Author: Chung, S.
Title: CBM progress monitoring in reading and foreign-language learning for secondary-school students
Issue Date: 2018-06-26

Chapter 5

Stability of maze reading slopes: Implications for use of CBM maze scores for tracking the progress of secondary-school students

Partly based on:

Siuman Chung

Christine Espin

Claire Stevenson

Manuscript in preparation

Abstract

Curriculum-based measurement (CBM) is a procedure that teachers can use to monitor student progress and evaluate the effectiveness of instruction. Despite the large body of literature supporting the technical adequacy of CBM measures, an area of concern remains the stability of the individual slopes or growth rates produced by the measures. Most research on this topic has focused on scores from CBM reading-aloud measures. In this paper, the stability of slopes produced by scores from CBM maze measures is examined. Specifically, we examine (1) the influence of duration, schedule, and variation in maze scores on the stability of CBM maze reading slopes, and (2) the effects of assuming a linear vs. nonlinear growth rate on SEE. For research question 1, stability, operationalized as the standard error of the slope (SEb), was calculated for different durations, schedules, and levels of SEE. Results revealed that the denser the schedule and lower the SEE, the shorter duration needed to achieve stable slopes. Large instability was seen in the first few weeks of data collection, regardless of schedule and SEE. For research question 2, the effects of assuming linear vs. nonlinear growth rates were examined. Assuming a nonlinear growth rate did not improve SEE. Implications for teachers' use CBM maze scores as indicators of growth in reading are discussed.

Introduction

Curriculum-Based Measurement (CBM) is a tool that teachers can use to monitor students' learning progress and evaluate the effectiveness of instruction on that progress (Deno, 1985, 2003). Short, timed CBM measures are administered frequently, and the scores are used to represent students' academic progress. For each student, CBM scores are plotted in a progress graph in which a linear growth line, referred to as the slope line, is drawn. Teachers read and interpret the graphs to determine whether the student is improving/growing at a desirable rate, and thus, whether the instruction is effective. If growth is not at a desirable rate, instruction is changed. If growth is at a rate better than expected, the goal is raised.

Over the last 40 years, a large amount of research has been conducted on the technical adequacy of CBM measures across various academic areas, including reading, mathematics, written expression, and content-area learning (see, for example, reviews by Wayman et al., 2007; Foegen et al., 2007; McMaster & Espin, 2007; Espin et al., in press, or see also Chapter 2). The majority of this research has focused on the technical adequacy of CBM scores as performance level indicators, that is, on whether "static" scores from the measures adequately represent students' levels of performance. This research is referred to by L. S. Fuchs (2004) as *Stage 1* research. Much less research has been conducted at what L. S. Fuchs (2004) refers to as *Stage* 2. Stage 2 research focuses on the technical adequacy of CBM scores as progress or growth indicators, that is, on whether the slopes generated from scores on frequently administered CBM probes provide stable, sensitive, and valid estimates of growth (see Wayman et al., 2007).

The paucity of research at Stage 2 might be due in part to the technical challenges associated with carrying out this research. It is only in recent years that advancements in technology and statistical methods have made it possible to conduct in-depth and comprehensive investigation of the properties of CBM slopes. Nonetheless, as early as the mid-1980s and 1990s there were attempts to examine slope. For example, Good and Shinn (1990) and Shinn, Good, and Stein (1989) compared ordinary-least-squares and split-middle techniques for estimating reading growth with CBM reading-aloud scores, and found that ordinary-least-squares was more accurate and precise for growth estimation than was the split-middle technique. Skiba, Deno, Marston, and Wesson (1986) reported on the characteristics of level, slope, and standard error of estimate (SEE) for CBM reading-aloud data for students in grades 2 through 5, providing typical SEE levels at each grade level. L. S. Fuchs & Fuchs (1992) examined the ratio of SEE to the mean growth rate for scores from two different types of CBM reading measures: reading-aloud and maze-selection. (A maze is a text in which every seventh word is replaced by a multiple-choice item with one correct and two incorrect choices. Students read silently and select words as they read.) Results of L. S. Fuchs and Fuchs (1992) demonstrated that the ratio of SEE to mean growth was similar for scores from the two measures.

In recent years, advances in technology and statistics have allowed for a more comprehensive look at the technical characteristics of CBM slopes. Most of this research has focused on students at the elementary-school level and on CBM reading-aloud measures (see Petscher, Cummings, Biancarosa, & Fien, 2013; Wayman et al., 2007, and work by Christ and colleagues reviewed later in the paper). Much less research has focused on students at the secondary-school level or has included CBM maze-selection measures.

