New metrics, measures, and uses for fluency data: an introduction to a special issue on the assessment of reading fluency

Gina Biancarosa · Kelli D. Cummings

Abstract The primary objective of this special issue is to synthesize results from recent reading fluency research endeavors, and to link these findings to practical uses of reading curriculum-based measurement (R-CBM) tools. Taken together, the manuscripts presented in this issue discuss measurement work related to new metrics of indexing student reading progress, the application of R-CBM tasks to traditionally underrepresented groups in the literature, specifically students in Grades 7–8 disaggregated by first-language and special education status, and new uses of aggregate reading fluency data to assist with school-level reporting. All manuscripts link to practical applications for R-CBM users and researchers. In this introduction to the issue, we provide an overview of the critical pieces of each article, discuss how and where the article fits in the broader context of reading research generally and R-CBM research specifically, and set the stage for how these articles can serve as a springboard for subsequent phases of R-CBM research.

Introduction

Reading researchers are most familiar with the construct of reading fluency through theories of automaticity (LaBerge & Samuels, 1974) and efficiency (Perfetti, 1985) in reading. These theories suggest that fast, accurate reading is a prerequisite for

G. Biancarosa (✉)
Department of Educational Methodology, Policy, and Leadership, College of Education, University of Oregon, 102 Lokey Education, Eugene, OR 97403-5267, USA
e-mail: ginab@uoregon.edu

K. D. Cummings
Center on Teaching and Learning, University of Oregon, 1600 Millrace Drive, Suite 207, Eugene, OR 97403, USA

reading comprehension and that when reading is not effortless enough, the ability to make and sustain meaning from what one reads is impaired, thereby explaining fluency's strong correlations with comprehension. Indeed, Smith and Holmes (1971) argued that "unless the reader reads fast enough … he is not going to comprehend what he is reading simply because his memory system will not be able to retain, organize, and store the fragmentary information in any efficient way" (p. 412).

Most educators, however, become familiar with the construct of fluency by exposure to the use of curriculum-based measures (CBMs), particularly those in reading. Curriculum-based measurement is a standardized, fluency-based assessment practice that quantifies student academic performance during a specified, brief period of time, often less than 10 min. CBM has strong roots in special education, with its primary original purpose as a tool to evaluate the achievement gains of students in basic skill areas, such as reading, writing, and arithmetic (Deno, 2003). A major use of CBM today includes student-level universal screening. In this practice, students are assessed three to four times per year in order to identify those who may be at risk for not meeting state standards on high-stakes, end-of-year assessments (Mellard, McKnight, & Woods, 2009).

For most CBMs, fluency metrics are critical elements of the technology. Constraining performance to a standard unit of time allows scores on the measures to be comparable across students and over time within a relatively brief assessment window. Current best practices recommendations for CBM in the field of school psychology include: (a) the use of CBM universal screening data for all students (Ikeda, Neesen, & Witt, 2008; Stewart & Silberglitt, 2008), (b) progress monitoring with alternate forms of CBMs to make response-to-intervention decisions about individual students (Hixson, Christ, & Bradley-Johnson, 2008; Fuchs & Fuchs, 2008), and (c) the use of aggregate CBM data in school-level formative evaluation (Braden & Tayrose, 2008; Kaminski, Cummings, Powell-Smith, & Good, 2008). Numerous recent articles demonstrate the robust psychometric properties of common reading CBMs (R-CBMs), and of oral reading fluency (ORF) screening CBMs in particular, in a number of different contexts and for different populations and purposes (e.g., Ardoin, Christ, Morena, Cormier, & Klingbeil, 2013; Christ, Zopluoglu, Monaghen, & Van Norman, 2013; Espin, Wallace, Lembke, Campbell, & Long, 2010; Francis, Santi, Barr, Fletcher, Varisco, & Foorman, 2008; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Petscher & Kim, 2011).

Nonetheless, a number of questions about the use of ORF data and other R-CBMs remain contested in the field. The purpose of this special issue of *Reading & Writing* is to provide insight into measurement considerations in the domain of fluency-based R-CBMs that speak too many of these questions.

