School Psychology Quarterly, Vol. 20, No. 3, 2005, pp. 304–317

Story Retell: A Fluency-Based Indicator of Reading Comprehension

Greg Roberts
The University of Texas at Austin

Roland Good
University of Oregon

Stephanie Corcoran
University of Alabama

This article presents a fluency-based measure of reading comprehension. A part of the Vitals Indicators of Progress (VIP) system, the measure outlined here represents an alternate form to the retell-fluency measure in the Dynamic Indicators of Basic Early Literacy System (DIBELS). Measures of retell fluency provide an efficient, fluency-based tool for teachers to use in identifying the handful of children whose oral reading fluency may not adequately represent comprehension. When used in tandem with oral fluency measures, retell measures provide a vehicle for more reliably targeting classroom and school-level resources and for maximizing the efficiency and effectiveness of early reading instruction.

Twenty years of research in early reading instruction underscore the importance of assessing student progress at regular intervals throughout the school year (Deno, 1985; Deno, 1992; Shinn, 1998). Successful teachers systematically collect and use data on students' abilities with important early literacy skills (Fuchs & Fuchs, 1999) to plan and alter instruction, at the group level and for individual students (Deno, 1986; Fuchs, Fuchs, & Hamlett, 1989; Stoner, Scarpati, Phaneuf, & Hintze, 2002). For these teachers, decisions about what to teach and to whom, using which grouping strategy, etc. are based on evidence collected during these recurring assessment periods. Curriculum-based measurement offers a useful platform for this type of teacher activity.

Curriculum-based measurement is a time- and resource-efficient, classroom-

Address correspondence to Greg Roberts, Ph.D., The University of Texas at Austin (D0100), Austin, TX 78712; E-mail: gregroberts@mail.utexas.edu

oriented means of indexing academic standing at points in time and of quantifying progress over time (Deno, Fuchs, Marston, & Shin, 2001). Curriculum-based systems provide reliable and relevant data for estimating status and charting growth (Deno, 1985; Fuchs & Deno, 1991; Marston, 1989), for making instructional decisions and refining instructional plans (Marston, Mirkin, & Deno, 1984), and for evaluating program effectiveness (Deno, 1986; Stoner et al., 2002). A growing body of research supports the technical adequacy of curriculum-based measurement (e.g., Fuchs & Deno, 1991; Good & Jefferson, 1998; Markell & Deno, 1997).

This article presents a curriculum-based measure of early reading comprehension using a retell fluency format. The retell fluency measure is one element of the more comprehensive Vital Indicators of Progress (VIP) used by a number of large school districts across the country to monitor the progress of early and beginning readers. Vital Indicators of Progress (VIP) are part of the Voyager Universal Literacy (VUL) Program, which is a research-based program for teaching reading in kindergarten through third-grade classrooms. VIP was developed by Dr. Roland Good of the University of Oregon as an alternate form of the Dynamic Indicators of Basic Early Literacy (Kaminski & Good, 1996), which is widely used to monitor early reading progress within classroom settings. The Dynamic Indicators of Basic Early Literacy System (DIBELS) and VIP are based on the ongoing program of research of Good, Kaminski, and their colleagues at the University of Oregon, and the benchmarks that serve as foundations of both systems are based on the most current work on early literacy.

CURRICULUM-BASED MEASUREMENT OF READING COMPREHENSION

The most widely used curriculum-based measure of reading competence and reading comprehension is oral reading fluency (Good, Simmons, & Kame'enui, 2001). Generally operationalized as the number of words read correctly in one minute (Fuchs & Fuchs, 1999), oral reading fluency is a reliable and remarkably efficient predictor of elementary-school students' scores on more traditional (and more lengthy) measures of reading ability (Fuchs & Deno, 1991; Fuchs, Fuchs, Hosp, & Jenkins, 2001) and reading comprehension (Fuchs, Fuchs, & Maxwell, 1988; Jenkins, Fuchs, Espin, van den Broek, & Deno, 2000), with bivariate correlations generally in the .6 to .9 range (Deno, Mirkin, & Chiang, 1982; Good & Jefferson, 1998). Performance on measures of oral reading fluency is widely used for intra- and inter-individual comparisons (Fuchs et al., 2001) and may represent a more sensitive indicator of individual and group change than more broadly conceptualized (i.e., the commercially available measures of reading) measures of reading comprehension (Marston, Fuchs, & Deno, 1985). There is also evidence suggesting that oral reading reading fluency scores may reliably forecast performance on high-stakes, state-mandated tests of

reading ability (Crawford, Tindal, & Stieber, 2001; Buck & Torgeson, 2003), though this area of research continues to develop.

