variables, the unstandardized regression coefficients (B) and intercept, the standardized regression coefficient (β), the standard errors of estimates (S_{ee}) and R^2, and adjusted R^2 after entry of all three independent variables for the African American sample. As with the full sample, all variables were normally distributed with the exception of SES, which was again positively skewed and leptokurtic. In addition, all assumptions of multiple regression were satisfied and no multivariate outliers were observed.

In the first step of the analysis, age and CBM oral reading fluency scores combined

Table 3

Hierarchical Regression Analysis of Age, CBM Oral Reading Fluency, and SES on Reading Comprehension Scores of African American Students

Variables	WJ	Age	CBM	SES	B	S_{ee}	b
Age	.46				.177	.379	.047
	(.89)						
CBM	.76	.57			.009	.014	.718
	(.65)	(.82)					
SES	.37	.22	.46		.001	.015	.032
	(.93)	(.97)	(.89)				
				Intercept = 14.159			
Mean	23.58	9.84	82.43	42.15			
(*SD*)	(4.64)	(1.24)	(36.81)	(29.29)			
					R^2 = .579		
					Adjusted R^2 = .558		
					R = .761		
					Standard Error of Estimate = 3.09		

Note. Values in parentheses associated with correlations represent standard errors.

accounted for 58% of the observed variation in the reading comprehension scores of African American participants, R^2 = .58, F_{inc} (2, 62) = 42.46, p < .001. At the second step, SES did not contribute significantly to the prediction of reading comprehension scores (R^2 = .001). Again, as with the full model analysis, the results of this reduced analysis indicated only age and CBM oral reading fluency scores combined significantly in predicting reading comprehension scores for African American students, and the influence of SES was negligible.

An analysis identical to the African American sample was conducted for the Caucasian sample. Table 4 provides the correlations between the variables, the unstandardized regression coefficients (B) and intercept, the standardized regression coefficient (β), the standard errors of estimates (S_{ee}) and R^2, and adjusted R^2 after entry of all three independent variables for the Caucasian sample. All variables were normally distributed with no outliers detected. Likewise, all assumptions of multiple regression were satisfied. As with the African American sample, results of the reduced model multiple regression analysis for Caucasian students indicated that only age and CBM oral reading fluency scores at Step 1 significantly predicted reading comprehension scores, R^2 = .30, F_{inc} (2, 68) = 14.23, p < .001], and that SES at Step 2 did not contribute significantly to the prediction of reading comprehension scores (R^2 = .003).

Finally, to examine whether the combined influence of age, CBM oral reading fluency, and SES were biased with respect to predicting reading comprehension ability across the two groups, separate z-tests were conducted for African American and Caucasian students comparing their multiple regression prediction results to the combined group average. Figure 1 presents the prediction lines for all three

Table 4

Hierarchical Regression Analysis of Age, CBM Oral Reading Fluency, and SES on Reading Comprehension Scores of Caucasian Students

Variables	WJ	Age	CBM	SES	B	S_{ee}	b
Age	.33				.006	.382	.021
	(.94)						
CBM	.54	.58			.005	.012	.527
	(.84)	(.81)					
SES	.10	.05	.08		.001	.018	.054
	(.99)	(.99)	(.99)				
					Intercept = 17.291		
Mean	22.61	9.54	82.45	44.65			
(*SD*)	(3.72)	(1.23)	(37.75)	(20.96)			
						R^2 = .298	
						Adjusted R^2 = .267	
						R = .546	
						Standard Error of Estimate = 3.19	

Note. Values in parentheses associated with correlations represent standard errors.

groups.[1] As can be seen, the prediction lines for both the African American and Caucasian students are very similar and closely aligned to that predicted for the overall group. Results of the *z*-test group comparisons indicated that African American and Caucasian students did not differ significantly with respect to either slope or intercept as compared to the overall group prediction. Furthermore, when compared directly, neither group differed with respect to slope or intercept. The results of this analysis suggest that each of the two separate groups (i.e., African American and Caucasian) belonged equally to the same overall parent population and that no differential predictive bias of reading comprehension scores was evident based on a student's age, CBM oral reading fluency abilities, SES, or ethnicity.