New metrics

Whether and how prosody ought to figure into the assessment of oral reading has been debated since the inception of R-CBMs and even inspired a special National Assessment of Educational Progress study in 2002 (Daane, Campbell, Grigg, Goodman, & Oranje, 2005). Although most definitions of reading fluency incorporate

reading with expression or prosody (e.g., Hudson, Pullen, Lane, & Torgesen, 2009; Kuhn, Schwanenflugel, & Meisinger, 2010; Rasinski, Rikli, & Johnston, 2009), R-CBMs have traditionally excluded an explicit measure of this aspect of fluency. Rationales for this exclusion range from the modest added value of assessing prosody when predicting later reading comprehension to the difficulty of measuring prosody with reliability. For example, rating scales such as the Multidimensional Fluency Scale (Rasinski et al., 2009) have only achieved 86 % inter-rater agreement within 2-points for the sum of four 4-point rating scales. When prosody is measured in addition to more typical ORF CBM metrics, it may add and even mediate partially the predictive relations between ORF and reading comprehension (Schwanenflugel, Hamilton, Kuhn, Wisenbaker, & Stahl, 2004). More recently, however, researchers have begun to explore means of measuring prosody that do not require subjective judgments or rubrics. Although contemporary research has led to advances in the practical and reliable measurement of prosody (Benjamin et al., 2013), current CBM systems have yet to incorporate these more precise prosody measures. The article by Schwanenflugel, Westmoreland, and Benjamin in this issue represents some of the cutting-edge work being done in this area.

In our second article on the topic of new metrics in fluency-based R-CBMs, Petscher, Mitchell, and Foorman (this issue) evaluated student response time on a computer–adaptive test of passage and list-word reading fluency. They then compared student reading scores using classical-test theory (CTT) and item-response theory (IRT) versus a conditional item-response model. This work is important because student response time is likely to become increasingly relevant as R-CBMs are translated into computer–administration formats. Additionally, evaluations of this type are particularly relevant to the next iteration of psychometric study for R-CBMs. Many critiques of R-CBMs center on their absolute reliability and validity across testing forms, occasions, and raters, as well as whether and how alternate forms of common R-CBMs ought to be equated. To date, however, the vast majority of work has favored CTT rather than more modern approaches to psychometrics. For example, only a handful of studies have examined R-CBMs from the perspective of Generalizability theory (Hintze & Pelle Petitte, 2001; Hintze, Owen, Shapiro, & Daly, 2000; Poncy, Skinner, & Axtell, 2005). These studies have added confidence that individual R-CBMs are reliable across individuals, forms (i.e., different passages), groups (as defined by special education status or grade level), and occasions when multiple sources of measurement error are considered simulta-neously, but more work is needed in this area particularly in the domain of school-based assessments with multiple passage forms (e.g., Cummings, Park, & Schaper, 2012; Francis et al., 2008). With respect to IRT methods in particular, at the time of this publication only one study beyond the manuscript in the current issue has attempted to apply IRT to fluency-based R-CBMs (Zopluoglu, 2013).

New measures

Another area of debate centers on the validity of ORF as a universal screener. Many question the validity of R-CBM measures of ORF as an index of overall reading achievement for students in the middle grades and beyond (Rasinski et al., 2009;

Slocum, Street, & Gilberts, 1995). Many also question the validity of ORF for students who are English Learners (ELs), arguing that limited proficiency in English can be confounded with or mask reading-specific difficulties on such measures (e.g., Geva & Farnia, 2012; Lesaux & Siegel, 2003). In both instances, arguments are made for the importance of triangulating ORF data with other measures. In this issue, D. Baker and colleagues examine these questions and more in a sample of Grade 7 and 8 students using a traditional 1-year prediction study with an R-CBM measure of ORF and two additional measures: the number of errors made while reading aloud and an assessment of multiple-choice reading comprehension (RC). This article adds importantly to the small but growing number of R-CBM studies dealing with either middle grade readers or ELs (e.g., Barth et al., 2012; Knight-Teague, Vanderwood, & Knight, 2014).