Information processing models provide a framework for understanding the relationship of oral reading fluency and reading comprehension are generally discussed. The prevailing theory is that decoding and comprehension both require cognitive resources, and the more processing capacity devoted to decoding, the less capacity available for understanding what has been decoded (LaBerge & Samuels, 1974; Stanovich, 2000). Recent extensions (Perfetti, 1995) of this view suggest that comprehension-related tasks may be subject to varying levels of automaticity, much like word recognition and decoding-oriented tasks. Inferential processes, for instance, may become increasingly automatic, freeing up additional cognitive resources for even deeper levels of text processing (Fuchs et al., 2001; Logan, 1997), suggesting that fluency may be more than a proxy for reading competence and comprehension; instead, oral reading fluency may relate to reading ability in more substantive ways (Nathan & Stanovich, 1991).

ORAL READING FLUENCY AND RETELL FLUENCY

Whichever cognitive mechanism underlies the relationship of oral fluency and reading comprehension, the reality is that children who read with greater speed and accuracy (i.e., with greater fluency) tend to read more skillfully and with greater comprehension than children who read less fluently. Given that oral reading fluency provides an efficient and reliable indicator of reading ability and progress, it is reasonable to question the need for another fluency-based measure of reading comprehension. The case for a measure of retell fluency is based on several considerations: (1) teachers' and administrators' concerns with the face validity of oral reading fluency as a measure of comprehension, (2) concern with teaching and practicing a *misrule,* and (3) a desire to provide direct linkage to the National Reading Panel's report, *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction* (National Reading Panel, 2000). These are discussed in the following paragraphs.

Face Validity of Fluency Measures and Its Impact on Endorsement and Use

Oral reading fluency as a measure of reading comprehension enjoys considerable support among substantive and applied researchers working on topics related to curriculum-based measurement. Practitioners, however, appear to be less convinced (Shinn, Good, Knutson, Tilly, & Collins, 1992). Indeed, concern with the face validity of oral reading fluency as a measure of comprehension is an acknowledged barrier to its wider use by classroom teachers (Shinn, et al., 1992), a reality that may persist until teachers are presented with evidence beyond what may be available in scholarly sources (Shinn et al., 1992). A potential

source of such evidence is "hands-on" experience with both a fluency-based measure that more closely resembles traditional comprehension-related tasks *and* the traditional oral reading task. Asking teachers to collect data on both types of measures will provide opportunities to directly observe the considerable correspondence of retell and oral reading fluency. Over time, teachers may grow more confident in the validity of oral reading fluency *without* retell as an equally reliable and more efficient indicator of student ability (Fewster & MacMillan, 2002) and begin to rely more directly on its use for monitoring students' reading progress.

Concerns with the validity of oral reading fluency are reflected in the prevalence of teacher-reported cases of *word calling,* where children read aloud as fluently as same-aged peers but "can't understand a word of what they read." While reliable estimates of its prevalence are not available (excluding cases of hyperlexia and related spectral-disorder conditions), cases where fluency is unrelated to comprehension are thought to be rare, with teacher reports likely representing a considerable overestimate of actual incidence (see Hamilton, 2001). A retell fluency measure may help to clarify for teachers the relationship of oral reading fluency and comprehension by identifying students' abilities on a more comprehension-like task within a fluency framework. A retell fluency measure may also help to identify the relatively few children who, in fact, fit the pattern where reading comprehension is not consistent with oral reading fluency. Measures of retell fluency may also provide a means for beginning to generate reliable estimates of this pattern in the population of beginning and early readers.