Discussion

The purpose of this study was to examine the differential predictiveness of CBM oral reading fluency on the reading comprehension skills of African American and Caucasian students. Results suggested that the CBM oral reading fluency metric was not biased with respect to ethnicity or SES. That is, once developmental level (i.e., age), oral reading fluency, and SES were accounted for, ethnicity predicted very little with respect to reading comprehension abilities. Moreover, the data were consistent with the extant empirical literature, indicating that CBM in reading is a strong indicator of generalized reading performance including reading comprehension (Deno, 1985; Deno, Mirkin, & Chiang, 1982; Hintze et al., 1997).

Although these findings are in contrast to those of Kranzler et al. (1999), it would appear premature to qualify CBM oral reading fluency measures as totally without bias. Although no differences were observed between the individual slopes and intercepts across the two groups, in the current study a model based

on age, CBM oral reading fluency, and SES explained much more in the way of observed variation for African Americans then it did for Caucasians (upwards of 25% more variation). These findings, combined with those of Kranzler and colleagues, clearly point to the fact that other variables must be operating in predicting reading comprehension from CBM performance across students of varying ethnic backgrounds. In examining the patterns of correlations it would appear that for African American students the relationship between CBM and reading comprehension performance was much stronger than for Caucasian students. Indeed, the magnitude of the differences between these variables explains nearly all of the difference in the R^2 values between the two models. Whether this relationship is real or just an artifact of the current sample remains to be seen.

So, which model is better? Is CBM a biased predictor of reading comprehension? Currently, the answer to each of these questions would appear to be "neither." At least one plausible explanation for the differences in findings between the current study and those of Kranzler and colleagues (1999) lies in the developmental nature of the methodology and analysis. Because all students were assessed with common CBM reading passages, developmental level, represented by age, was allowed to enter into the regression analysis. Without doubt, oral reading fluency is highly developmental and increases as a function of age (Fuchs & Deno, 1992; Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993; Hintze, Daly, & Shapiro, 1998; Hintze & Shapiro, 1997). Indeed, Hintze et al. (2000) found that the combination of grade and oral reading fluency as measured with CBM reading passages accounted for upwards of 70% of observed variation in reading performance. Omitting the critical variable of developmental level from any analysis using CBM in reading increases the likelihood that other variables, though they may be associated with reading comprehension at some level, have a

