Read Writ (2012) 25:887–915
DOI 10.1007/s11145-011-9304-0

Within-year oral reading fluency with CBM: a comparison of models

Joseph F. T. Nese · Gina Biancarosa · Daniel Anderson · Cheng-Fei Lai · Julie Alonzo · Gerald Tindal

Published online: 18 March 2011

Abstract This study examined the type of growth model that best fit within-year growth in oral reading fluency and between-student differences in growth. Participants were 2,465 students in grades 3–5. Hierarchical linear modeling (HLM) analyses modeled curriculum-based measurement (CBM) oral reading fluency benchmark measures in fall, winter, and spring with grade level and student characteristics (including special education and Limited English Proficiency status) as covariates. Results indicated that a discontinuous growth model fit the data better than a linear growth model, with greater growth in the fall than in the spring. Oral reading fluency growth rates also differed by grade and student characteristics. Implications for school practice and research are discussed.

Keywords Oral reading fluency · Curriculum-based measurement · Growth modeling

Modeling

Curriculum-based measurement (CBM) generally, and oral reading fluency (ORF) specifically, are widely used as benchmark and progress-monitoring measures, but little empirical work has attempted to examine intra-individual change in CBM ORF. Although norms exist for different benchmark periods (i.e., fall, winter, and spring), these are cross-sectional norms. Those studies that have examined growth have largely adhered to linear models, which assume a constant rate of growth

J. F. T. Nese (✉) · D. Anderson · C.-F. Lai · J. Alonzo · G. Tindal
Behavioral Research and Teaching, University of Oregon, 175 Education, 5262, Eugene, OR 97403, USA
e-mail: jnese@uoregon.edu

G. Biancarosa
University of Oregon, 102 Education, 5267, Eugene, OR 97403-5267, USA

over time. This study attempts to add to our understanding of what growth in CBM ORF benchmark measures looks like in grades 3–5, and whether student characteristics including grade-level predict systematic differences in growth.

CBM is an assessment process designed to give teachers feedback on the effectiveness of their instruction in producing academic learning gains (Deno, 1985). CBM uses direct assessments to evaluate student performance, is based on general grade-level outcome measures, and helps guide teachers' instruction by identifying specific student deficits. CBM can be used in a number of different ways, but is primarily used in the Response to Intervention (RTI) framework for benchmark screenings and progress monitoring (Deno, 2003; Keller-Margulis, Shapiro, & Hintze, 2008). Educators can administer a universal benchmark screening assessment to identify students at-risk for low achievement, and also monitor the progress of students targeted for intervention (Shinn, 2007).

Benchmark screening tests are typically administered to all students in a school in the fall, winter, and spring as part of a universal screening process. Most CBM systems offer benchmark ORF measures that assess all students three times per year for universal screening; general education progress monitoring; and adequate yearly progress (AYP) accountability. Schools using an RTI model often use the results of these tri-annual benchmark assessments for early identification of students at-risk for poor learning outcomes, and also to make inferences about the progress of the entire school population, classrooms, or specific subgroups of students. Academic decisions (e.g., instruction or intervention) are then based on the progress, or growth, students make in comparison to the expected rate of improvement. Depending on the academic progress tracked over time, or growth rate, the identified students may be selected for increasingly intensive interventions, including special education evaluation and/or placement (Fuchs, Mock, Morgan, & Young, 2003).

CBM has been rigorously researched and a considerable number of studies have shown the psychometric features of CBM to be robust (Deno, Fuchs, Marston, & Shin, 2001; Fuchs, 2004; Good & Jefferson, 1998; Kranzler, Brownell, & Miller, 1998; Shinn, Good, Knutson, Tilly, & Colllins, 1992). ORF is the most researched CBM and research has shown strong, positive correlations between CBM ORF and reading comprehension (range .70 s to .80 s; Kranzler et al., 1998) and state standardized reading tests (range .60–.75; Shapiro, Keller, Santoro, & Hintze, 2006). With the increased attention to and use of CBM, it is important that growth patterns on these measures be more closely examined.

Interpreting oral reading fluency growth

A limited number of researchers have examined typical reading fluency growth using CBM. Table 1 summarizes the results of these studies.

The first study to investigate typical academic growth patterns with CBM was that of Fuchs, Fuchs, Hamlett, Walz, and Germann (1993). Using Ordinary Least Squares regression (OLS), the authors estimated weekly growth in ORF for a small sample of students ($n = 117$) in grades one to six who were tested monthly.