New uses

The increasingly widespread use of R-CBMs has also led to new questions as to how researchers and schools can use R-CBM data for drawing inferences beyond individual students. At issue is whether R-CBM data can yield insights regarding the efficacy of schools and the extent to which variability in student performance can be attributed to the nesting of students in larger groups. That is, can aggregated data supply information on how well schools are meeting the needs of different populations of readers? And to what extent are differences in student R-CBM scores attributable to differences between classrooms, schools, and districts? Cummings, Stoolmiller, S.K. Baker, Fien, and Kame'enui (this issue) address the first question by presenting one method for examining school-level reading achievement using R-CBM. Kim, Petscher, and Foorman (this issue) tackle the second question by parsing the variance in student scores from universal screening CBMs, but with a different measure (a Maze silent reading task) and a large range of student grade levels (i.e., Grades 3–10). The goals of the Kim and colleagues study were twofold: (a) assess the degree to which variation in maze scores exist between students, class, schools, and districts and (b) validate the maze task in terms of the unique variance that it adds to predicting spring reading comprehension scores. Although neither of these two articles directly address the issue of using student data for teacher evaluation systems, as allowed by recently granted flexibility in meeting certain requirements of the No Child Left Behind Act (U.S. DOE, 2012, n.d.), they both suggest that a great deal of variability exists between schools in terms of student performance on R-CBMs. Whether the differences between schools (let alone teachers and classrooms) are reliable enough for high stakes use remains an open question, but as both articles make clear, these differences are substantial enough to warrant attention from a formative program evaluation perspective (e.g., Betebenner, 2009; Earl & Fullan, 2003).

Summary

This issue of *Reading & Writing* highlights future uses and innovations in fluency-based R-CBMs in education and addresses several outstanding issues in and

critiques of the measurement properties of fluency-based R-CBMs. The questions addressed by the articles in this special issue are by no means the only questions remaining about R-CBM, but each tackles an understudied area in which new knowledge contributes to our use and development of R-CBMs.

Investigations such as those presented in this issue hold implications for the field of reading research quite broadly. The articles by Schwanenflugel and colleagues (this issue) and by Petscher and colleagues (this issue) have important implications for how reading fluency can be measured with more precision, thereby yielding more information about the construct. Schwanenflugel and colleagues suggest prosody belongs among the salient, definitional aspects of reading fluency. Petscher and colleagues provide a modern approach to the measurement of the accuracy and speed dimensions of reading fluency. Likewise, the article by D. Baker and colleagues (this issue) suggests that measuring reading fluency is not for the elementary grades alone, but can provide important information regarding the reading performance of middle grade students, if not equally effectively for all subgroups. Finally, the articles by Cummings and colleagues (this issue) and by Kim and colleagues (this issue) demonstrate reading fluency has utility beyond understanding individual differences among readers, but can also be used to understand aggregate differences between classrooms and schools.

Together the studies presented in this special issue provide readers with a window on the future of fluency-based R-CBMs in education, along with an understanding of the psychometric properties that make the interpretation and use of fluency metrics unique. R-CBMs have been essential tools not only for exploring fundamental questions regarding proficient reading and its development, but also more practically for supporting implementation of a prevention-oriented, response-to-intervention approach to reading instruction. As such, the expanded perspectives on R-CBMs represented in this issue also hold implications for the everyday decisions made with fluency data within models of prevention and early intervention. Indeed, each article details such implications in its discussion section. In addition, Christ and Ardoin (this issue) close out the special issue by commenting on each of the articles and offering additional insights into their implications, as well as identifying future topics yet to be tackled in fluency-based R-CBM research.

References

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding Curriculum-Based Measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51*, 1–18. doi:10.1016/j.jsp.2012.09.004.

Barth, A. E., Stuebing, K. K., Fletcher, J. M., Cirino, P. T., Romain, M., Francis, D., et al. (2012). Reliability and validity of oral reading fluency median and mean scores among middle grade readers when using equated texts. *Reading Psychology, 33*(1–2), 133–161. doi:10.1080/02702711.2012.631863.

Benjamin, R. G., Schwanenflugel, P. J., Meisinger, E. B., Groff, C., Kuhn, M. R., & Steiner, L. (2013). A spectrographically grounded scale for evaluating reading expressiveness. *Reading Research Quarterly, 48*, 105–133. doi:10.1002/rrq.43.

Betebenner, D. (2009). Norm-and criterion-referenced student growth. *Educational Measurement: Issues and Practice, 28*(4), 42–51. doi:10.1111/j.1745-3992.2009.00161.x.

Braden, J. P., & Tayrose, M. P. (2008). Best practices in educational accountability: High-stakes testing and educational reform. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 5, pp. 575–588). Bethesda, MD: National Association of School Psychologists.

Christ, T. J., Zopluoglu, C., Monaghen, B. D., & Van Norman, E. R. (2013). Curriculum-based measurement of oral reading: Multi-study evaluation of schedule, duration, and dataset quality on progress monitoring outcomes. *Journal of School Psychology, 51*, 19–57. doi:10.1016/j.jsp.2012.11.001.