The proposal is *not* that retell fluency replace oral reading fluency; evidence in favor of the eminence of oral reading fluency as an index of reading comprehension and reading competence is compelling. Instead, retell fluency is proposed as an additional tool for teachers to use in diagnosing and addressing the instructional needs of struggling students. If it also serves as a means for increasing teacher support for fluency-based measures in general, and for oral reading fluency in particular, all the better.

Ancillary Effect of Fluency Measures and Its Impact on Students

There is evidence that regular use of fluency-based measures may have an indirect effect on student learning (Good et al., 2001), whether by prompting students' attention or by directing teachers' instructional focus to skills that underlie fluency measures of reading comprehension. Assuming an effect, it becomes important that children understand (or, at least, not *mis*understand) the lesson that may be embedded in fluency-based tasks. This concern, that students (and teachers) may focus too intently on the rate aspect of the fluency equation (i.e., accuracy/rate), thus sacrificing accuracy, or that they may see fluency as an end rather than a means, perhaps to the detriment of comprehension, warrants research. In the interim, a fluency measure that includes a task more readily identifiable as comprehension-oriented may provide some protection against the possibility of learning or practicing a misrule.

Linkage with Report of the National Reading Panel

A final point to make in defense of retell fluency concerns an interest in providing an explicit link to the core components in the National Reading Panel's (2000) report. The report is clear on the elements of early reading and early reading instruction, and VIP and DIBELS map explicitly onto the first three (alphabetic principle, phonemic awareness, and fluency) of the *five big ideas.* Retell fluency provides an explicit link that corresponds to the fourth core component, reading comprehension.

RETELL VERSUS OTHER "COMPREHENSION-LIKE" FORMATS

Story retell is preferred over other comprehension-like formats (e.g., cloze and question-response) for several reasons (see also Fuchs et al., 2001). First, production-type responses potentially increase the variability in a set of scores. In general, the more behavior a child generates representing the construct of interest, the greater the chances of detecting post-intervention change, and retell produces a large-sized sample of comprehension behaviors. Cloze and question-response generate much less behavior.

A second reason for preferring retell is its time efficiency. Cloze formats require reading a great deal of text for every scorable response (i.e., a filled in blank). Also, cloze requires the child to perform a second task, apart from oral reading (i.e., the passage used as a measure fluency cannot be used for the cloze task), increasing testing time, dramatically in many cases. While question-response formats can use the same passage used to measure fluency, the student generally has to complete the entire passage before answering the questions, a lengthy, punishing ordeal for some students.

The difficulty in writing good questions is another reason for preferring retell to question-response. Writing good questions can be difficult for several reasons. One problem is background knowledge—student responses to questions in traditional question-response formats may be as related to their levels of background knowledge as to their comprehension of the reading passage. Verbal ability may also predict a unique portion of variance in response scores, apart from comprehension skill. Further, curriculum-based measures, when used to monitor student progress on a given skill, require repeated administrations of *equivalent* measures. Given the difficulty of creating two parallel tests using a question-response format, producing multiple equivalent forms may be problematic.

Finally, retell fluency as a measure of reading comprehension is more easily linked to instruction than question-response and cloze formats. Retelling can be taught, modeled, and practiced more easily than cloze and question-response tasks. Retell may also fit more readily into an instructional sequence. If one assumes that higher-level comprehension skills rely on more basic abilities (Per-

fetti, 1995), being able to retell what one has read may represent a critical gateway to more sophisticated levels of text comprehension.

METHOD

Participants

The data used in this study were collected during the administration of post-test measures in the evaluation of an early reading intervention program being used in an urban, southeastern U.S. school district. First graders from six schools were included in the sample. Participating schools served low-income, Title I populations. Cases with valid data on the *Woodcock Diagnostic Reading Battery* and on the VIP measures, including the measure of retell fluency, were included in the analysis. A total of 86 such cases was available. Ninety-percent of students were African American and all received free or reduced lunches.