Table 1 Summary of research estimates of annual and seasonal weekly reading fluency growth in grades 3–5

	n	Analysis	Reading fluency CBM weekly growth (cwpm)								
			Grade 3			Grade 4			Grade 5		
			Annual	Fall-Winter	Winter-Spring	Annual	Fall-Winter	Winter-Spring	Annual	Fall-Winter	Winter-Spring
Fuchs et al. (1993)	197	OLS; linear	1.00	–	–	0.85	–	–	0.50	–	–
Deno et al. (2001)	907	OLS; linear	0.94	–	–	1.03	–	–	0.58	–	–
AIMSweb© (Pearson 2011)	574,973–779,526	Linear	1.00	1.22	0.78	0.83	0.94	0.72	0.81	0.83	0.78
Hasbrouck and Tindal (2006)[a]	49,933	Descriptive	1.13	1.31	0.94	0.91	1.13	0.69	0.91	1.06	0.75
Graney et al. (2009)	427	ANOVA; grade by time	0.75	0.71	0.79	0.71	0.49	1.07	0.66	0.45	0.95
Christ et al. (2010)[a]	<3,808[b]	HLM; discontinuous	1.18	1.38	0.97	0.99	1.19	0.78	0.98	1.12	0.83
Present study[a]	2,465		0.95	1.78	0.12	0.91	1.36	0.45	0.67	0.53	0.81

[a] Fall-Winter and Winter-Spring growth parameters were estimated based on a 36 week school year; weekly growth scores were calculated by dividing the difference between benchmark scores by the number of weeks between benchmark assessments

[b] The precise sample size for grade 3–5 students could not be determined from the manuscript

The authors reported that a linear model fit the data better than a quadratic model; however, it is important to note that the quadratic term was significant at $p < 0.05$, revealing a slightly negatively accelerating pattern of within-year growth. An analysis of variance (ANOVA) tested for grade-level effects and found that growth differed significantly between grades. Grade 3 students reportedly gained 1.0 correctly read words per minute (cwpm) each week. This gain rate decreased to 0.85 cwpm per week in grade 4 and 0.50 words per week in grade 5. Consistent with reading theory, the authors noted an inverse relation between slope and grade for ORF, indicating that students made their most dramatic growth in the early grades.

Eight years later, Deno, et al. (2001) used CBM to compare the academic growth rates of students with learning disabilities to those of students in general education. The sample was composed of four local education agencies (LEAs) across the U.S., three of which administered ORF CBM benchmarks in fall, winter, and spring, while the fourth administered them weekly.[1] Regarding the reliability of the measures, the authors assumed what most do in this line of research, that "CBM passage difficulty remained constant across the school year so that changes in scores reflected changes in reading proficiency rather than changes in passage difficulty" (p. 512). Based on the conclusions of Fuchs et al. (1993), Deno et al. (2001) assumed within-year growth to be linear and used OLS to estimate weekly fluency growth rates of 1.18 cwpm per week for grade 3, 1.10 for grade 4, and 0.58 for grade 5. The authors also reported that weekly growth rates for students with learning disabilities were consistent at 0.58 across grades 3–5.

According to Deno et al. (2001) "it should be emphasized that growth on CBM tasks *within* any school year is satisfactorily represented by a linear model" (p. 519, emphasis in the original). However, since the Fuchs et al. (1993) study, the results of nearly every study examining typical academic growth within a school year as measured by CBM have suggested a non-linear model best represents growth in ORF (Ardoin & Christ, 2008; Christ, Silberglitt, Yeo, & Cormier, 2010; Graney et al., 2009; Hasbrouk & Tindal, 2006).

Hasbrouk and Tindal (2006) collected ORF information that spanned 23 states and included samples ranging from 3,496 to 20,128 students per grade level. The report presented ORF norms for students in the fall, winter, and spring at the 10th, 25th, 50th, 75th, and 90th percentiles for each of grades 1 through 8 (excepting grade one fall). Although within-year individual growth patterns were not explicitly examined, the authors reported norms that demonstrated greater growth rates from fall to winter than from winter to spring for students at the 50th percentile in each of grades two through eight (with the exception being grade seven where the growth in the two periods was nearly identical).

