

Decision-Making Accuracy of CBM Progress-Monitoring Data

Journal of Psychoeducational Assessment
2018, Vol. 36(1) 74–81
© The Author(s) 2017
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0734282917729263
journals.sagepub.com/home/jpa



John M. Hintze¹, Craig S. Wells¹, Amanda M. Marcotte¹,
and Benjamin G. Solomon²

Abstract

This study examined the diagnostic accuracy associated with decision making as is typically conducted with curriculum-based measurement (CBM) approaches to progress monitoring. Using previously published estimates of the standard errors of estimate associated with CBM, 20,000 progress-monitoring data sets were simulated to model student reading growth of two-word increase per week across 15 consecutive weeks. Results indicated that an unacceptably high proportion of cases were falsely identified as nonresponsive to intervention when a common 4-point decision rule was applied, under the context of typical levels of probe reliability. As reliability and stringency of the decision-making rule increased, such errors decreased. Findings are particularly relevant to those who use a multi-tiered response-to-intervention model for evaluating formative changes associated with instructional intervention and evaluating responsiveness to intervention across multiple tiers of intervention.

Keywords

assessment of interventions/outcomes, diagnostic classification models, measurement, response to intervention (RTI)/multi-tiered system of supports (MTSS), statistical analyses, measurement

Curriculum-based measurement (CBM) is a set of standardized and specific measurement procedures that can be used to index student performance in the basic skill areas of early literacy and reading, early numeracy and mathematics computation and application, spelling, and written expression (Deno & Fuchs, 1987; L. S. Fuchs & Deno, 1991; A. L. Reschly, Busch, Betts, Deno, & Long, 2009; Shinn, 1989). As a variant of curriculum-based assessment, CBM uses dynamic general outcome measures of basic skill areas for making educational decisions such as screening for risk, instructional planning, and evaluating the effects of intervention efforts (Hintze, 2009; Shinn & Bamonto, 1998). When used within a response-to-intervention (RTI) model, CBM is well suited for (a) universal screening and obtaining point-estimates of basic skill performance to identify students potentially at-risk for academic problems, and (b) monitoring student responsiveness to instruction over time (Hintze, 2009).

When used to index student progress in response to intervention, CBM has shown to be sensitive to student change over time (e.g., L. S. Fuchs & Fuchs, 1986a, 1986b; Hintze & Christ, 2004;

¹University of Massachusetts, Amherst, MA, USA

²University at Albany, NY, USA

Corresponding Author:

John M. Hintze, University of Massachusetts, Furcolo S156, 819 N Pleasant St., Amherst, MA 01003, USA.
Email: hintze@educ.umass.edu

Hintze, Daly, & Shapiro, 1998; Hintze, Owen, Shapiro, & Daly, 2000; Nese et al., 2013). Despite this strong literature base, however, little is known with regard to the precision with which educators use such information in making accurate decisions, particularly those of a formative nature. While a number of studies have demonstrated that, when provided with such data, educators make more frequent decisions with respect to intervention effectiveness (Fuchs & Fuchs, 1986b; D. Fuchs, Mock, Morgan, & Young, 2003; L. S. Fuchs, Deno, & Mirkin, 1984; L. S. Fuchs & Fuchs, 1998), few studies have examined the accuracy of these decisions as a function of observed patterns in progress-monitoring data for individual students.

Two approaches to evaluating progress-monitoring data are commonly described in practitioner-oriented textbooks and in the peer-reviewed literature to determine student responsiveness to instruction using CBM data. The first, *trend line decision rules*, entails an interventionist comparing a child's estimated growth slope—regressing the student's CBM scores onto time—to a goal line of progress (Ardoin, Christ, Morena, Cormier, & Klingbeil, 2013; Hosp, Hosp, & Howell, 2007). Research has demonstrated that interventionists should be mindful of using such approaches, as establishing reliable slope differences using CBM requires more data than practitioners may assume is necessary (Christ, Zopluglu, Long, & Monaghan, 2012; Christ, Zopluglu, Monaghan, & Van Norman, 2013).

