

# Florida State University Libraries

---

Faculty Publications

The Florida Center for Reading Research

---

2013

## Advanced (Measurement) Applications of Curriculum-based Measurement in Reading

Yaacov M. Petscher, Kelli Cummings, Gina Biancarosa, and Hank Fien



**Advanced (measurement) applications of curriculum-based measurement in reading**

Yaacov Petscher

Florida Center for Reading Research

Florida State University

Kelli D. Cummings

Gina Biancarosa

Hank Fien

University of Oregon

### Abstract

The purpose of this article is to provide a commentary on the current state of several measurement issues pertaining to curriculum-based measures of reading (R-CBM<sup>1</sup>). We begin by providing an overview of the utility of R-CBM, followed by a presentation of five specific measurements considerations: 1) the reliability of R-CBM oral reading fluency, 2) issues pertaining to form effects, 3) the generalizability of scores from R-CBM, 4) measurement error, and 5) linearity of growth in R-CBM. We then conclude with a presentation of the purpose for this issue and broadly introduce the articles in the special issue. Because oral reading fluency is one of the most common measures of R-CBM, much of the review is focused on this particular type of assessment; however, the issues presented extend to other assessments of R-CBM.

---

<sup>1</sup> In a discussion among the authors about whether to represent curriculum-based measures in reading as CBM-R (the more historic representation) or R-CBM (a more modern characterization), we opted to use the RAND function in Excel to decide which acronym was to be used. As shown here, the lot fell to R-CBM; however the acronyms can be used interchangeably.

Oral reading fluency (ORF) is one of many curriculum-based measures of reading (R-CBM) that serves as an indicator of overall reading achievement (Fuchs, Fuchs, Hosp, & Jenkins, 2001). Research has demonstrated the importance of fluently reading connected text as a theoretical construct (Hudson, Pullen, Lane, & Torgesen, 2009; LaBerge & Samuels, 1974) and scores from fluency measures have been validated as proxies for measuring reading comprehension (Buck & Torgesen, 2003; Ridel, 2007; Roehrig et al., 2008; Shapiro, Solari, & Petscher, 2008). As a result, measures of ORF have widespread utility in their capability as screeners of future reading comprehension difficulties, and as tools to monitor student progress in overall reading skills. Along with a decade of attention to accountability, the recent spread of Response to Intervention (RtI) models as a means of identifying and monitoring students at risk for reading failure has made R-CBM and ORF in particular more popular than ever (Cummings, Atkins, Allison, & Cole, 2008).

Whereas ORF measures used to be idiographic, and truly based in local curricula (e.g., Hintze, Shapiro, Conte, & Basile, 1997; Shinn, Gleason, & Tindal, 1989), a number of standardized measurement systems with common sets of reading passages have since been developed; for example, AIMSweb, DIBELS, easyCBM, and Edcheckup are all widely used R-CBM measurement systems that include a measure of ORF. Although the procedures, prompts, and passages can vary among measurement systems, certain practices have become standard across systems. Typical R-CBM scoring conventions include (a) a count of the total *words read correctly*, (b) mispronunciations, substitutions, omissions, and transpositions are counted as *errors*, (c) inserted or repeated words are ignored, (d) students are allowed to "self-correct" a previously mispronounced word if they do so within three seconds, and (e) if students hesitate or struggle with a word for three seconds, an examiner provides the word but marks it as an error on

the test form (Hosp, Hosp, & Howell, 2007; Stecker et al., 2005). In addition, most systems offer schools benchmark passages for occasional status checks (i.e., screening purposes) and for more frequent checks of response to intervention (i.e., progress monitoring). R-CBM data may be aggregated at the classroom, grade, or school level to provide quarterly indicators of school health that can be used to allocate resources (Cummings, et al., 2008).

Regardless of the system used, R-CBM research across the last 25 years has focused on investigating (a) the technical features of the static score (i.e., baseline or aggregate score(s)), (b) the technical features of the slope (i.e., time series performance), and (c) the instructional utility of the measure (Deno, 2003; Fuchs, 2004). However, the increasingly widespread and novel use of ORF for higher stakes accountability decisions has brought closer scrutiny to an expanded set of technical adequacy and validity information for ORF. Many of these concerns hinge on psychometric issues, such as reliability and error of measurement for estimates of students' status and growth.