Measures

Reading passages, retell protocols, and scoring procedures. Two passages were developed for this study. First drafts of each passage were analyzed using the MicroPower & Light Software system to estimate readability. All available indices (Dale-Chall, Flesch, FOG, Powers, SMOG, FORCAST, Frye, and Spache) were computed, and the Spache readability index was used to revise and refine passages. Target readability for passages was 2.0 to 2.3 on the Spache scale. Passages that were too difficult were revised by substituting single-syllabic words for multi-syllabic, by substituting high-frequency words for potentially difficult words, and by shortening long sentences. Passages below the target range were revised in the more difficult direction using the same three considerations.

The outcome of primary interest was the number of words used by a child in correctly retelling the story. Test administrators estimated the number of words used by each child in accurately recounting the passages by discretely moving a finger across a sequence of evenly spaced numerals according to the pace established by a given child's appropriate retelling of what was read. The tracking procedure ended when a child stopped retelling the story. The procedure paused when he or she provided inaccurate or irrelevant information. The total word count did *not* include the number of words used in inaccurately retelling the story or in telling a story other than the one read.

Woodcock Diagnostic Reading Battery. The Woodcock Diagnostic Reading Battery (WDRB) is a widely used measure of student reading achievement. It includes ten subtests organized into seven reading and reading-related clusters. For purposes of the larger evaluation, the letter-word identification, word attack, and passage comprehension subtests were administered. The other subtests, including the measure of reading vocabulary, were not administered due to the limited

time allowed for the testing of students. Scores on letter-word identification and passage comprehension combine to yield a measure of the Broad Reading Cluster, with median reliability of .95 (Woodcock Diagnostic Reading Battery: Examiner's Manual, 1997). Scores from letter-sound identification and word attack represent the Basic Reading Skills Cluster, and the median reliability is .96. The Reading Comprehension Cluster, which requires scores from the reading vocabulary subtest, was not calculated. Readers interested in details on the WDRB's properties are referred to the Examiner's Manual (1997).

Procedures

As described elsewhere, data were collected in May 2002 as part of a larger evaluation of Voyager's Universal Literacy program, a beginning reading program being used in a large urban school district in the southeastern United States. Six schools participated in the evaluation. Local school officials selected participating schools. All schools were Title I-qualified. Individual students were pulled from classrooms for approximately 20 minutes to complete the assessments. The oral reading and retell fluency measures of the VIP were administered along with the letter-word identification, word attack, and passage comprehension subtests of the WDRB.

Two advanced-level graduate students collected the data used in this study. The administrators were trained on administration of VIP oral reading fluency, VIP retell fluency, and the WDRB. Interrater reliability on oral reading and retell fluency was informally assessed using a series of four passages administered to four different children. Administrations by the lead administrator were audiotaped, and the second administrator scored the audiotaped sessions. Comparison of the respective protocols suggested a very high level of agreement across administrators; when comparing students' status (i.e., low risk, some risk, high risk), agreement was 100%.

RESULTS

VIP Oral Reading Fluency and Retell Fluency Reliability

Two VIP passages were included in the study to obtain an estimate of the alternate-form reliability of the VIP oral reading fluency and VIP retell fluency. The alternate-form reliability of VIP oral reading fluency was .89 for one passage. Using the Spearman-Brown prophecy formula, the reliability of three passages was estimated to be .96. Since the recommended use of the VIP oral reading fluency is to administer three passages and use the middle score, the estimated reliability of three passages is the most appropriate reliability to consider. The reliability of VIP oral reading fluency for one passage is sufficient for screening, and the reliability of the aggregate of three VIP oral reading fluency passages is sufficient for important educational decisions.

The alternate-form reliability of the VIP retell fluency was .57. The estimated reliability of three passages was .80, sufficient for screening decisions (Salvia & Ysseldyke, 1988). When seven passages were aggregated, for example in progress monitoring, the estimated reliability of VIP retell fluency was .90. In this study two passages were averaged, with an expected reliability of the average being about .73, sufficient for group or research purposes. For practice, the recommendation is to administer VIP retell fluency passages during the benchmark assessment. For students whose performance raises concerns, two retests with three additional VIP retell fluency passages each would yield a very high reliability score ($r = .92$).