In contrast, L. S. Fuchs, Fuchs, and Hamlett (1990) and Shapiro (2004) recommended a *data point decision rule* (Hosp et al., 2007): whereby any time four data points are observed consecutively above or below an aimline that has been configured using peer performance and/or normative standards, then either a change in the long-term goal or instructional intervention should be made. Although this decision-making heuristic has certain face validity, what is not known is the extent to which such decisions are made on the basis of chance occurrences in the data as opposed to reliable trend and the resultant effects on diagnostic accuracy. Importantly, if observed data are variable and thus less reliable (e.g., as a function of the standard error of estimate [SEE]), practitioners might prematurely discontinue or augment an otherwise effective intervention due to what appears to be intervention effectiveness brought about by unreliability of measurement. In the latter, and often more serious case, a student who is declared to be unresponsive to current supports may be recommended for evaluation—a costly, time-consuming process—and/or be recommended to receive more intensive services that separate the student from their peers in Tier I (primary) instruction. Regardless, minimizing both types of errors by means of application of valid decision-making rules serves to optimize student outcomes and make existing RTI service delivery frameworks efficient. In one of the few empirical studies on the topic of data point decision rules, Van Norman and Christ (2016a) investigated the validity of a 3-point decision rule, deriving results after parameters were specified from a large extant database. Their findings indicated that in all but the most extreme circumstances, such a rule-of-thumb was invalid for instructional decision making. Unresolved, however, is the accuracy of more stringent criteria, including the frequently referenced 4-point rule suggested by L. S. Fuchs et al. (1990).

This problem is due in part to a general lack of guidance as to what constitutes expected responsiveness to various interventions and the absence of agreed upon decision rules for when to make progress-monitoring or instructional changes (Ardoin et al., 2013). In the absence of such guidelines, educators often make decisions on the basis of visual analysis, which can be unreliable with single-case data of a developmental nature (DeProspero & Cohen, 1979; Harbst, Ottenbacher, & Harris, 1991; Mercer & Sterling, 2012; Ottenbacher, 1990; Park, Marascuilo, & Gaylord-Ross, 1990; Van Norman & Christ, 2016b). Another contributing factor is that the residual error—the SEE—surrounding the regression line varies based on the quality of the probes employed (Ardoin & Christ, 2009). When the SEE is relatively high, there will be more “bounce” in the data, resulting in a need for extended progress monitoring, unbeknownst to the interventionist. However, how varying, plausible levels of the SEE directly relates to shifting criteria for decision making within the context of data point decision rules is unknown.

Given these questions, the purpose of the current study was to examine the efficacy of the number of consecutive data points that are needed that fall above a trend line required to make accurate decisions regarding treatment effect while at the same time minimizing the likelihood that interventions and/or goals would be prematurely altered due to construct irrelevant information.

Method

A Monte Carlo simulation study was implemented to assess the false and true positive rates of five data point decision rules for determining whether a student is not making typical progress compared to an aimline defined by an expected growth of two-words correct per minute (wcpm) per week. The rules for determining whether a student was not making expected progress were based on the number of consecutive data points above or below the trend line, referred to as the x -point rule.

Using the likely range of SEEs, as reported by Christ and Silbergliitt (2007)¹ and Ardoin and Christ (2009), 20,000 time-series data sets were simulated per condition to model student reading growth of either two wcpm/wk or one wcpm/wk across 15 consecutive weeks of intervention. These values roughly correspond to those specified in the extant literature as indicative of typical and ambitious growth during the early elementary period (cf. Deno, Fuchs, Marston, & Shin, 2001; L. S. Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993; Nese et al., 2013). The two wcpm/wk data were used to assess the rate at which students are falsely identified as not progressing at two wcpm/wk (i.e., false positive rate; Model I) given the respective rule and SEE. The one wcpm/wk data were used to assess the rate at which students are correctly identified as not progressing at two wcpm/wk (true positive rate; Model II), that is, the student's aimline, or instructional goal, was two wcpm/wk.

Varying the SEE allowed for the examination of decision-making and diagnostic accuracy as a function of varying levels of reliability while holding performance growth and length of progress monitoring in weeks constant. In other words, reducing the SEE approximates the condition of observing an individual student's growth slope using probes of progressively higher quality and, ipso facto, containing less measurement error. If a student's true growth is superior to that of the aimline, increasing the SEE, and thereby scatter about the regression line, will likely result in an increasing frequency of false positive observations, which collectively may result in an instructional decision-making error when applying an x -point rule.

A total of 20,000 randomly generated 15-week progress-monitoring data sets were produced for each level of three factors: SEE (2 to 18 in increments of 2), x -point rule (3-, 4-, 5-, 6-, and 7-point rule), and growth rate (two- and one-word per week), making the study a fully crossed $9 \times 5 \times 2$ design. The data generation process consisted of a standard linear model with both fixed and random elements, defined as

$$wcpm_{ij} = \beta_{0i} + \beta_{1i}week_{ij} + \varepsilon_{ij}, \quad (1)$$

where β_0 was equal to 0, β_1 was specified as either growth of one or two wcpm per week depending on condition, and the residuals, ε_i , were specified as $\varepsilon_i \sim \text{Normal}(0, \sigma_l^2)$ for variance l , which corresponds to the defined levels of the SEE for the i th student at time point j . The 20,000 iterations per condition specified exceed that of other relevant simulation studies (e.g., Christ et al., 2013).