Research on CBM Reading Growth at the Secondary-School Level

A handful of studies have examined the technical adequacy of reading growth trajectories produced by frequent (weekly) CBM measurement at the secondary-school level (Chung, Espin, & Stevenson, 2018, or see also Chapter 4; Espin et al., 2010; Tichá et al., 2009).[8] Espin et al. (2010) and Tichá et al. (2009) examined the reliability, sensitivity, and validity of reading growth produced by scores from two CBM measures – reading aloud (also referred to as oral reading fluency) and maze – for 31 and 35 8th-grade students respectively. For reading aloud, students read aloud for 1 to 3 min from a passage, and the number of words read correctly (WRC) was scored. For maze, students read silently for 2 to 4 min from a maze passage, and the number of correct maze choices (CMC) was scored. In both studies, growth was monitored for 10 weeks, with one reading-aloud and one maze probe administered per week.

Alternate-form reliability coefficients, calculated across scores on adjacent weeks, were on average $r = .86$ for 1-min reading aloud and $r = .80$ for 3-min maze scores (Espin et al., 2010).[9] Hierarchical linear modeling (HLM) was used to examine sensitivity and validity of the growth trajectories. Sensitivity analyses focused on whether the mean growth rate was significantly different from zero, which would indicate that the measures were sensitive to change in performance. Scores from reading-aloud reflected minimal to no change over time, whereas the scores from maze reflected significant growth at 2-, 3-, and 4-minute time frames (Espin et al., 2010; Tichá et al., 2009). The validity of the growth trajectories was then examined to determine whether growth on CBM maze was related to performance level and growth on the criterion measures. The CBM maze growth trajectories were found to be significantly related to performance level and growth on standardized reading measures and to reading group membership (Espin et al., 2010; Tichá et al., 2009).

Although the results from Espin et al. (2010) and Tichá et al. (2009) provided initial support for the use of maze to monitor reading growth for secondary-school students, the studies involved small sample sizes and included only 10 weeks of progress monitoring.

[8] Other studies at the secondary-school level have examined growth, but have not included weekly measurements (Codding et al., 2015; Tolar et al., 2014; Tolar et al., 2012; Yeo et al., 2012). However, frequent data collection is necessary if CBM progress monitoring data are to be used for ongoing evaluation of instruction for students with learning difficulties as illustrated in the opening paragraph of this paper. Thus, in the current study we only reviewed the studies that included frequent measurements.

[9] In the study by Tichá et al. (2009), the alternate-form reliabilities were reported for scores on reading-aloud and maze measures that have been administered at one measurement point, and were on average $r = .96$ for 1-min reading-aloud scores and $r = .87$ for 3-min maze scores.

Thus, in a subsequent study, Chung et al. (2018, or see also Chapter 4) replicated this early research with a larger sample and for a longer progress-monitoring time period. In addition, Chung et al. (2018, or see also Chapter 4) fit both linear and nonlinear growth models to the data. Participants in the study were 452 7^{th}-grade students who completed one 2-min maze task each week for 23 weeks. Alternate-form reliability, calculated for maze scores on adjacent administration weeks, was on average r = .75. Results of multilevel analyses revealed that both linear and nonlinear (logistic) growth models fit the data, although the fit was better for the nonlinear than the linear growth model. For both linear and nonlinear models, level and growth on the maze were found to be significantly related to scores on multiple criterion measures, including school level, dyslexia status, and performance level and growth on three external reading measures. The results of Chung et al. (2018, or see also Chapter 4), Espin et al. (2010), and Tichá et al. (2009) provide support for the alternate-form reliability, sensitivity, and validity of maze scores as indicators of reading growth for secondary-school students. However, missing from these studies is consideration of one of the most essential factors in CBM growth measurement – stability. Stability refers to the magnitude of measurement error associated with reading slopes for individual students. The larger the measurement error associated with a slope, the less likely it is that the slope is reflecting "true" growth (Hintze & Christ, 2004). Stability is important within CBM growth measurement to answer questions related to how often and how long students must be monitored before the slopes generated by the CBM measures are "trustworthy". These questions lie at the core of CBM progress monitoring because for teachers to use CBM progress data to inform instruction, they must be able to use the data in a "timely fashion" to make instructional decisions about the effectiveness of their interventions, and to change instruction if it is not effective. In recent years, Christ and colleagues (Christ, 2006; Christ, Monaghen, Zopluoglu, & van Norman, 2012; Christ, Zopluoglu, Long, & Monaghen, 2012; Christ et al., 2013) have examined the question of slope stability for CBM reading-aloud measures at the elementary-school level.