### **Reliability of R-CBM ORF**

Generally speaking, reliability estimates for R-CBM ORF measures are uniformly high. For example, when estimated using three or more passages administered concurrently, alternate form reliability for a single passage is usually estimated at .80 and above and for the set of passages is approximately .90 and above (Baker et al., 2008; Biancarosa, Bryk, & Dexter, 2010; Howe & Shinn, 2002; Roberts, Good, & Corcoran, 2005). As a result, and perhaps due to historical reasons (e.g., Marston, 1989) most, though not all, R-CBM systems take the median or mean score from three passages administered concurrently. In a 2011 review of National Center for Response to Intervention (NCRTI) website, the authors identified 11 unique R-CBM screening tools. Of those, six required that two or more reading passages to be administered for

screening assessment purposes. For progress-monitoring purposes, five R-CBM tools required the administration of multiple passages per assessment period. Although the administration of multiple passages per assessment period improves the static reliability of the average score, it is not yet known if this practice sufficiently addresses some of the other concerns about ORF validity.

### **Form Effects and R-CBM ORF**

As is evident from the lower alternate form reliability for single passage test administrations compared to multiple passage administrations, reliability suffers due to passage-specific differences that cannot be eliminated entirely through careful text authoring (e.g., using readability formulae) nor the typical ORF measure development process (Ardoin, Suldo, Witt, Aldrich, & McDonald, 2005; Ardoin, Williams, Christ, Klubnik, & Wellborn, 2010; Betts, Pickart, & Heistad, 2009; Christ & Ardoin, 2009). Indeed, recent research has demonstrated that from a progress-monitoring perspective, these passage-specific, or form, effects can have a profound influence on the trajectory of monitored students (Ardoin & Christ, 2009; Francis, Santi, Barr, Fletcher, Varisco et al., 2008; Hintze & Christ, 2004) as well as diagnostic accuracy (Petscher & Kim, 2011). As a result, there are now calls for equating of ORF probe passages. In fact, researchers have begun to use equated ORF scores in their studies (e.g., Barth, Stuebing, Fletcher, Cirino, Francis et al., 2012; Denton, Barth, Fletcher, Wexler, Vaughn et al., 2011), but the trade-offs among various methods of equating and the costs associated with them have only begun to be investigated (Albano & Rodriguez, 2012; Christ & Ardoin, 2009; Francis et al., 2008; Griffiths, VanDerHeyden, Skokut, & Lilles, 2009).

### **Generalizability and Dependability of R-CBM ORF**

Studies examining ORF progress-monitoring measures from a Generalizability (G) theory perspective have demonstrated that although relative generalizability is very good, absolute generalizability is much lower. The advantage of G theory studies is that they account for multiple sources of unreliability (or error) simultaneously and also allow for projection of reliability under different circumstances. In the first such G theory study (Hintze, Owen, Shapiro, & Daly, 2000), researchers accounted for variance due to the measurement materials (i.e., passages based on a long-term versus a short-term goal level of difficulty), the grade level of students (second- through fifth-grade), and assessment occasions (twice per week over either eight or ten weeks) and found that the reliability of ORF for assessing growth for an individual student was .90 when using long-term goal materials (i.e., end of grade level) and .80 when using both instructional level and long-term goal materials. Projecting the reliability when only a single passage source is used for two grade levels, dependability dropped to .82 when using long-term goal materials and .67 when using both instructional level and long-term goal materials (Hintze et al., 2000). To date, the only other G theory study obtained reliability of slope estimates using 20 grade-level passages (range of coefficients = .81 - .97; Poncy, Skinner, & Axtell, 2005). In contrast to Hintze et al., Poncy and colleagues (2005) examined the influence of passages alone on reliability. Together these studies suggest that alternate form reliability may overestimate true reliability because traditional psychometric calculations for reliability (i.e., correlation coefficients) do not take into account the possibility of absolute score differences across forms (e.g., Albano & Rodriguez, 2012).