VIP Oral Reading Fluency and Retell Fluency Validity

The most interpretable validity coefficients to consider are concurrent, criterion-related validity with respect to the WDRB Broad Reading Cluster at post-test. The individual passages correlated .75 and .72 with the Broad Reading Cluster, and the average of two VIP oral reading fluency passage scores correlated .76 with the Broad Reading Cluster, explaining about 57% of the variance in Broad Reading Cluster standard scores.

The individual VIP retell fluency passage scores correlated .47 and .43 with the Broad Reading Cluster, and the average of the two VIP retell fluency passages correlated .51 with the Broad Reading Cluster (26% of the variance explained) and .61 with the VIP oral reading fluency average. It is likely that these validity coefficients are attenuated by the reliability of the VIP retell fluency measure, and that the validity of the median of three VIP retell fluency measures administered in a benchmark assessment would be higher.

When VIP oral reading fluency and retell fluency were considered together in predicting Broad Reading Cluster standard scores, adding retell fluency contributed a very small amount of additional explained variance, 58% combined vs. 57% explained by VIP oral reading fluency alone.

VIP Retell Fluency Comprehension Check

One purpose of VIP retell fluency is to provide a check on oral reading fluency as a measure of reading proficiency for children who may be fluent readers but may be unable to recount what they read. The first step in this analysis was to examine the relation between VIP oral reading fluency and VIP retell fluency (see Figure 1).

The open dots in the area bounded by the solid line represent cases where the relationship of oral reading fluency and retell fluency may be less than theory would suggest. Cases falling within the boundary line are children who read at least 40 words correctly per minute, but had a retell score of no more than .25 of their oral reading fluency score. Most pronounced are the two children reading 78 and 101 words correctly per minute, but who are retelling about four words on average about what they had just read (Figure 1). The area of concern is ini-

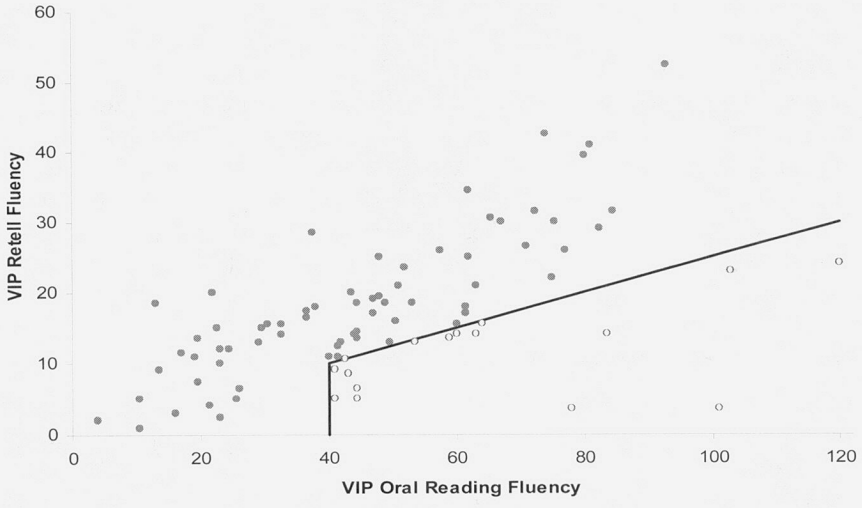

Figure 1. Relation between VIP oral reading fluency and VIP retell fluency showing area of concern about children's comprehension.

tially established on a rational basis. It seems important that students be able to tell about what they have read. When there is a large inconsistency between what they are able to read aloud from a passage and what they are able to tell about what they have read, it seems an instructionally relevant concern.

If the VIP retell fluency is providing a comprehension check, then the relation between VIP oral reading fluency and the Broad Reading Cluster should be stronger for children with consistent retell than for children whose retell is not consistent with their oral reading fluency. The relation between VIP oral reading fluency and Broad Reading Cluster scores for the two groups of children is illustrated in Figure 2.