Stability of Slopes for Scores from CBM Reading-Aloud Measures

In 2006, Christ and colleagues began a series of studies designed to examine the stability of slopes generated by CBM reading-aloud scores, and to determine factors that influenced that stability (Christ, 2006; Christ, Monaghen, et al., 2012; Christ, Zopluoglu, et al., 2012; Christ et al., 2013). The factors examined in the studies included duration, schedule, and variation in scores. Duration was the number of weeks of progress monitoring. Schedule was the frequency of progress monitoring, and included the number of occasions (the number of measurements, per week or per month), and the number of observations (the number of passages administered per occasion). Variation was the intra-individual variation in the data around the slopes, and was reported as either SEE or residual variance in scores.

In the first in this series of studies, Christ (2006) examined the effects of duration and SEE on the stability of the slopes. Results revealed that the lower the SEE and the longer the data-collection duration, the more stable the slopes; however, the relation between these variables was curvilinear. Thus, with only a few weeks of data collection (for example, 2 to 4 weeks), SEE had a large influence on the stability, whereas for longer durations (for example, 12 to 15 weeks), SEE had minimal influence on stability.

In subsequent studies, Christ and colleagues (Christ, Monaghen, et al., 2012; Christ, Zopluoglu, et al., 2012; Christ et al., 2013) used a simulated data set with reading-aloud scores for 9000 hypothetical students. Duration and variation were again examined, and a third factor, schedule of data collection, was also considered (Christ et al., 2013). Schedules of data collection included one, two, three or five measurement per week or month (occasions) with either one or three measures administered per occasion (observations per occasion). In addition, reliability and validity coefficients of the slopes were estimated for different durations, schedules, and levels of score variation.

In general, the results of the simulation studies with regard to duration and variation were similar to those found in Christ (2006). The stability of the reading-aloud slopes showed a similar curvilinear trend, with high levels of instability found for shorter progress-monitoring durations and stability increasing rapidly with longer durations. Lower levels of score variation resulted in more stable, reliable and valid reading slopes (Christ, Monaghen, et al., 2012; Christ, Zopluoglu, et al., 2012; Christ et al., 2013). With regard to schedule, results revealed that the denser the schedule, the more stable, reliable, and valid the reading slopes. For instance, when score variation was low, reliable and valid slopes were seen after only 8 weeks of data collection with 5 to 6 observations per week, after 10 weeks of data collection with 3 observations per week, and after 13 weeks with 3 observations per month (Christ et al., 2013).

In sum, the studies by Christ and colleagues demonstrate that duration, schedule, and variation in scores affect the stability of the slopes generated by CBM reading-aloud scores for elementary-school students. What is not yet known is whether and to what extent such factors influence the stability of the slopes generated by CBM maze scores for secondary-school students. In addition, the studies done by Christ and colleagues all assumed linear growth models, yet growth on CBM measures may be nonlinear. For example, studies have shown that growth on CBM measures is greater from fall-to-winter than from winter-to-spring (e.g., Christ et al., 2010; Keller-Margulis, Mercer, Payan, & McGee, 2015; Nese et al., 2012; Shin, Espin, Deno, & McConnell, 2004). Assuming a nonlinear growth model might result in a decrease in score variation.

Purpose of the Study and Research Questions

The purpose of the present study is to examine the stability of the growth trajectories generated by scores on frequently administered CBM maze measures. The study involves a reanalysis of data from Chung et al. (2018, or see also Chapter 4), and examines the influence of duration, schedules of data collection, and variation (levels of SEE) on slope stability. We adopt the approach used by Christ and colleagues in their examination of slope stability for CBM reading-aloud measures, but extend their work by examining the effects on SEE of assuming a nonlinear growth model. Two research questions are addressed:

1. What are the effects of duration, variation in maze scores, and schedule of data collection on the stability of CBM maze slopes?
2. Can standard error of the estimate be decreased by assuming a nonlinear (logistic) growth rate rather than a linear growth rate?