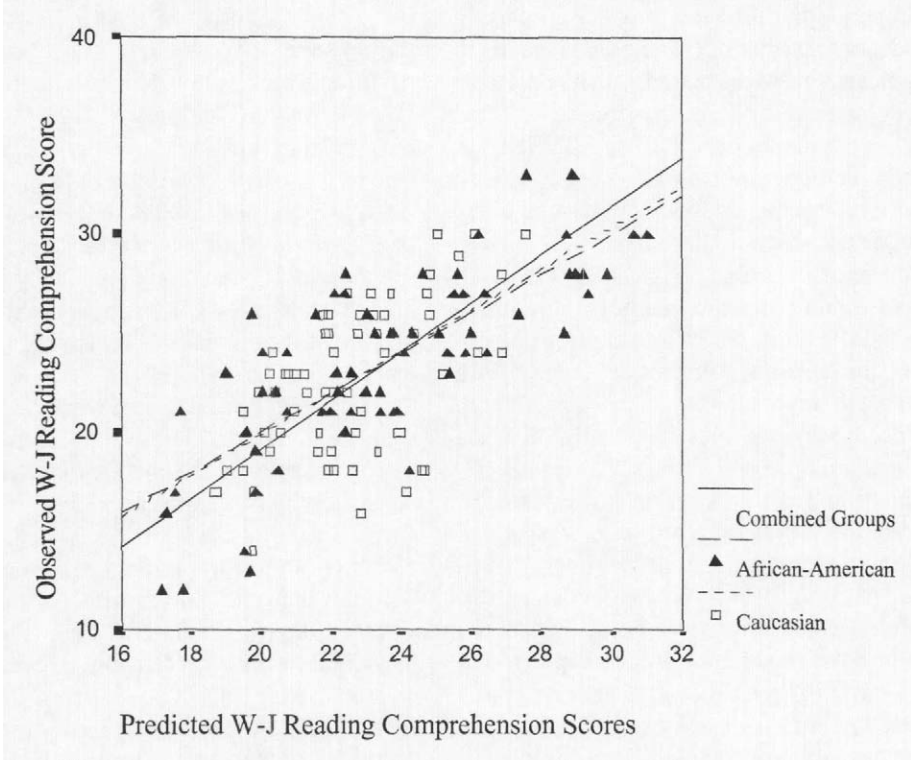

Figure 1. Scatter plot of regression lines for full model (Combined Groups) and reduced models by ethnicity (African American and Caucasian students).

greater chance of explaining significant amounts of variance. Allowing developmental level and age into the model decreases the total amount of variance available for which other variables are able to compete. As such, variables that appear significant in one model may account for negligible amounts of variance in another. Results of the current study support previous findings suggesting that age is a critically important determinant of reading fluency.

As a result, this study found that the CBM oral reading fluency metric was not a biased indicator of reading comprehension as a function of ethnicity. As stated by Jensen (1980) and elaborated by Kranzler et al. (1999), tests are unbiased to the extent that both majority and minority groups share the same or similar regression lines of prediction. That is, CBM oral reading fluency scores predicted the same or similar criterion reading comprehension scores regardless of whether a student was African American or Caucasian. The apparent lack of over- or underprediction across groups as reported in this study provides conflicting evidence to the notion that the CBM oral reading fluency metric is biased with respect to its ability to predict similar constructs.

For practitioners who use CBM in reading in their educational assessment practices, findings from this study suggest that relatively similar decisions would be made for African American and Caucasian students alike when trying to predict reading comprehension skills from CBM information. Moreover, the lack of significant differences with respect to intercept diminishes concern with respect to content bias. That is, in this case it does not appear that the CBM test development and administration procedures favored the majority group with respect to indexing absolute oral reading fluency abilities. Although the type of curriculum in which the students were being instructed might have moderated this, it appears that African American and Caucasian students performed alike in their oral reading abilities, irrespective of their comprehension abilities.

At a more general practice level, results of this study also lend further support to the use of CBM in reading as part of a larger problem-solving educational decision-making model (Deno, 1989). Using the data in the current study, practitioners should arrive at similar norm-referenced decisions regardless of whether a student was African American or Caucasian. The use of tests in educational evaluation continues to be controversial. Adopting the use of psychometrically sound, unbiased tests can help address the alleged discrimination in assessment, disability classification, and special education placement of ethnic or racial minorities. Results of the current study support the use of CBM in reading for making screening, referral, and resource allocation decisions at a nomothethic level for students in Grades 2 through 5 regardless of whether they were African American or Caucasian.

Although the current study has attempted to build on and extend the current knowledge base with respect to CBM and bias, the results must nevertheless be interpreted in light of a number of limitations that limit the generalizability of the findings. First, with respect to the issue of bias, the results of the current study do not indicate that CBM in reading is totally without bias. The current study has focused solely on the differential predictive validity of the measure. And, as previously indicated, it would appear that the current model left open many other variables that could potentially explain reading comprehension performance in a way that might suggest that CBM reading is indeed a biased indicator.

Though it would appear that the current study exercised adequate levels of internal control and validity, arguments for threats to construct validity of putative cause and effect and external validity can be made. Indeed, all the variables in the current study were measured with only one type of instrument. Although the current study operated under the assumption that reading performance could be accurately measured with fluency and short-answer/response formats, such measures may represent a mono-operation and method bias. If this were the case, the construct validity of the results could be called into question. Moreover, se-

lection bias with respect to the participants may have been evident. As reported, although mean differences for SES between the two groups were not evident, the distribution of SES within groups was not the same. That is, the range of variability for Caucasian students was greater than that of the African American students. Results of the current study represent the performance of one ethnically diverse school in the Northeast. It remains an empirical question as to whether similar results would be obtained in other schools, districts, or regional locales. Replicating the study with samples that better represent SES patterns in the larger population might lead to totally different results. Nonetheless, selection bias and its interaction with setting could have affected the current results.