Consistent with the prediction, VIP oral reading fluency displayed a stronger relation with the Broad Reading Cluster for students with consistent retell (r = .81, 65% of variance explained) than for students with inconsistent retell (r = .42, 17% of variance explained), and the difference in relation was significant, $t(1) = -3.30$, $p < .01$. However, several caveats are important to mention.

First, some of the difference in relation may be due to the restriction in range for the inconsistent group. Due to the definition of inconsistent retell, children with oral reading fluency below 40 were not included. Second, although the relation between oral reading fluency and Broad Reading Cluster was less strong for the inconsistent retell group, it was not until VIP oral reading fluency was above 80 that there was a consistent over-prediction of Broad Reading Cluster scores, and then only for a small number of children. Third, regardless of VIP retell flu-

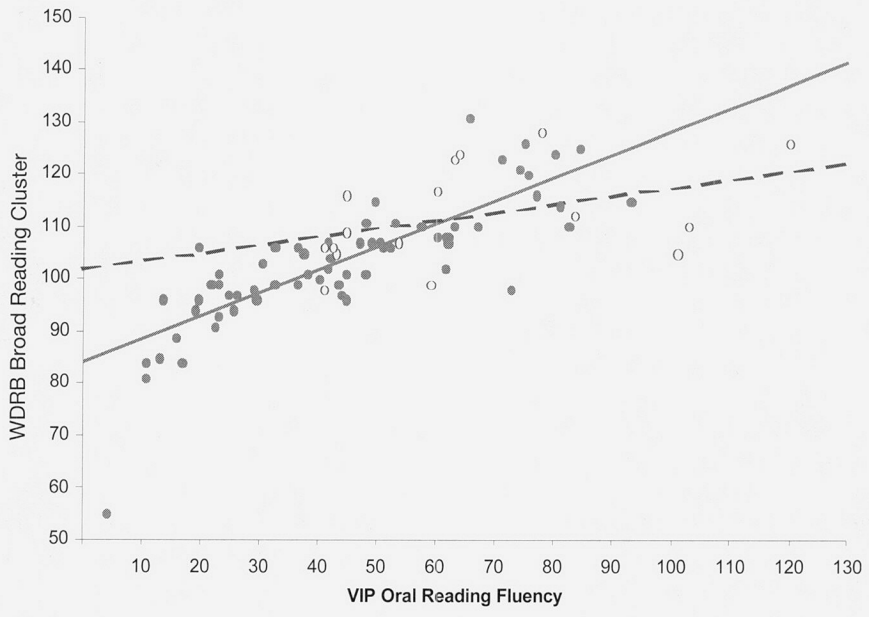

Figure 2. Relation between VIP oral reading fluency and WDRB Broad Reading Cluster for students with consistent retell (solid circles and dashed line) and for students with inconsistent retell (open circles and solid line).

ency, all children who were reading over 40 words correctly per minute at the end of first grade scored at or above a standard score of 96, 39th percentile, on the Broad Reading Cluster. Thus, the fundamental conclusion that students who read more than 40 words correctly per minute are "reading healthy" is not compromised by any combination of VIP oral reading fluency and VIP retell fluency obtained by the children in this study. Fourth, some children obtained relatively high VIP oral reading fluency scores, but unexpectedly lower Broad Reading Cluster scores. However, they did not have inconsistent VIP retell fluency scores. For example, one student read 73 words correctly per minute (a relatively high VIP oral reading fluency score) and received a Broad Reading Cluster score of 98 (a more middle score), but the student's VIP retell fluency was roughly consistent with that student's VIP oral reading fluency.

DISCUSSION

The results of this study provide modest support for adding a measure of retell fluency to an existing battery of fluency-based measures. At the least, inclusion of retell fluency does not compromise the validity of oral reading fluency; a

more favorable interpretation is that retell fluency provides an efficient, fluency-based tool for teachers to use in identifying the handful of children whose oral reading fluency may not adequately represent comprehension. When used in tandem with oral fluency measures, retell measures provide a vehicle for more reliably targeting classroom and school-level resources and for maximizing the efficiency and effectiveness of early reading instruction.