Method

The current study was a reanalysis of data from Chung et al. (2018, or see also Chapter 4). A short overview of the participants, instruments and procedure relevant to this study are described here. For more elaborate details, see Chung et al. (2018, or see also Chapter 4).

Participants

Participants were 452 (233 male) Dutch 7th-grade students from three secondary schools. Students were M = 12.63 (SD = 0.63; range: 12-15) years old. Participants spanned a wide range of performance/school levels. In the Netherlands, secondary schools are divided into four levels: practical education, pre-vocational, senior general secondary, and pre-university education (ordered from highest to lowest, Dutch Ministry of Education, Culture and Science, 2005). The number of participants per level were: practical (n = 55), pre-vocational (n = 321), and senior general secondary/pre-university (n = 76). Students with dyslexia (12%) were represented in all school levels.

Instruments and Procedure

Student growth was measured using CBM maze tasks. A maze is a reading passage in which every seventh word is deleted and replaced by a multiple-choice item with the correct word and two distractors. Students read silently through the passage for 2 min selecting words as they read. The number of correct choices was scored.

Mazes were administered and scored through an online progress-monitoring system called *Mazesonline* (http://www.mazesonline.nl). Students completed one maze per week for a period of 23 weeks. The maze was not administered during vacation or exam weeks, resulting in a total passage set of 17 maze passages.

Data Analyses

Variation in maze scores

Based on the observed maze scores, the beginning level of performance at the start of the study (intercept) and the rate of growth (slope) were calculated for each individual student, after which the variation around the slope (i.e., standard error of the estimate; SEE) was estimated per student. The SEE was estimated by applying the following sample estimation formula:

$$SEE_i = \sqrt{\frac{\Sigma(Y_{it} - Y'_{it})^2}{n_i - 2}}\,, \qquad (1)$$

where Y_{it} was the observed maze score of individual i at time t, and Y_{it}' was the predicted maze score of individual i at time t, and n_i was the total number of maze tasks administered to individual i. We calculated the mean SEE from the sample, and used this level of SEE to represent typical levels of SEE for 7th-grade students completing mazes on a weekly basis. We then varied SEE relative to this typical level.

Stability of maze slopes

To examine the influence of duration, schedule and various levels of SEE on the stability of maze growth rates, we calculated the standard error of the slope (SEb; stability) by applying the following formula for various scenarios:

$$\text{SEb} = \frac{SEE}{SD_x\sqrt{n}}\,, \qquad (2)$$

where *SEE* was the standard error of the estimate, x was the duration, and n was the number of maze tasks collected. The SEb was calculated in Excel for SEEs ranging from two to five, and durations of up to 30 weeks.[10] Schedules included one observation twice a week, once a week, once in two weeks, and once a month.

Linear vs. nonlinear growth

To address research question 2, SEE was compared for linear vs. nonlinear (logistic) growth. For both functional forms (linear and nonlinear), two approaches were used to estimate the predicted slope lines of individual students: ordinary least squares regression (OLS) and multilevel modeling (ML; Hox, 2010). Both estimation approaches represent individual learning growth, but differ in practicality and the accuracy of "true" learning growth. While

[10] We first calculated the daily SEb, and then converted this to weekly SEb by multiplying the daily SEb by 7. For all schedules, the first administration was set on day one. The twice a week schedule was set with intervals of 3 to 4 days, thus administrations occurred on day 1, 4, 8, 11, 15, etc. The once a week schedule was set with intervals of 7 days; thus administrations occurred on day 1, 8, 15, 22, 29, etc. The once every two weeks schedule was set with intervals of 14 days, thus administrations occurred on day 1, 15, 29, 43, 57, etc. The once a month schedule was set with intervals of 30 to 31 days, thus administrations occurred on day 1, 32, 62, 93, 123, etc. Missing values were imputed by interpolating the data with Excel.

ML is more accurate in representing "true" growth, because the method controls for missing data and dependence between measurements (Hox, 2010), OLS is more practical, because it can be used more easily in the educational practice by teachers. OLS slopes can be graphed from individual student's scores by using, for instance, free templates that are available online (e.g., http://www.measuredeffects.com). For graphing ML slopes, additional data from a large representative sample is required, and thus, in practice, less easy to use. Because both arguments – accuracy of "true" learning progress and practical use – are important, both estimation approaches were included in the analyses. The difference in SEE for linear vs. nonlinear growth were conducted separately per estimation approach (OLS and ML). We will describe both estimation approaches in more detail.