Second, both this study and the Kranzler et al. (1999) study focused only on using CBM oral reading fluency measures for making between-individual decisions. Although using CBM in this manner is consistent with three of the five steps in the problem-solving model, neither study has addressed whether bias in outcomes would be observed over time using CBM to monitor the progress of students for the purpose of making within-individual decisions. Although some research has suggested differential outcomes as a result of students' knowledge about are being timed (Derr & Shapiro, 1989; Derr-Minneci & Shapiro, 1992), other mediating variables such as examiner and test demand characteristics have not been fully explored. For example, would African American students perform similarly or differently for African American examiners as compared to Caucasian examiners? Would differences be observed in assessing the student within the classroom as opposed to a pull-out situation? Finally, what are the social consequential effects of using local norms in the decision-making process for students from minority groups? It is not uncommon to observe demographic differences across schools within the same district. What effects does this have on educational decision making? Should local norms be developed at the school level or aggregated at the district level? These are just a sampling of the types of questions that need to be explored with respect to the use of CBM and bias.

Finally, a third limitation rests in the nature of the CBM reading passages and how they were used in the current study. Although for research purposes it is not uncommon to use common sets of reading passages across grade levels (cf. Fuchs & Deno, 1992; Hintze et al., 1997), in practice students initially would be administered grade-appropriate passages and then passages of increasing and decreasing difficulty for the purposes of identifying and certifying academic skill weaknesses. To do so, however, would eradicate a researcher's ability to isolate age as a critical variable. Although controlling the passages in this manner improved the internal validity of the current work, clearly the generalizability of the results is limited to school settings where this would not occur.

In summary, although far from conclusive, results of the current study suggest that CBM in reading does not evidence differential predictive bias of reading comprehension skills across African American and Caucasian students in elementary grades 2 through 5 when age is taken into account as an important consideration. Importantly, research in the area of CBM must acknowledge the fact that the unit of analysis—oral reading fluency—is developmental in nature and, therefore, must be accounted for in any prediction analysis. Notwithstanding previous research, it appears that CBM in reading, as part of a problem-solving educational decision-making model, does not provide biased findings as a function of ethnicity. Although continued research in this area is sorely needed, CBM as a direct form of academic assessment continues to appear to be a sensitive form of assessment in the local curriculum for both African American and Caucasian elementary-age students.

Footnotes

[1] The reader should note that the intercepts in Figure 1 appear visually different from the intercepts reported in Tables 2 through 4 because the intercepts in Figure 1 are a result of the bivariate correlation between observed and predicted WJ scores, not the full hierarchical model.

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Lawrence Erlbaum Associates.

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

Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 1-17). New York: Guilford.

Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19,* 1-15.

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

Derr, T. F., & Shapiro, E. S. (1989). A behavioral evaluation of curriculum-based assessment of reading. *Journal of Psychoeducational Assessment, 7,* 148-160.

Derr-Minneci, T. F., & Shapiro, E. S. (1992). Validating curriculum-based measurement in reading from a behavioral perspective. *School Psychology Quarterly, 7,* 2-16.

Fuchs, L. S. (1993). Enhancing instructional programming and student achievement with curriculum-based measurement. In J. J. Kramer & J. C. Conoley (Eds.), *Curriculum-based measurement* (pp. 65-103). Lincoln, NE: University of Nebraska-Lincoln, Buros Institute of Mental Measurements.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57,* 488-500.

Fuchs, L. S., & Deno, S. L. (1992). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58,* 232-242.

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,* 449-46.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Ferguson, C. (1992). Effects of expert system consultation within curriculum-based measurement using a reading maze task. *Exceptional Children, 58,* 436-450.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22,* 27-48.

Hintze, J. (2000). *PASS power analysis.* Kaysville, UT: NCSS Statistical Software.