There may be desirable by-products, as well. For teachers, use of a retell fluency measure may be a pathway to increased understanding and more confident use of the traditional oral reading fluency measure. Also, if retell fluency in fact represents a gateway to more sophisticated levels of comprehension, and if it is "teachable," as suggested in the introduction to this article, there may be instructional value in knowing and monitoring a student's abilities in this respect.

For investigators, retell fluency may provide a vehicle for better understanding the relationship of fluency to reading comprehension. One-minute measures of oral reading fluency are highly correlated to more lengthy measures of comprehension, although the reasons for this relationship are not clear. This lack of clarity suggests that while oral reading fluency may be the preeminent fluency-based indicator of reading competence, its value for *diagnosing* the sources of a given student's reading difficulties may be limited. More skill-specific measures, like retell fluency, may represent a vehicle for specifying factors that mediate the relationship of fluency and comprehension.

Curriculum-based measurement offers advantages to practicing school psychologists, as well. Concerns about the utility of commercially available norm-referenced tests are well documented in the school psychology literature (see Deno, 1985; Howell, 1986; Peverly & Kitzen, 1998; Rosenfeld, 1987; Rosenfeld & Shinn, 1989; Shapiro, 1990; Shapiro & Elliott, 1999). Of particular interest is their limited value for making instructional recommendations. Because the underlying constructs for these measures are often vague, test items may fail to adequately address instructionally important skills, which diminishes the usefulness of results for identifying specific instructional needs and for developing or modifying instructional plans. More practically, these measures cannot be used repeatedly to assess and monitor progress, and they are expensive and time consuming. Curriculum-based measures, like story retell fluency, are less subject to these limitations; cost-effective, time-efficient, and instructionally focused, they represent a useful tool for a profession that spends more time on assessment than on any other single task (Hos & Reschly, 2002).

Readers should interpret the results presented in this article with some caution. Probability sampling was not possible, which represents a threat to the external validity of the study and may limit the generalizability of its findings. Further, the relatively small-sized sample is homogenous on important variables. Because participants were recruited in low-income, poor-performing public schools from a single community, they may not adequately represent the population of all first graders or all first graders from the community in question. Fu-

ture studies of retell fluency will address issues related to generalizability. This kind of sampling limitation can complicate interpretation of inferential analyses, a possibility to consider when using such methods and when interpreting their results. Future study of story retell fluency should involve diverse samples, assembled, when possible, using probability sampling strategies.

References

Buck, J., & Torgeson, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test* (FCCR Tech. Rep. No. 1). Florida Center for Reading Research, Tallahassee, FL.

Crawford, L., Tindal, G., & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment, 7*(4), 303–323.

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

Deno, S., Mirkin, P., & Chiang, B. (1982). Identifying valid measures of reading. Exceptional Children, 49, 35–45.

Deno, S. (1992). The nature and development of curriculum-based measurement. *Preventing School Failure, 36*(2), 5–10.

Deno, S., Fuchs, L., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30*(4), 507–524.

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

Fewster, S., & MacMillan, P. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing. *Remedial and Special Education, 23,* 149–156.

Fuchs, L., & Deno, S. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57,* 488–501.

Fuchs, L., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28,* 659-671.

Fuchs, L., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9(2),* 20–28.

Fuchs, L., Fuchs, D., Hosp, M., & Jenkins, J. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5,* 239–256.

Fuchs, L., Fuchs, D., & Hamlett, C. (1989). Effects of instrumental use of curriculum-based measurement to enhance instructional programs. *Remedial and Special Education, 10,* 43–52.

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

Good, R., Simmons, D., & Kame'enui, E. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*(3), 257–288.

Hamilton, C. (2001). The "word caller" phenomenon: Teachers' judgments on reading comprehension and fluency (Doctoral dissertation, University of Oregon, 2001). *Dissertation Abstracts International, 62*(06), 2027.

Hos, J. L., & Reschly, D. J. (2002). Regional differences in school psychology practice. *School Psychology Review, 31,* 11–29.

Howell, K. W. (1986). Direct assessment of academic performance. *School Psychology Review, 15,* 324–335.