A year later, Silberglitt and Hintze (2007) conducted a study that examined the ORF growth rates of students based on deciles of achievement. Using students whose scores were between the 50–59th decile as the reference group, the authors found that the top decile and the bottom two deciles all had slopes of improvement

[1] Given that the authors had a database with growth measured both weekly and with tri-annual benchmarks, it would have been interesting to compare the estimates between these samples; unfortunately, no such analysis was completed.

significantly lower than the reference group. The discrepancy between growth rates of different student groups is consistent with the literature on this topic, showing that students in special education (i.e., the bottom two deciles) have lower growth rates than students in general education (Deno et al., 2001; Graney, Missall, Martinez, & Bergstrom, 2009). Similar to Deno et al. (2001) and Fuchs et al. (1993), the authors operated under the assumption of within-year linear growth. However, they used a more advanced statistical technique—hierarchical linear modeling (HLM), which can account for the nested nature of ORF data (e.g., repeated measures within students, students within classrooms, or classrooms within schools) and thereby produces more accurate standard errors than does OLS.

In their study, Ardoin and Christ (2008) used "a standard set of descriptive statistics and correlations" to examine the growth rate of 86 students in second grade on CBM benchmark assessments (p. 115). Using a multivariate analysis of variance (MANOVA), they found the academic gains of students from the fall to winter benchmarks were significantly greater than the gains from winter to spring. Although the sample size was small and results cannot be interpreted as the standard, they do provide additional evidence that within-year growth should be further investigated and not assumed to be linear.

Graney et al. (2009) examined within-year growth of students in grades 3–5 over a two-year period, with 442 students across all grades in year one and 456 students across grades in year two. Because the study took place in one elementary school, the authors could better control the testing dates and *precisely* calculate weekly growth rates; similar studies often must divide each gain score by an *approximation* of the number of weeks between testing administrations. Similar to Ardoin and Christ (2008), however, the small sample limited the external validity of the study. While most research has found greater growth from fall to winter than from winter to spring, Graney et al. (2009) found statistically greater growth from winter to spring than from fall to winter. It is important to note that not only did the resulting growth estimates directly conflict with all other reported studies, but so did the results that showed no differences in growth between grades. It is commonly demonstrated in reading fluency research and theory that greater growth occurs in the earlier grades than in later (Fuchs et al., 1993). These discrepant results, however, could be attributed to a confounding intervention program within the school that promoted data use and evidence-based interventions. That is, if the school was engaged in a concerted and collaborative effort to use CBM data to intervene quickly and effectively, it is feasible that student growth patterns could have been positively changed.

Using the same AIMSweb CBM system as Graney et al. (2009), Christ, Silberglitt, Yeo, and Cormier (2010) evaluated the rate of ORF growth with 3808 students, grades two through six, in five school districts. The authors concluded that ORF growth was better estimated with a piecewise model than with a linear model, that more growth occurred in the fall than in spring, and that more growth was observed in earlier grades. Similar to the results of Deno et al. (2001) and Graney et al. (2009), the authors also found that annual growth was more substantial within the general education population, and that the seasonal effect was less pronounced within the special education population. Results showed mean weekly fall fluency cwpm growth rates of 1.38, 1.19, and 1.12, and weekly spring fluency growth rates of 0.97, 0.78, and 0.83 for grades 3, 4, and 5, respectively.

The research behind the annual growth of benchmark reading fluency is sparse, but it appears that a linear model may not be the best for describing ORF growth. Most studies have found or provided evidence for faster growth from fall to winter than from winter to spring, for faster growth among general education than special education, and for faster growth for lower grades than higher grades. Recent findings are adding to the understanding that growth may not be linear, but a conclusive pattern is not yet evident. Based on the findings of these prior CBM ORF studies, we hypothesize that a non-linear growth model will better fit the within-year growth for students in grades 3–5 but are not in the position to make causal or correlative speculations about why growth is not linear.