### **Error of Measurement for ORF**

Related to the issue of reliability, is the consideration of the standard error of measurement (SEM) associated with ORF assessment. SEM provides a transparent metric for

evaluating the degree of measurement error associated with a student's particular raw score. Poncy and colleagues (2005) offered the first peer-reviewed publication to address the SEM associated with R-CBM, and these results have since been replicated by other researchers (e.g., Christ & Ardoin, 2006; Christ & Silberglitt, 2007). The SEM associated with a single ORF score has been found to range from five to 15 wcpm, with a median SEM of 10 wcpm (Christ & Silberglitt, 2007). Even when passage effects are controlled through a Generalizability (G) theory approach, the SEM can range from three or four to as many as 12 wcpm (Poncy et al., 2005). Furthermore, a line of research by Christ and colleagues has demonstrated a profound impact of SEM on estimates of slope and, ultimately, decisions about adequate student progress (Christ, 2006; Ardoin & Christ, 2009; Christ, Zopluoglu, Long, & Manghen, 2012). The work of Christ and colleagues reveals that when the SEM is *small* (i.e.,  $< 6$ ), slope estimates show little error as long as ORF is measured six or more times. However, when the SEM is larger, variation in slope estimates can also be quite large; with unsatisfactory psychometric properties until at least 10 or more ORF measures have accumulated. It is worth noting that both (a) small SEMs ( $< 6$ ) are rarely observed in practice and (b) that waiting for a satisfactory number of data points to address unreliability in a student data set has the potential to either delay or misalign intervention support with student need. Ensuring that R-CBM is validated for the intended use of important educational decisions requires a renewed focus on methods used to evaluate test administration.

### **Linearity of R-CBM ORF Growth**

Adding to the complexity of measurement modeling, is the problem that change in ORF over time may not follow a linear trajectory nor be the same for readers of all abilities, which most technical manuals and research, including the aforementioned work, assume (see also, Speece & Ritchey, 2005). Research over the past decade suggests that ORF growth may in fact

be curvilinear with faster growth observed earlier in the academic year and slower growth later (e.g., Al Otaiba, et al., 2009; Christ, Silberglitt, Yeo, & Cormier, 2010; Crowe, Connor, & Petscher, 2009; Deno, Fuchs, Marston, & Shin, 2001; Logan & Petscher, 2010). In addition, ORF slopes vary depending on the skill level of readers at the beginning of the school year, receipt of special education services, the varying type and quality of instructional support that is available to students across the school year, and a number of other individual characteristics (Puranik, Petscher, Al Otaiba, Catts, & Lonigan, 2008; Silberglitt & Hintze, 2007; Wang, Algozzine, Ma, & Porfeli, 2011). The results of this line of research imply that applying the same standard for expected growth, and a linear standard at that, may be a flawed practice. Considering that research has demonstrated that growth in ORF uniquely explains individual differences in reading comprehension (Kim, Petscher, Schatschneider, & Foorman, 2010), it is critically important that the functional form of growth is correctly specified.

### **Purpose of the Special Issue**

Deno, Mirkin, and Chiang (1982) identified R-CBM as *the critical CBM task* and a robust indicator of overall reading competence, and four decades of research have demonstrated R-CBM's value for providing reliable and valid screening and monitoring information for first grade and beyond (e.g., Wayman, Wallace, Wiley, Tichá, & Espin; 2007). However, research on R-CBM for new purposes is not yet complete. The purpose of this special issue is to bring together a set of articles that highlight at least some of the novel measurement issues facing R-CBM researchers in the 21<sup>st</sup> century. Issues of form equivalency, methodologies for estimating linear and non-linear change over time, and using R-CBM tools as indicators of teacher effects are the three key areas on which we will focus.

In the first article, Stoolmiller, Biancarosa, and Fien explore passage equivalence from a robust and novel measurement perspective. The authors apply a set of linear and non-linear factor analytic models to scores from oral reading fluency passages, and discuss the appropriate methods for equating given the type of measurement model applied. Cummings, Park, and Bauer follow this paper by examining the persistence of form effects in the DIBELS Next Oral Reading Fluency measures in first through sixth grades; evaluating the value of different equating strategies to control for the differences in average fluency rates across the passages. The third article in this series by Christ, Monaghan, Zopluoglu, and Van Norman evaluate growth and diagnostic accuracy of R-CBM scores using pre-post assessment data ranging from six week to 20 week intervals. The fourth article, by Kamata, Nese, Patarapichayathan, and Lai, use scores from passage reading fluency to demonstrate how piecewise growth models and piecewise growth mixture models may be applied to data where only three time points exist per individual. The fifth article, by Hart, Taylor, and Schatschneider, presents a unique methodology using ORF scores to demonstrate how classrooms may causally affect later performance in fluency tasks. Though this design has traditionally been used in medical research, this paper represents the first application in education to account for school effects above the student level.