OLS consisted of only a within-individual level, whereas ML consisted of both within- and between-individual levels. At the within-individual level, the following formulas were applied to estimate the predicted maze scores for linear and nonlinear (logistic) growth respectively:

$$Y'_{it} = \pi_{0i} + \pi_{1i} * maze\ session_{it}, \text{ and} \quad (1)$$
$$Y'_{it} = \pi_{0i} + \pi_{1i}\ (\ln(maze\ session_{it} + 1)) \quad (2)$$

where the π_{0i} and π_{1i} were respectively the intercept and slope of the predicted growth line in both formulas. The between-individual level expanded the formulas above with:

$$\pi_{0i} = \beta_{00} + u_{0i} \quad (3)$$
$$\pi_{1i} = \beta_{10} + u_{1i} \quad (4)$$

where β_{00} was the overall mean intercept to predict π_{0i} from a between-individual level for individual i. The u_{0i} was the error term for π_{0i}, thus the difference between the overall mean intercept and the intercept of individual *i*. The β_{10} was the mean slope to predict π_{1i} from a between-individual level for individual i. The u_{1i} was the error term for π_{1i}, thus the difference between the overall mean slope and slope of individual *i*.

Differences in standard errors of estimate (SEEs) were examined with the non-parametric Wilcoxon Signed Rank test. Wilcoxon Signed Rank tests were used because assumptions for the paired samples t-tests were not met. The formula used to estimate SEE consisted of an observed maze score Y_{it} and a predicted maze score Y_{it}' of individual *i* at time *t*, see equation 1.

The intercepts and slopes for linear and nonlinear growth for OLS and ML estimation were calculated by using the packages plyr (Wickham, 2011) and lme4 (Bates & Maechler, 2010) in the software program R. The SEEs were calculated in plyr in R or manually with Excel. The Wilcoxon Signed Rank tests were performed in IBM SPSS.

Results

Selecting "Typical" and "Optimal" Standard Error of the Estimates (SEEs)

We first determined which values to use to represent "typical" and "optimal" SEEs. The standard error of the estimate (SEE) of reading slopes for the weekly administered 2-min maze passages for the sample was $M = 4.21$ ($SD = 1.67$), with a range of 0.92 – 13.01 SEE, with 73.5% of the cases falling below an SEE = 5. Based on the obtained mean, SEEs of 4 and 5 were selected to represent "typical" SEEs, whereas SEEs of 2 and 3 were used to represent more "optimal" SEEs. (Recall that a lower SEE indicates that there is less error between observed and "true" scores).

Stability of Reading Slopes

To examine the stability of maze slopes, the standard error of the slope (SEb) was calculated for different durations, levels of SEEs, and schedules. We report on the effects of duration and SEE, and on duration and schedule, on the stability of maze slopes.

Duration and SEE

In Figure 5.1 the SEb's are plotted for weekly-administered maze passages with more optimal (SEE= 2 and 3) and typical (SEE= 4 and 5) SEEs. At all levels of SEE, curves showed a steep decelerating trend in the first weeks, followed by stable straight lines thereafter. Stable lines were achieved sooner (i.e., SEb was lower) for lower levels of SEE than for higher levels of SEE. For instance, given an optimal SEE of 2, the SEb at week 8 was 0.26, and reductions in SEb were minimal (smaller than 0.05) each week thereafter. In contrast, given a typical SEE of 4, the SEb at week 8 was 0.52, and it was not until week 10 that the weekly reduction was smaller than 0.05 per week. At week 22 the differences in SEb for different levels of SEE were negligible (less than 0.10).

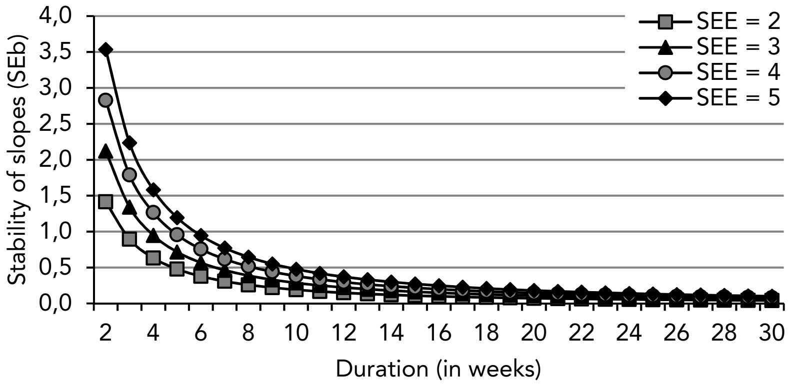

Figure 5.1. Stability of the slopes (SEb) for different SEEs on weekly administered maze passages