Perhaps the three most recognized CBM systems, easyCBM, DIBELS, and AIMSweb, offer three yearly ORF benchmarks which are intended to be administered to all students in a school. Schools use their students' results on these benchmark assessments and/or the results presented by the system developers to establish risk indicators. For example, the easyCBM system allows schools to create their own cut-points for student identification, while DIBELS provides cut-points to identify students as 'at risk,' 'some risk,' or 'low risk.' DIBELS cut-points were based on a bootstrapping technique using conditional probabilities of success on a criterion (e.g., other DIBELS measures), where 'at risk' decisions resulted in a low probability of future reading proficiency (approximately 0.15, ±0.05), 'some risk' decisions resulted in an even probability of future reading proficiency (approximately 0.50, ±0.05), and 'low risk' decisions resulted in high probability of future reading proficiency (approximately 0.85, ±0.05) (Cummings et al. 2008). While easyCBM and DIBELS do not offer explicit guidance for expected growth rates between benchmark assessments,[2] as noted earlier AIMSweb reports national *weekly* growth norms that are based on the assumption of linearity. In short, until recently the underlying assumption of research of linear growth has made its way into practice, yet more recent research indicates this assumption may be faulty. As a result, schools may be holding students to inaccurate expectations for growth. If some periods typically witness faster growth then use of a constant benchmark rate could lead to inadequate provisions of support to students at risk, while in other periods where growth is typically slower then use of a constant benchmark rate could lead to overuse of intensive instructional supports for students at risk.

In response to these issues and the mixed research results to date, our study examines one district's data to directly compare linear and discontinuous models of ORF benchmark growth, and opens the discussion about the potential and limitations of using a quadratic growth model. This study contributes to the slowly building evidence on within year ORF growth. In addition, while previous studies have largely used an OLS approach to modeling growth, the current study uses HLM to control for the nesting of repeated measures within students to produce

[2] The DIBELS website offers a draft paper on progress monitoring practices that implies an expectation for linear growth through the use of an "aimline" that is converted into a weekly growth rate goal on sample figures (Kaminski, Cummings, Powell-Smith, & Good, 2007).

more robust estimates of growth rates, as did Silberglitt and Hintze (2007) and Christ et al. (2010). Finally, most studies have not examined how growth differs for students with different background characteristics beyond educational setting, or general and special education. The current study addresses this issue by examining whether student characteristics are predictive of differences in growth.

More knowledge of reading fluency growth patterns can increase the accuracy of the decisions made by teachers and also inform the methodology of researchers' evaluations of educational programs. In addition, more accurate measures of change over time are needed to monitor student growth and ensure adequate learning progress as well as intervention effectiveness. The goal of this study was to answer the following research questions.

1. Does a non-linear growth model (i.e., discontinuous) fit grade 3–5 data on within-year CBM oral reading fluency measures better than a linear growth model?
2. Using the most appropriate growth model identified by the first research question, how do CBM oral reading fluency growth slopes differ between students based on sex, free/reduced lunch status (FRL), ethnicity, special education services (SpEd), and Limited English Proficiency services (LEP)?

Methods

Participants and setting

This study took place in a school district in a Pacific Northwest state with 2,465 students in Grades 3, 4, and 5 (Table 2). All demographic data was taken from school records. Of the total sample, 48% were female, and 62% qualified for free or reduced-price lunch (FRL), 19% received special education services, and 6% were receiving Limited English Proficiency (LEP) services.

The ethnic make-up of the sample was as follows: 71% White, 14% Latino, 8% multi-ethnic, less than 3% American Indian/Alaskan Native, less than 2% Asian/Pacific Islander, less 2% was Black, and 1% declined to report. Students who declined to report their ethnicity were deleted from the analyses. Because of the lack of ethnic diversity in the sample, the ethnicity variable was collapsed into two categories: ethnic non-minority students, which made up 71% of the sample and included only White students[3]; and ethnic minority students, which included American Indian/Alaskan Native, Asian, Black, Latino, and multi-ethnic students.

Measures

The dependent variable in the study was student scores on the easyCBM Passage Reading Fluency Benchmark measures for grades 3–5 (Alonzo, Tindal, Ulmer, &

[3] An examination of the data revealed that it would not be appropriate to group Asian/Pacific Islander students with White students.