Jenkins, J., Fuchs, L., Espin, C., van den Brock, P., & Deno, S. (2000). *Effects of task format and performance dimension on word reading measures: Criterion validity, sensitivity to impairment, and context facilitation.* Paper presented at the Pacific Coast Research Conference, San Diego, CA.

Kaminski, R., & Good, R. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25,* 215–227.

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

Logan, G. (1997). Automaticity and reading: Perspectives from the instance theory of automatization. *Reading and Writing Quarterly, 13,* 123–146.

Markell, M., & Deno, S. (1997). Effects of increasing oral reading: Generalization across reading tasks. *The Journal of Special Education, 31*(2), 233–250.

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., Fuchs, L., & Deno, S. (1985). Measuring pupil progress: A comparison of standardized achievement tests and curriculum-related measures. *Diagnostique, 11*(2), 77–90.

Marston D., Mirkin, P., & Deno, D. (1984). Curriculum-based measurement: An alternative to traditional screening, referral, and identification. *Journal of Special Education, 18,* 109–117.

Nathan, R., & Stanovich, K. (1991). The causes and consequences of differences in reading fluency. *Theory Into Practice, 30,* 176–184.

National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Perfetti, C. (1995). Cognitive research can inform reading education. *Journal of Research in Reading, 18*(2), 106–115.

Peverly, S. T., & Kitzen, K. R. (1998). Curriculum based assessment of reading skills: Consideration and caveats for school psychologists. *Psychology in the Schools, 35*(1), 29–47.

Rosenfeld, S. A. (1987). *Instructional consultation.* Hillsdale, NJ: Erlbaum.

Rosenfeld, S. A., & Shinn, M. R. (Eds.). (1989). Mini-series on curriculum based assessment. *School Psychology Review, 18,* 299–370.

Salvia, J., & Ysseldyke, J. (1998). Using estimated true scores: A response to Sabers, Feldt, and Reschly. *Journal of Special Education, 22*(3), 367–373.

Shapiro, E. S. (1990). An integrated model for curriculum based assessment. *School Psychology Review, 19,* 331–349.

Shapiro, E. S., & Elliott, S. N. (1999). Curriculum-based assessment and other performance-based assessment strategies. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology.* New York: Wiley.

Shinn, M. (1998). *Advanced applications of curriculum-based measurement.* New York: The Guilford Press.

Shinn, M., Good, R., Knutson, N., Tilly, W., & Collins, V. (1992). Curriculum-based measurement reading fluency: A confirmatory factor analysis of its relation to reading. *School Psychology Review, 21,* 459–479.

Stanovich, K. (2000). *Progress in understanding reading: Scientific foundations and new frontiers.* New York: The Guilford Press.

Stoner, G., Scarpati, S., Phaneuf, R., & Hintze, J. (2002). Using curriculum-based measurement to evaluate intervention efficacy. *Child & Family Behavior Therapy, 24*(1/2), 101–112.

Woodcock, R. W., & Johnson, M. B. (1997). *Diagnostic Reading Battery (WDRB).* Allen, TX: DLM Teaching Resources.

Action Editor: Melissa Bray

Greg Roberts, Ph.D., is a Research Psychologist at The University of Texas at Austin. His areas of interest include measurement, program evaluation, and early literacy. He is affiliated with the Vaughn Gross Center for Reading and Language Arts, the Central Center for Reading First Technical Assistance, and The University of Texas at Austin. The School Psychology Program at the University of Texas at Austin is accredited by the APA.

Roland H. Good, Ph.D., is an Associate Professor at the University of Oregon. He teaches graduate courses in measurement, statistics, and research design and conducts research on Dynamic Indicators of Basic Early Literacy Skills (DIBELS™). Good was co-principle investigator for the Early Childhood Research Institute on Measuring Growth and Development, and he was a member of the Secretary's Reading Leadership Academy, Assessment Committee. The School Psychology Program at the University of Oregon is accredited by the APA.

Stephanie Corcoran is a school psychologist with Shelby County Schools in Alabama. She is currently completing her doctorate at the University of Alabama. Her research interests include the implementation and evaluation of evidence-based intervention and prevention programs.