Duration and schedule

In Figure 5.2 the stability of the slopes for different schedules are plotted given a typical SEE of 4. For all schedules, curves showed a steep decelerating trend in the first weeks, followed by stable straight lines thereafter. Stable lines were achieved sooner (i.e., SEb was lower) for higher density schedules. For instance, given a dense schedule of data collection (twice per week), the SEb at week 9 was 0.34, and reductions in SEb were minimal (smaller than 0.05) each week thereafter. In contrast, given a less dense, but more typical schedule of data collection (once per week), the SEb at week 9 was 0.44, and it was not until week 10 that the weekly reduction was smaller than 0.05 per week. At week 27 the differences in SEb for different data collection schedules were negligible (less than 0.10).

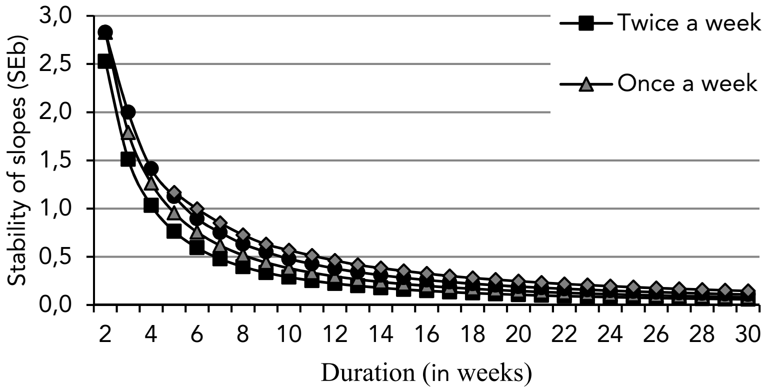

Figure 5.2. Stability of the slopes (SEb) for different schedules and SEE = 4

Linear vs. Nonlinear Growth

In Table 5.1 descriptive statistics are provided for the intercepts, slopes, and SEE for linear and nonlinear growth as estimated with ordinary least squares regression (OLS) and multilevel modeling (ML). Average SEEs ranged from 4.2 to 4.5, and were similar for linear and nonlinear growth. The obtained SEE was higher for ML estimation (SEE = 4.5) than for OLS estimation (SEE = 4.2). The range of the SEE was large, ranging from approximately 1 to 15. Results revealed that SEEs did not differ significantly for linear and nonlinear growth using an OLS estimation ($Z = -.82$, $p = .41$) or ML estimation ($Z = -1.19$, $p = .24$).

Table 5.1
Intercepts, slopes, and SEEs for linear and nonlinear slopes with OLS and ML estimation

	M	*SD*	Min	Max
OLS				
Intercept linear	25.70	7.35	9.47	52.32
Slope linear	0.16	0.39	-1.19	2.20
Intercept nonlinear	24.77	7.29	6.43	51.06
Slope nonlinear	1.16	2.20	-7.06	10.87
SEE linear	4.21	1.67	0.92	13.01
SEE nonlinear	4.23	1.68	1.01	12.64
ML				
Intercept linear	25.71	6.40	11.36	47.66
Slope linear	0.15	0.14	-0.40	0.62
Intercept nonlinear	24.70	5.83	11.87	44.70
Slope nonlinear	1.20	0.94	-2.14	4.22
SEE linear	4.50	1.74	0.94	14.50
SEE nonlinear	4.50	1.75	1.14	14.52

Discussion

This research addressed the second stage of CBM research (L. S. Fuchs, 2004) - the technical adequacy of CBM slopes for secondary-school students. Specifically, we examined the stability of the reading slopes produced by maze scores and factors that affect that stability. We focused on the influence of duration, schedule, and variation of maze scores (SEE) on the stability of maze slopes, and examined whether assumption of a nonlinear growth model would result in reduced variation in the data.