Table 2 Descriptive statistics

Variables	Mean (SD)
Level-1	$n = 5{,}187$
Passage reading fluency	124.44 (45.26)
Level-2	$n = 2{,}465$
Grade 3	0.31 (0.46)
Grade 5	0.34 (0.47)
Female	0.48 (0.50)
FRL status	0.62 (0.49)
Ethnic minority	0.28 (0.45)
Special education status	0.19 (0.39)
LEP	0.06 (0.24)

Level-2 means represent proportion of students with a particular characteristic

Glasgow, 2006). The easyCBM ORF measures were collected in a multiple-time-point design administered during the fall, winter, and spring of the 2008–2009 school year. The passages used in these ORF measures were written specifically as part of the easyCBM progress monitoring and assessment system (Alonzo & Tindal, 2007a). A trained assessor administered the tests to the students individually. Students were shown a 250-word narrative passage and were given 60 seconds to orally read as much of the passage as they could. The trained assessor followed along as the student read, indicating on the test protocol each word the student read incorrectly. If a student hesitated for more than three seconds, the assessor would provide the correct word, prompt the student to continue and mark the word as read incorrectly. Student self-corrections were marked as correct responses. At the end of the allotted time, the assessor marked the last word read and calculated the total number of words read correctly to arrive at the student's score, in words read correctly per minute.

This study used a single different reading passage for each benchmark observation. The ORF benchmark measures used in this study were developed to be of equivalent difficulty for each grade level (for description of methods and results see Alonzo & Tindal, 2007b). Reading passages were written according to word count and grade-level guidelines and reviewed by the lead researcher as well as an experienced teacher. Data on passage difficulty, such as the Flesch-Kincaid readability estimates, were used to bring the passages into closer alignment; correlations and mean differences between the different forms of the measures were analyzed using a repeated measures ANOVA analysis and these results were used to obtain information on the relative difficulty of each passage. Passages that were most similar in difficulty were identified, and the difficulty levels of the remaining passages were either increased (based on more average words read correctly than the other passages included in the pilot testing) or decreased (based on fewer average words read correctly than the other passages included in the pilot testing). To improve the reliability of score interpretations, all alternate forms (20) for each grade level were administered to the same group of students over the course of one week, and each student was administered all ORF forms (20) by the same tester to reduce variability caused by different testers. Alternate-form reliability coefficients

Table 3 Oral reading fluency (ORF) benchmark means (standard deviations) by grade and season

Grade	Fall Mean (SD)	Winter Mean (SD)	Spring Mean (SD)
3	74.48 (35.41)	106.90 (39.94)	107.48 (39.24)
4	102.31 (36.58)	122.89 (38.43)	130.91 (41.49)
5	134.48 (42.81)	143.38 (39.72)	156.01 (40.18)

Table 4 Correlations of student-level variables

	Female	FRL	Ethnic minority	SpEd
Female	–			
FRL	.01	–		
Ethnic minority	−.02	.25	–	
SpEd	−.16	.11	−.03	–
LEP	−.00	.19	.41	.05

Correlations equal to and above absolute 0.11 are significantly different at the 0.01 level. All variables were dichotomous, with female, FRL, ethnic minority, special education, and LEP students as the reference groups

during field testing of ORF passages ranged between 0.88 and 0.96 (Alonzo, Mariano, Nese, & Tindal, 2010). The means and standard deviations for the ORF passages used in our study are presented in Table 3 by grade and season.

The following student demographic covariates were obtained from school records and included in the student level of the model: sex, ethnicity, FRL, SpEd, and LEP. Table 4 displays the correlations of the student-level variables.

Analyses

HLM (Raudenbush & Bryk, 2002) offers a robust method for studying individual change and accounting for between-student differences. Stage (2001) outlined some statistical advantages of HLM that are not available with other statistical techniques. These advantages include the potential for improved accuracy in slope estimation; the flexibility to test the effects of independent variables on the initial status of the slope and the change in slope over time; and a statistical test for both group effects and individual variation in growth. Another important advantage is that HLM, using maximum likelihood, can include all students who have been observed at least once, and results can be interpreted as if no missing data were present as long as the data are missing at random (Raudenbush & Bryk, 2002). In this consideration, HLM has an advantage over the multivariate repeated-measures analyses that are most often used in the CBM growth studies in which the number and spacing of time points must be invariant across students (Raudenbush & Bryk, 2002). Even with the use of HLM, however, an appropriate growth model is needed because the accuracy of the growth parameters may still be compromised if, for example, a linear model is used to estimate observed within-year growth that is non-linear.

A multilevel data analysis method, HLM (Raudenbush & Bryk, 2002), was used to analyze the data. The outcome, Y_{ti}, was defined as student ORF rate, defined as words read correctly at time t for student i. The ORF benchmark measures were administered to the students during the fall, winter, and spring within one school year. Because two different growth models were explored, the time variable was measured in different units for each analysis. For the linear growth model time was constructed in *months*, so that the fall data point (September), or the initial status was coded zero, winter (January), was coded four, and spring (May), was coded eight. For the discontinuous growth model, vectors were created to indicate the change from *fall-winter* and the change from *winter-spring*. Across grades, the correlation between ORF scores and time, in months, was 0.22.