Once constructed, levels of decision-making diagnostic accuracy based on the proportion of 20,000 data sets that were correctly identified as either making progress or not making progress at two wcpm/wk was examined across the various levels of reliability and x -point rules.

Table 1. False and True Positive Rates Associated With Varying Levels of Error and the Number of Consecutive Data Points Falling Above or Below the Students Trend (Slope) Line.

		3-point rule	4-point rule	5-point rule	6-point rule	7-point rule
False positive rate						
SEE	2	0.74	0.36	0.14	0.05	0.02
	4	0.86	0.49	0.23	0.10	0.04
	6	0.89	0.55	0.27	0.12	0.05
	8	0.90	0.57	0.29	0.13	0.06
	10	0.91	0.59	0.29	0.13	0.06
	12	0.92	0.60	0.31	0.14	0.06
	14	0.92	0.60	0.31	0.15	0.07
	16	0.92	0.61	0.31	0.15	0.07
	18	0.93	0.61	0.32	0.15	0.07
True positive rate						
SEE	2	1.00	1.00	1.00	1.00	1.00
	4	1.00	1.00	1.00	0.99	0.98
	6	1.00	1.00	0.98	0.93	0.85
	8	1.00	0.98	0.91	0.80	0.67
	10	0.99	0.93	0.82	0.66	0.52
	12	0.98	0.89	0.73	0.55	0.40
	14	0.96	0.84	0.65	0.47	0.32
	16	0.95	0.79	0.59	0.40	0.26
	18	0.93	0.75	0.54	0.35	0.23

Note. The false positive rate was based on the proportion of observations that were falsely identified as not improving at two-words per week according to the respective rule in any point in a 15-week period. The true positive rate was based on the proportion of observations that were correctly identified as not improving at two words per week according to the respective rule in any point in a 15-week period. The observations in this case were improving at only one word per week. SEE = standard error of estimate.

Specifically, for Model I, if consecutive simulated data points, equal to the x -point decision rule, fell below a two wcpm/wk aimline, then the case was declared a false positive (i.e., the student was declared nonresponsive to intervention when they actually were responding to intervention). Similarly, for Model II, if consecutive data points, equal to the x -point decision rule, fell below a two wcpm/wk aimline, then the case was declared a true positive (i.e., the student was correctly identified as a nonresponder).

Results

Table 1 presents the true and false positive rates for SEEs ranging from 2.00 to 18.00 and for five x -point rules. For example, if a practitioner exercised a decision rule whereby anytime four consecutive data points were below the trend line, and the SEE associated with the trend line was 8, they would correctly conclude the student was not making expected progress (i.e., true positive) about 98% of the time when that student's slope of progress was actually one wcpm/wk. Conversely, however, practitioners would also incorrectly conclude in about 55% of decision-making occasions that a student's rate of growth was not two wcpm/wk when in fact it was two wcpm/wk. In this case, the variability in the data brought about by the rather large SEE leads to a biased indicator of growth. Formatively, decisions associated with the first example are not problematic as it would likely be the case that an instructional change would be recommended where it is likely necessary. In the second example, however, formative instructional changes would be indicated when no change in instruction is necessary. In the second example, the

student is actually responding positively to the intervention and the intervention would be prematurely terminated or altered when in fact it should probably have been continued unchanged.

Inspection of Table 1 demonstrates that as precision of measurement increases (i.e., reducing the SEE and thus increasing reliability), the false positive rate decreases. Here again for example, if a practitioner exercised a decision-making rule whereby if four consecutive data points were all above or below the trend line that the intervention would be changed; she/he would be making an incorrect decision in about 60% of the cases if the observed slope of improvement had an associated SEE in the range of 12.00 to 14.00. However, if precision of measurement could be improved and an SEE of 2.00 obtained, decision-making errors could be reduced to approximately 36%. In this case, practitioners would reduce their likelihood of making an incorrect decision by 24% by improvements in their measurement system.