Results revealed that the stability of maze slopes was affected by duration, schedule, and variation in maze scores (SEE). The denser the schedule and the lower the variation in scores, the shorter the duration needed to achieve stable maze slopes. Relatively large instability was seen in the first few weeks of data collection, regardless of schedule or variation in scores. These results are similar to those of Christ and colleagues in the sense that duration, schedule, and variation in scores affected the stability of the slopes (Christ, 2006; Christ et al., 2013), and that stability increased rapidly with longer duration.

With regard to the effects of SEE and duration, given a schedule of once a week data collection, it was not until approximately weeks 8 to 11 that the effects of SEE began to decrease for maze slopes. At week 12, the decrease of stability for each week was minimal. Such results suggest that if monitoring is being conducted with a known passage set that has been shown to produce relatively low levels of SEE for a given population, then instructional decisions could reasonably be made on the basis of slope after a relatively short period of time, for example, perhaps after only 9 or 10 weeks of progress monitoring. If, however, the SEE for a passage set is unknown for a given population, then it would be advisable to monitor the progress for a longer period of time, for example, 12 to 14 weeks, before making instructional decisions on the basis of the maze slope.

Decisions about duration can also be influenced by the schedule of data collection. Given an SEE of 4, if data are collected on a "thin" schedule, for example, once every two weeks or once a month, then it is necessary to collect data over a longer period of time before slopes stabilize. However, for denser schedules of data collection, for example, once or twice a week, slopes stabilize earlier (that is, stability does not decrease substantially with additional weeks of data collection). Thus, for thin schedules of data collection, slopes do not stabilize until about 12 to 13 weeks, whereas for dense schedules of data collection, slopes stabilize at about 10 to 11 weeks. Differences in slope stability between once- and twice-a-week data collection disappear fairly quickly, with only with minimal differences seen between the two schedules of data collection at 10 to 11 weeks.

Although we found that after approximately 12 weeks, slope stability levels off, it is still worrisome that the standard error of the slope given a "typical" SEE of 4 was about 0.3 for once- or twice-a-week data collection. Although 0.3 seems quite small, it is large when one considers that the mean rate of growth for the 7th-grade students included in this study was approximately 0.2 correct maze choices per week (Chung et al., 2018, or see also Chapter 4). Such variation could lead to incorrect instructional decisions for some students. For instance, if a student (with an average amount of variation in his or her slope) has an observed growth rate of 0.3 correct maze choices, the "true" rate of growth might lie between 0 to 0.5. An educator might decide to not change the student's instruction based on the observed growth rate of growth of 0.3 given that this growth exceeds the typical rate of growth for students at this age; however, if the "true" rate of growth for the student is actually 0, then the correct decision would be to change instruction. Similarly, educators may unnecessarily change instruction that has been effective if the student's "true" growth rate exceeds their observed growth rate. An additional consideration is the fact that the *average* SEE was approximately 4. As mentioned earlier, about 25% of the sample had SEEs of 5 or higher, with some students having SEEs as high as 13. The chances of incorrect instructional decision-making is especially large for students with much variation in their slopes.

To address problems of variation, one can implement longer progress-monitoring durations with denser schedules of data collection, but it is also important to try and reduce the maze variation seen in slopes. As can be seen in the current study, the less variation, the faster stable maze slopes are achieved. One way to reduce variation in maze slopes is to provide optimal conditions within the classroom when CBM mazes are administered. For instance, educators should be sure to administer the probes according to guidelines and as intended, to provide a quiet setting in which students can complete the CBM mazes with enough concentration, to motivate students' to do their best, and to carefully score and graph the CBM maze scores (Haladyna & Downing, 2004). In addition, there is also room for improvement for developers and/or (future) researchers to reduce variation in maze scores via attention to the characteristics of the passages used to monitor progress. One factor that likely influences the variation in maze scores is the extent in which frequently administered passages are equivalent to one another. Approaches to addressing problems related to passage characteristics have been described for scores from CBM reading-aloud measures, and include removing outlier passages from a passage set in which scores on passages had the largest Euclidean distance (Christ & Ardoin, 2009), or by statistical equating of the scores in the passage set (Albano & Rodriguez, 2012; Santi, Barr, Khalaf, & Francis, 2016).