A two-level hierarchical model represented student reading fluency growth with intra-individual growth trajectory represented at level-1 and variation in these trajectories across students represented at level-2.

In all models discussed below, dummy vectors representing grade level were modeled at every level-2 term (i.e., π_{0i}, π_{1i}, and π_{2i}), regardless of whether the error parameter was fixed or random. Students in grade 4 were the reference group, and dummy vectors were created to represent students in grade 3 and grade 5. These terms allowed for modeling ORF variance across grade levels.

Level-1 model comparison

The analyses began with a null model (i.e., no variables entered into the model at any level). Next, two different growth models were constructed separately to determine a preferred model: (1) a linear growth model, and (2) a curvilinear growth model. In these two growth models, the discrete time variables were modeled at level-1 to help determine the best model of growth and to gather baseline statistics for evaluating subsequent models.

Because there were only three testing observations, only two random parameters could be estimated without over-saturating the model. Given the grade-level spread of the sample, the intercept was always allowed to vary randomly, meaning that only one other random effect could be simultaneously modeled. All terms of the linear growth model could be explored simultaneously. In contrast, the fall-winter and the winter-spring terms could not both be estimated in the same discontinuous growth model, and thus random effects were alternately freed and fixed in order to determine the optimal specifications of each type of growth model.

Linear growth. First, as oral reading fluency measures are collected tri-annually, a linear growth model was employed at level-1 to represent linear growth across the three occasions, with time parameterized as the month of testing.

$$Y_{ti} = \pi_{0i} + \pi_{1i}\text{month}_{ti} + e_{ti} \quad (1)$$

$$\pi_{0i} = \beta_{00} + \beta_{01}\text{Third}_i + \beta_{02}\text{Fifth}_i + r_{0i} \quad (2)$$

$$\pi_{1i} = \beta_{10} + \beta_{11}\text{Third}_i + \beta_{12}\text{Fifth}_i(+r_{1i}) \quad (3)$$

In this model, π_{0i} represented the initial ORF status performance for student i at time zero (i.e., fall of the 2008-2009 school year); π_{1i} represented the linear growth rate for student i over the school year; month_{ti} represented time in months; and e_{ti} represented the deviation at time t of student i's ORF score from the predicted score based on the model. Please note that the parentheses included in the equations presented throughout are meant to indicate which parameters were tested as both random and fixed.

In the first linear model run, only the intercept (π_{0i}) included a random coefficient (r_{0i}), and in the second linear model run, both the intercept (π_{0i}) and linear growth parameter (π_{1i}) included a random coefficient (r_{0i} and r_{1i} respectively).

Discontinuous growth. Next, a discontinuous growth model was employed at level-1 to represent discontinuous growth across the three testing occasions, wherein *month,* representing month of testing, was replaced by dummy vectors indicating the change from fall to winter and the change from winter to spring, such that winter_{ti} was coded zero for the fall observation only and one for the winter and spring observations and spring_{ti} was coded zero for the fall and winter observations and one only for the spring observation.

$$Y_{ti} = \pi_{0i} + \pi_{1i}\text{winter}_{ti} + \pi_{2i}\text{spring}_{ti} + e_{ti} \tag{4}$$

$$\pi_{0i} = \beta_{00} + \beta_{01}\text{Third}_i + \beta_{02}\text{Fifth}_i + r_{0i} \tag{5}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}\text{Third}_i + \beta_{12}\text{Fifth}_i(+r_{1i}) \tag{6}$$

$$\pi_{2i} = \beta_{20} + \beta_{21}\text{Third}_i + \beta_{22}\text{Fifth}_i(+r_{2i}) \tag{7}$$

In this model, π_{1i} represented the change in ORF scores from fall to winter; and π_{2i} represented the additional change in ORF scores from winter to spring.

Two discontinuous models were compared, in which the first discontinuous model specified the intercept (π_{0i}) and the fall-winter change parameter (π_{1i}) as random coefficients, and the second specified the intercept (π_{0i}) and the winter-spring change parameter (π_{2i}) as random.