Using typical SEE estimates in the range of 6 to 10, roughly corresponding to estimates of reliability reported by Christ and Silberglitt (2007), Ardoin and Christ (2009), and Hintze and Christ (2004), it can be seen that reductions in false positive rates can be significantly reduced by more cautiously requiring that more consecutive data points fall above or below the trend line. Here, commonly used 3- and 4-point decision-making rules lead to unrealistically high levels of false positives which result in prematurely altering an otherwise effective intervention. Requiring a minimum of five or six consecutive data points reduces errors in decision making by a minimum of 50%. The increased precision in decision making, however, comes at the expense of more stringent data-based decision-making rules.

Discussion

The purpose of the current study was to examine the diagnostic accuracy associated with the number of consecutive data points (x -point rule) falling above or below a trend line while holding the overall number of data points collected during intervention and the SEE of the trend line constant. Results not only converge with those of Van Norman and Christ (2016a) in casting a 3-point decision rule in doubt, but go further to suggest even 4- or 5-point rules may be invalid given high levels of the SEE (cf. Ardoin & Christ, 2009; Christ & Silberglitt, 2007; Hintze & Christ, 2004; Thornblad & Christ, 2014). If practitioners base their instructional decisions on progress-monitoring data with large amounts of error, as indicated by the SEE, the accuracy of their decisions will be sacrificed. Conversely, when sufficient data are collected of a reliable nature, the accuracy of making data-based formative decisions improves.

The findings from this study are particularly relevant for those who use a multi-tiered RTI model whereby continuous progress monitoring is used to evaluate the effects of targeted and intensive interventions as commonly delivered in Tiers II and III. In such models, the intensity of interventions and services are increased in intensity only after a child's response to such intervention has been shown to be less than adequate (Brown-Chidsey & Steege, 2010; National Association of State Directors of Special Education, 2005; D. J. Reschly, Tilly, & Grimes, 1999). In such multi-tiered models, errors in decision making are exacerbated not only by the possibility of inadvertently altering an otherwise effective intervention; but by possible changes in placement that might result as a function of under-responsiveness to intervention. In these situations, students might be prematurely moved to more intensive and/or restrictive educational environments when a change is not warranted. In these situations, errors in decision making not only affect instructional decisions but placement decisions as well.

Second, results of this study weigh heavily on the concept of treatment validity as conceptualized by L. S. Fuchs and Fuchs (1998). This approach to learning disability classification rests on the premise that students are not classified unless and until it has been demonstrated empirically that they are not benefiting from the general education curriculum. As defined, treatment validity refers to the extent to which assessment procedures contribute to beneficial outcomes for

individuals (Cone, 1989; Hayes, Nelson, & Jarrett, 1987). The central feature of treatment utility is that there needs to be a clear and unambiguous relationship between the data collected during assessment and intervention decisions. Moreover, the decisions made on the basis of hierarchically arranged assessment procedures must demonstrate incremental validity as well (Sechrest, 1963). As a multi-tiered RTI model rests on the sequential application of more intensive interventions, it is essential that the progress-monitoring assessment data validly reflect the incremental changes that occur as a function of students' response to intervention.

Lastly, results of this study relate directly to the notions of evidential and consequential bases of assessment and decision making as it relates to validity, relevance and utility, and the social consequences of assessment (Messick, 1995). Here, the consequential aspects of validity include evidence and rationales for evaluating the intended use and unintended consequences of score interpretation. Simply, negative effects of decision making should not derive from any source of test invalidity (Messick, 1989). In the case of progress monitoring, low rates of growth should not occur because the assessment is missing something relevant to the focal construct that, if present, would have permitted affected students to display their competence or response to intervention. Likewise, low rates of growth should not occur because measurement contains something irrelevant that interferes with students' rate of growth or response to intervention over time.

In summary, the purpose of this study was to examine the diagnostic accuracy associated with decision making as is typically done with CBM approaches to progress monitoring. Results indicated that the propensity of errors in decision making are strongly related to the level of reliability of measurement and the progressive stringency of the applied decision-making rules that are used to determine intervention responsiveness. Furthermore, commonly endorsed decision-making criteria (e.g., a 4-point decision rule), given common levels of the SEE, are likely not sufficiently reliable for use in practice. Those who use forms of CBM progress monitoring to formatively evaluate the effects of instruction should pay careful attention to both the reliability of progress-monitoring measurement and how intervention effectiveness is determined. This is perhaps even more important in multi-tiered RTI models where not only are intervention effectiveness evaluated but educational placement and the intensity of intervention as well.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note

1. Note that Christ and Silbergliitt (2007) reported likely values of the standard error of measurement, not the standard error of estimate. However, they suggest that these values "might be used to estimate the likely magnitude of the standard error of the estimate" (p. 143).