Although reducing the variation in maze slopes may increase the stability, our data reveal that even for low levels of SEE, for example SEE = 2, the standard error of the slope is still approximately 0.15. Considering a mean growth rate of 0.2, this standard error might still be considered to be too large. It will be important in future research to examine ways to increase the rates of growth typically seen for maze data. One obvious approach is to adjust the maze passage slightly. For instance, by increasing the number of possible maze choices per passage by, for example, replacing every fifth word in the passage rather than every seventh word, scores per passage would be greater. It also might be possible to administer longer maze passages with a longer administration time, or to administer more passages at each occasion. A different approach would be to support the maze score with an additional score from a different CBM probe, and to combine the two scores as one indicator of reading progress. Both approaches would increase the score on the CBM progress measure, and possible also the growth rates. However, it is not yet known whether the amount of variations in the slopes would remain constant, or whether such measures would be reliable and valid indicators of reading level and growth, thus additional research should be conducted if these approaches would be considered.

A final consideration with regard to growth rates is to examine growth under "optimal" rather than under "typical" conditions. In our study, we did not implement reading interventions for participating students. It may be that under "optimal" instructional conditions where students receive evidence-based reading instruction, their rates of growth on the CBM measures would be much larger. Supporting this idea is research by Deno et al. (2001) who demonstrated that growth on CBM reading measures was much larger for students with learning disabilities under optimal instructional conditions than under typical instructional conditions.

A second purpose of our study was to examine whether assumption of a nonlinear growth rate would result in reduced SEE. Recent studies have suggested that CBM growth measures might not be linear but nonlinear (Christ et al., 2010; Chung et al., 2018, or see also Chapter 4; Keller-Margulis et al., 2015; Nese et al., 2012; Shin et al., 2004); If students' reading growth is in fact nonlinear, then the variation in scores around the nonlinear growth line (SEE) might be smaller. Our results revealed, however, that variation was not reduced by assuming a nonlinear growth model. The variation in maze scores were the same for both linear and nonlinear growth models, regardless of whether OLS or ML estimation approaches were used.

Our results are seemingly at odds with the original data analyses by Chung et al. (2018, or see also Chapter 4), which demonstrated that a nonlinear growth model provided a better fit to the data than a linear growth model. The seemingly conflicting results might be explained by the choice of analysis method. Our study used the Wilcoxon Signed Rank test that tests differences based on matched pairs of residuals (SEE linear and nonlinear), whereas Chung et al. (2018, or see also Chapter 4) compared fit indices that were based on the sum of residuals for the two full growth models (AIC, BIC, Deviance of linear vs. nonlinear). At the very least, the results of the present study support the use of a linear growth model for monitoring students' progress with CBM measures. A linear growth model is more practical in terms of generating progress graphs and ease of interpretation for educators. It is worthwhile to note however, that in a recent study that compared linear and nonlinear growth for students, not all students showed one type of growth trajectory. On reading-aloud scores, some students followed a linear learning curve, whereas other students followed a nonlinear learning pattern (van Norman & Parker, 2016). Such results indicate that variation in scores might be smaller if the individual growth trajectories were taken into account. The issue of individual growth trajectories should be considered in future research.

Limitations

A limitation of the study was the selection of "typical" levels of the variation in maze scores (SEE). These levels would preferably be set based on a range of studies and data sets to be able to gauge what the expected amount of variation in maze scores is for secondary-school students. However, to date there are only a handful studies that have examined reading growth for secondary-school students. Of those studies, only one has reported the SEE of the reading slopes (Tolar et al., 2012), but in that study reading growth was based on only five measurements occasions throughout the school year. If more data were available, "typical" levels of SEE might be different. That said, results related to the effects of duration, variation, and schedule on the stability of the slopes would not change. More research is needed on the use of CBM reading measures to monitor the growth of secondary-school students, and especially to establish levels of SEE for student growth.

Conclusions and Future Directions

The current study extended CBM research at the secondary-school level, and expanded the research conducted at the second stage (technical adequacy of CBM reading slopes) of CBM measure development. This study was the first to examine the stability of maze slopes for secondary-school students and to examine an approach for reducing variation in maze scores. The development of CBM progress monitoring in reading for secondary-school students is a relatively new area of research, and many questions remain. Our study illustrates the need to examine methods for reducing variation in maze scores and thereby improving slope stability. It also illustrates the importance of examining growth under optimal rather than just typical instructional conditions. Finally, our present study focused solely on the technical characteristics of scores from the CBM measures. It did not address the essential question of CBM progress monitoring: *Does teacher use of CBM progress-monitoring result in improved reading instruction for and improved performance of students with learning disabilities?* Using the CBM progress data for instructional decision-making is the essence of CBM. This issue has yet to be addressed at the secondary-school level.