### **Conclusion**

The articles represented here begin to delve more deeply into recent calls to comprehensively explore issues of the reliability and validity of fluency scores. Because R-CBM, especially oral reading fluency, will continue to be used by practitioners and researchers, it behooves us to continually study methodologies and practices for improving the psychometric quality of the scores from such assessments. Advances in educational measurement have been occurring at a rapid pace, yet even some relatively older innovations developed in the last ten to

fifteen years [e.g., item response models with time parameters (Verhelst, Verstralen, & Jansen, 1997), latent change scores models (McArdle, 2001)], which hold some great promise for understanding individual differences, have seen little application in the field of assessment research. As accessible software packages and programs are developed for educational researchers, the feasibility of implementing such models is greatly enhanced. The potential promise of these advances in measurement is the ability to estimate more reliable student scores and remove multiple sources of bias. Subsequently, this augments the capacity to use R-CBM scores to correctly identify students' strengths and weaknesses, the need for additional services or accommodations, as well as monitoring progress.

## References

- Al Otaiba, S. , Petscher, Y., Pappamihiel, N. E., Williams, R. S., Drylund, A. K., & Connor, C. M. (2009). Modeling oral reading fluency development in Latino students: A longitudinal study across second and third grade. *Journal of Educational Psychology, 101*, 315-329.
- Albano, A.D., & Rodriguez, M.C. (2012). Statistical equating with measures of oral reading fluency. *Journal of School Psychology, 50*, 43-59.
- Ardoin, S. P., & Christ, T. J. (2009). Curriculum based measurement of oral reading: Estimates of standard error when monitoring progress using alternate passage sets. *School Psychology Review, 38*, 266-283.
- Ardoin, S. P., Suldo, S. M., Witt, J., Aldrich, S., & McDonald, E. (2005). Accuracy of readability estimates' predictions of CBM performance. *School Psychology Quarterly, 20*, 1–22.
- Ardoin, S. P., Williams, J. C., Christ, T. J., Klubnik, C., & Wellborn, C. (2010). Examining readability estimates' predictions of students' oral reading rate: Spache, lexile, and forecast. *School Psychology Review, 39*, 277–285.
- Baker, S., Smolkowski, K., Katz, R., Fien, H., Seeley, J., Kame'enui, E.J., et al. (2008). Reading fluency as a predictor of reading proficiency in low-performing high poverty schools. *School Psychology Review, 37*, 18-37.
- Barth, A., Stuebing, K. K., Fletcher, J. M., Cirino, P. T., Francis, D. J., & Vaughn, S. (2012). Reliability and validity of the median score when assessing the oral reading fluency of middle grade readers. *Reading Psychology, 33*, 131-161.

- Betts, J., Pickart, M., & Heistad, D. (2009). An investigation of the psychometric evidence of CBM-R passage equivalence: Utility of readability statistic and equating alternate forms. *Journal of School Psychology, 47*, 1-17.
- Biancarosa, G., Bryk, A. S., & Dexter, E. (2010). Assessing the value-added effects of Literacy Collaborative professional development on student learning. *Elementary School Journal, 111*, 7-34.
- Buck, J., & Torgesen, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test* [Electronic version]. Retrieved: [www.fcrr.org/Technical Reports/TechnicalReport1.pdf](http://www.fcrr.org/Technical%20Reports/TechnicalReport1.pdf).
- Christ, T.J. (2006). Short term estimates of growth using curriculum-based measurement of oral reading fluency: Estimates of standard error of the slope to construct confidence intervals. *School Psychology Review, 35*, 128-133.
- Christ, T. J., & Ardoin, S. P. (2009). Curriculum-based measurement of oral reading: Passage equivalence and probe-set development. *Journal of School Psychology, 47*, 55–75.
- Christ, T. J., & Silbergitt, B. (2007). Curriculum-based measurement of oral reading fluency: The standard error of measurement. *School Psychology Review, 36*, 130–146.
- Christ, T. J., Silbergitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based measurement of oral reading: An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review, 39*, 447-462.
- Christ, T. J., Zopluoglu, C., Long, J. D., & Monaghan, B. D. (2012). Curriculum-based measurement of oral reading: Quality of progress monitoring outcomes. *Exceptional Children, 78*, 356-373.

- Crowe, E.C., Connor, C.M., & Petscher, Y. (2009). Examining the core relations between poverty, reading curriculums, and first through third grade reading achievement. *Journal of School Psychology, 47*, 187-214.
- Cummings, K. D., Atkins, T., Allison, R., & Cole, C. (2008). Response to Intervention. *Teaching Exceptional Children, 40*, 24–31.
- Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37*, 184–192.
- Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J., Vaughn, S., Cirino, P. T., Romain, M., & Francis, D. J. (2011). The relations among oral and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific Studies of Reading, 15*, 109-135.
- Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30*, 507-526.
- Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49*, 36–45.
- Francis, D. J., Santi, K. S., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology, 46*, 315-342.
- Fuchs, L.S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33*, 188-192.

- Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*, 239-256.
- Griffiths, A-J., VanDerHeyden, A. M., Skokut, M., & Lilles, E. (2009). Progress monitoring in oral reading fluency within the context of RTI. *School Psychology Quarterly, 24*, 13-23.
- Hintze, J.M. & Christ, T.J. (2004). An examination of variability as a function of passage variance in CBM progress monitoring. *School Psychology Review, 33*, 204-217.
- 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., 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.
- Howe, K. B., & Shinn, M. M. (2002). *Standard reading assessment passages (RAPs) for use in general outcome measurement: A manual for describing development and technical features*. Eden Prairie, MN: Edformation.
- Hosp, M., Hosp, J., & Howell, K. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York: The Guilford Press.
- Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2009). The complex nature of reading fluency: A multidimensional view. *Reading and Writing Quarterly, 25*, 4-32.

- Kim, Y.S., Petscher, Y., Schatschneider, C., Foorman, B.R. (2010). Does growth in oral reading fluency matter in reading comprehension achievement? *Journal of Educational Psychology, 102*, 652-667.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 62*, 293–323.
- Logan, J.A.R & Petscher, Y. (2010). School profiles of at-risk student concentration: Differential growth in oral reading fluency. *Journal of School Psychology, 48*, 163-186.
- Marston, D.B. (1989). A curriculum-based measurement approach to assessing academic performance: 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.
- McArdle, J.J. (2001). A latent difference score approach to longitudinal dynamic structural analyses. In R. Cudeck, S. du Toit, D. Sorbom (Eds.), *Structural Equation Modeling: Present and Future*, , pp. 342–380. Lincolnwood, IL: Scientific Software International.
- Petscher, Y., & Kim, Y.S. (2011). The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *Journal of School Psychology, 49*, 107-129.
- Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2005). An investigation of the reliability and standard error of measurement of words read correctly per minute using curriculum-based measurement. *Journal of Psychoeducational Assessment, 23*, 326–338.
- Puranik, C., Petscher, Y., Al Otaiba, S., Catts, H.W., & Lonigan, C.J. (2008). Development of oral reading fluency in children with speech or language impairments: A growth curve analysis. *Journal of Learning Disabilities, 41*, 545-560.
- Ridel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly, 42*, 546-567.

- Roberts, G., Good, R., & Corcoran, S. (2005). Story retell: A fluency-based indicator of reading comprehension. *School Psychology Quarterly, 20*, 304-317.
- Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Not just speed reading: Accuracy of the DIBELS oral reading fluency measure for predicting high-stakes third grade reading comprehension outcomes. *Journal of School Psychology, 46*, 343-366.
- Shapiro, E., Solari, E., & Petscher, Y. (2008). Use of an assessment of reading comprehension in addition to oral reading fluency on the state high stakes assessment for students in grades 3 through 5. *Journal on Learning and Individual Differences, 18*, 316-328.
- Shinn, M.R., Gleason, M.M., & Tindal, G. (1989). Varying the difficulty of testing materials: Implications for curriculum-based measurement. *Journal of Special Education, 23*, 223-233.
- Silberglitt, B., & Hintze, J. (2007). How much growth can we expect? A conditional analysis of R-CBM growth rates by level of performance. *Exceptional Children, 74*, 71-84.
- Speece, D. L., & Ritchey, K. D. (2005). A longitudinal study of the development of oral reading fluency in young children at risk for reading failure. *Journal of Learning Disabilities, 38*, 387-399.
- Stecker, P.M., Fuchs, L.S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*, 795-819.
- Verhelst, N.D., Verstralen, H.H.F.M., & Jansen, M.G.H. (1997). A logistic model for time-limit tests. In W.J. van der Linden and R.K. Hambleton (Eds.) *Handbook of modern item response theory* (pp. 169-185). New York: Springer.

Wang, C., Algozzine, B., Ma, W., & Porfeli, E. (2011). Oral reading rates of second-grade students. *Journal of Educational Psychology, 103*, 442-454.

Wayman, M., Wallace, T., Wiley, H.I., Ticha, R., & Espin, C.A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education, 41*, 85-120.