References

- Ardoin, S. P., & Christ, T. J. (2009). Curriculum-based measurement of oral reading: Standard errors associated with progress monitoring outcomes from DIBELS, AIMSweb, and an experimental passage set. *School Psychology Review, 38*, 266-283.
- 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.

- Brown-Chidsey, R., & Steege, M. W. (2010). *Response to intervention: Principles and strategies for effective practice* (2nd ed.). New York, NY: The Guilford Press.
- Christ, T. J., & Silbergitt, B. (2007). Estimates of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review, 36*, 130-146.
- Christ, T. J., Zopluglu, C., Long, J. D., & Monaghan, B. D. (2012). Curriculum-based measurement of oral reading: Quality of progress monitoring outcomes. *Exceptional Children, 78*, 356-373.
- Christ, T. J., Zopluglu, C., Monaghan, 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.
- Cone, J. D. (1989). Is there utility for treatment utility? *American Psychologist, 44*, 1241-1242.
- 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., 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*, 507-524.
- DeProspero, A., & Cohen, S. (1979). Inconsistent visual analysis of intrasubject data. *Journal of Applied Behavior Analysis, 12*, 573-579.
- Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research and Practice, 18*, 157-171.
- 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., & Mirkin, P. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*, 449-460.
- Fuchs, L. S., & Fuchs, D. (1986a). Curriculum-based assessment of progress toward long- and short-term goals. *The Journal of Special Education, 20*, 69-82.
- Fuchs, L. S., & Fuchs, D. (1986b). Effects of systematic formative evaluation on student achievement: A meta-analysis. *Exceptional Children, 53*, 199-208.
- Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing identification of learning disabilities. *Learning Disabilities Research and Practice, 14*, 204-219.
- Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1990). Curriculum-based measurement: A standardized long-term goal approach to monitoring student progress. *Academic Therapy, 25*, 615-632.
- 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.
- Harbst, K. B., Ottenbacher, K. J., & Harris, S. R. (1991). Interrater reliability of therapists' judgements of graphed data. *Physical Therapy, 71*, 107-115.
- Hayes, S. C., Nelson, R. O., & Jarrett, R. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist, 42*, 963-974.
- Hintze, J. M. (2009). Conceptual and empirical issues related to developing a response-to-intervention framework. *Journal of Evidence-Based Practices for Schools, 9*, 128-147.
- 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., 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.
- Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York, NY: The Guilford Press.
- Mercer, S. H., & Sterling, H. E. (2012). The impact of baseline trend control on visual analysis of single-case data. *Journal of School Psychology, 50*, 403-419.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: Macmillan.

- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.
- National Association of State Directors of Special Education. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: Author.
- Nese, J. F. T., Biancarosa, G., Cummings, K., Kennedy, P., Alonzo, J., & Tindal, G. (2013). In search of average growth: Describing within-year oral reading fluency growth across grades 1-8. *Journal of School Psychology, 51*, 625-642.
- Ottenbacher, K. J. (1990). Visual inspection of single-subject data: An empirical analysis. *Mental Retardation, 28*, 283-290.
- Park, H., Marascuilo, L. A., & Gaylord-Ross, R. (1990). Visual inspection and statistical analysis of single-case designs. *Journal of Experimental Education, 58*, 311-320.
- Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47*, 427-469.
- Reschly, D. J., Tilly, W. D., & Grimes, J. (Eds.). (1999). *Special education in transition: Functional assessment and noncategorical programming*. Longmont, CO: Sopris West.
- Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement, 23*, 153-158.
- Shapiro, E. S. (2004). *Academic skills problems: Direct assessment and intervention* (3rd ed.). New York, NY: The Guilford Press.
- Shinn, M. R. (Ed.). (1989). *Curriculum-based measurement: Assessing special children*. New York, NY: The 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, NY: The Guilford Press.
- Thornblad, S. C., & Christ, T. J. (2014). Curriculum-based measurement of reading: Is 6 weeks of daily progress monitoring enough? *School Psychology Review, 43*, 19-29.
- Van Norman, E. R., & Christ, T. J. (2016a). Curriculum-based measurement of reading: Accuracy of recommendations from three-point decision rules. *School Psychology Review, 45*, 296-309.
- Van Norman, E. R., & Christ, T. J. (2016b). How accurate are interpretations of curriculum-based measurement progress monitoring data? Visual analysis versus decision rules. *Journal of School Psychology, 58*, 41-55.