Hintze, J. M., Daly, E. J., & Shapiro, E. S. (1998). An investigation of the effects of passage difficulty level on outcomes of oral reading fluency progress monitoring. *School Psychology Review, 27,* 433-445.

Hintze, J. M., Owen, S. V., Shapiro, E. S., & Daly, E. J. (2000). Generalizability of oral reading fluency measures: Application of G theory to curriculum-based measurement. *School Psychology Quarterly, 15,* 52-68.

Hintze, J. M., & Shapiro, E. S. (1997). Curriculum-based measurement and literature-based reading: Is curriculum-based measurement meeting the needs of changing reading curricula? *Journal of School Psychology, 35,* 351-375.

Hintze, J. M., Shapiro, E. S., Conte, K. L., & Basile, I. M. (1997). Oral reading fluency and authentic reading material: Criterion validity of the technical features of CBM survey-level assessment. *School Psychology Review, 26,* 535-553.

Hollingshead, A. B. (1975). *Two-factor index of social position.* Unpublished manuscript, Yale University, New Haven, CT.

Jensen, A. R. (1980). *Bias in mental testing.* New York: The Free Press.

Kranzler, J. H., Miller, M. D., & Jordan, L. (1999). An examination of racial/ethnic and gender bias on curriculum-based measurement in reading. *School Psychology Quarterly, 14,* 327-342.

Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What is it and why do it. In M. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford.

Mercer, J. R., & Ysseldyke, J. E. (1977). Designing diagnostic-intervention programs. In T. Oakland (Ed.), *Psychological and educational assessment of minority children* (pp. 70-90). New York: Brunner/Mazel.

National Research Council. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

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

Shinn, M. R., & Bamonto, S. (1998). Advanced applications of curriculum-based measurement: "Big ideas" and avoiding confusion. In M. R. Shinn (Ed.), *Advanced application of curriculum-based measurement* (pp. 1-31). New York: Guilford Press.

Silver, Burdett, & Ginn. (1991). *World of reading.* Morristown, NJ: Author.

Spache, G. (1953). A new readability formula for primary grade materials. *Elementary English, 53,* 410-413.

Spear-Swerling, L., & Sternberg, R. J. (1998). *Off track: When poor readers become "learning disabled."* Boulder, CO: Westview Press.

Woodcock, R. W., & Johnson, M. B. (1989). Woodcock-Johnson Psychoeducational Battery-Revised. Allen, TX: DLM Teaching Resources.

Ysseldyke, J. E., & Regan, R. (1980). Non-discriminatory assessment: A formative model. *Exceptional Children, 46,* 465-468.

John M. Hintze received his Ph.D. in School Psychology from Lehigh University in 1994 and is an Associate Professor of School Psychology at the University of Massachusetts at Amherst. His primary research interests are in the areas of school-based assessment, research design and methodology, and data analysis.

James E. Callahan III received his M.Ed. CAGS in School Psychology from University of Massachusetts at Amherst in 1997 and is currently a Ph.D. candidate in School Psychology at University of Massachusetts at Amherst. His primary research interests are in the areas of school-based assessment, system-wide intervention and prevention programming, program analysis, and skill-based interventions for minority students.

William J. Matthews received his Ph.D. in Clinical Psychology from the University of Connecticut in 1980 and is a Professor in the School Psychology Program at the University of Massachusetts at Amherst. His primary research interests are in the areas of teacher expectancy, school-based behavioral and academic assessment, and research design.

Kevin G. Tobin received his Ph.D in Clinical Psychology from the University of Rhode Island in 1988 and is a Clinical Psychologist with Pittsfield Public Schools in Pittsfield, MA. His primary research interests are in the areas of school-based assessment, the prevention of violence, drug use, and teenage pregnancy, and the effectiveness of different curricula and instructional methods on student learning.

Stacy A. Williams is a doctoral candidate in the School Psychology Program at the University of Massachusetts at Amherst. Her primary research interests are in the areas of African American identity development and methods to enhance the academic and social-emotional development of minority students.