Cummings, K. D., Park, Y., & Schaper, H. A. B. (2012). Form effects on DIBELS Next oral reading fluency progress-monitoring passages. *Assessment for Effective Intervention, 38*(2), 91–104. doi:10.1177/1534508412447010.

Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). Fourth-grade students reading aloud: NAEP 2002 special study of oral reading (NCES 2006-469). U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Washington, DC: Government Printing Office.

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*(3), 184–192. doi:10.1177/00224669030370030801.

Earl, L., & Fullan, M. (2003). Using data in leadership for learning. *Cambridge Journal of Education, 33*(3), 383–394.

Espin, C., Wallace, T., Lembke, E., Campbell, H., & Long, J. D. (2010). Creating a progress-monitoring system in reading for middle-school students: Tracking progress toward meeting high-stakes standards. *Learning Disabilities Research & Practice, 25*, 60–75. doi:10.1111/j.1540-5826.2010.00304.x.

Francis, D. J., Santi, K. L., 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. doi:10.1016/j.jsp.2007.06.003.

Fuchs, L. S., & Fuchs, D. (2008). Best practices in progress monitoring reading and mathematics at the elementary grades. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 5, pp. 2147–2164). Bethesda, MD: National Association of School Psychologists.

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.

Geva, E., & Farnia, F. (2012). Developmental changes in the nature of language proficiency and reading fluency paint a more complex view of reading comprehension in ELL and EL1. *Reading and Writing, 25*, 1819–1845. doi:10.1007/s11145-011-9333-8.

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*(1), 52–68.

Hintze, J. M., & Pelle Petitte, H. A. (2001). The generalizability of CBM oral reading fluency measures across general and special education. *Journal of Psychoeducational Assessment, 19*, 158–170.

Hixson, M., Christ, T. J., & Bradley-Johnson, S. (2008). Best practices in the analysis of progress-monitoring data and decision making. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 5, pp. 2133–2146). Bethesda, MD: National Association of School Psychologists.

Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2009). The complex nature of reading fluency: A multidimensional view. *Reading & Writing Quarterly, 25*, 4–32.

Ikeda, M. J., Neesen, E., & Witt, J. C. (2008). Best practices in universal screening. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 5, pp. 103–114). Bethesda, MD: National Association of School Psychologists.

Kaminski, R., Cummings, K. D., Powell-Smith, K. A., & Good, R. H. (2008). Best practices in using dynamic indicators of basic early literacy skills for formative assessment and evaluation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 5, pp. 1181–1204). Bethesda, MD: National Association of School Psychologists.

Knight-Teague, K., Vanderwood, M., & Knight, E. (2014). Empirical investigation of word callers who are English learners. *School Psychology Review, 43*(1), 3–18.

Kuhn, M., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and the definitions of fluency. *Reading Research Quarterly, 45*, 230–251.

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

Lesaux, N. K., & Siegel, L. S. (2003). The development of reading in children who speak English as a second language. *Developmental Psychology, 39*(6), 1005–1019.

Mellard, D., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 local schools. *Learning Disabilities Research & Practice, 24*(4), 186–195.

Perfetti, C. A. (1985). *Reading ability*. London: Oxford.

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. doi:10.1016/j.jsp.2010.09.004.

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.

Rasinski, T., Rikli, A., & Johnston, S. (2009). Reading fluency: More than automaticity? More than a concern for the primary grades? *Literacy Research and Instruction, 48*(4), 350–361.

Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M., & Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology, 96*, 119–129.

Slocum, T. A., Street, E. M., & Gilberts, G. (1995). A review of research and theory on the relation between oral reading rate and reading comprehension. *Journal of Behavioral Education, 5*, 377–398. doi:10.1007/bf02114539.

Smith, F., & Holmes, D. L. (1971). The independence of letter, word, and meaning identification. *Reading Research Quarterly, 6*, 394–415.

Stewart, L. H., & Silberglitt, B. (2008). Best practices in developing academic local norms. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 5, pp. 225–242). Bethesda, MD: National Association of School Psychologists.

U.S. Department of Education. (2012). ESEA flexibility. Retrieved 17 January, 2014, from http://www.ed.gov/esea/flexibility/documents/esea-flexibility-acc.doc.

U.S. Department of Education. (n.d.). Elementary and secondary education ESEA flexibility (website). Retrieved 17 January, 2014, from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html.

Zopluoglu, C. (2013). A comparison of two estimation algorithms for Samejima's continuous IRT model. *Behavior Research Methods, 45*(1), 54–65.