Fit indicators were examined to compare the growth models and their respective specifications; these included graphical comparisons of the observed and estimated data; variance explained in the criterion by the level-1 model, where larger values indicate better fit; and the deviance statistic, Akaike's information criterion (AIC), and Bayesian information criterion (BIC), where models with smaller values indicate better fit (Singer & Willett, 2003). Note that because the final growth models were not strictly nested and had the same number of estimated parameters, there were no degrees of freedom with which to conduct a statistical test of significance using the deviance statistics and thus deviance, AIC, and BIC statistics had to be weighed using a certain degree of subjective judgment (see Singer & Willett, 2003, pp. 116–122).[4] The final model selected served as the baseline model to which additional predictors of differences in growth between students were added.

[4] Singer and Willett (2003) describe comparing AIC and BIC statistics as "an 'art based on science'" (p. 122). It is not uncommon for different indices to suggest different models as better fitting the data. Thus, analysts must appeal to theoretical and practical concerns as well as tests of statistical significance.

Modeling between-student growth differences

After identifying an appropriate level-1 baseline model, student variables were added at level-2 to explain differences between students in the intercept and growth parameters. Student sex, FRL status, ethnicity, SPED qualification, and LEP qualification were added simultaneously as predictors of all level-1 coefficients.

Each of the student characteristics was coded dichotomously to represent the mean differences between females and males (reference group), between those receiving free/reduced lunch and those not (reference group), between ethnic minority students and ethnic non-minority students (reference group), between special education students and general education students (reference), and between LEP students and non-LEP students (reference). As a result, the intercept was constrained to be the expected fluency for a grade 4, non-ethnic minority, non-LEP male, who is not receiving free or reduced price lunch or in special education.

Results

Results of the comparison between the different growth models are presented first, followed by the results explaining the between-student differences in growth.

Growth model comparisons

The results of each growth model type are followed by the results across the model types. Results of the final, or baseline, model used for describing growth at level-1 are then presented.

Within-type model comparisons

Table 5 shows the fixed effects, random effects, and fit indicators of the level-1 model comparison analyses. The results of the linear models indicated that the second model with the random linear growth slope was preferred to the first model. The chi-square test of reduction in deviance indicated the second model fit significantly better than the first model given the added parameters ($\chi^2 = 37.18$, $df = 2$). In addition, the second model explained a bit more variance in ORF and had lower AIC and BIC values, additional signs that the more specified model was the better model. However, the chi-square test for the variance component of the linear growth rate parameter, π_{1i}, was not significant ($\chi^2 = 1722.259$, $df = 1{,}717$), suggesting that there were no significant individual differences among students' linear ORF growth rates.

The results of the discontinuous models yielded mixed results. Here, the first model (i.e., with the random fall-winter change slope) had lower deviance, AIC, and BIC values suggesting it fit the data better; however, the second model (i.e., with the random winter-spring change slope) explained 3% more of the level-1 variance in ORF. In addition, the chi-square test for the variance component of the fall-winter change parameter, π_{1i}, in the first model was not significant ($\chi^2 = 1076.78$,

Table 5 Level-1 model comparison of the parameters and fit indicators of the linear and discontinuous growth models

	Null model	Linear model[a]		Discontinuous model[b]	
		π_{0i} variance	π_{0i} & π_{1i} variance	π_{0i} & π_{1i} variance	π_{0i} & π_{2i} variance
Fixed effects					
Intercept, γ_{00}	125.36*	102.92*	103.11*	99.00*	98.71*
	(0.87)	(1.32)	(1.31)	(1.35)	(1.36)
Grade 3, γ_{01}		−21.61*	−21.74*	−24.81*	−24.52*
		(1.93)	(1.92)	(1.90)	(1.90)
Grade 5, γ_{02}		28.11*	28.12*	33.51*	33.67*
		(2.04)	(2.04)	(2.11)	(2.10)
1st Growth parameter, γ_{10}	–	3.66*	3.63*	23.77*	24.07*
		(0.12)	(0.11)	(0.86)	(0.87)
Grade 3, γ_{11}	–	−0.02	0.00	7.62*	7.40*
		(0.17)	(0.16)	(1.20)	(1.23)
Grade 5, γ_{12}	–	−0.63*	−0.63*	−14.89*	−14.93*
		(0.17)	(0.17)	(1.29)	(1.29)
2nd Growth parameter, γ_{20}	–	–	–	7.83*	7.81*
				(0.74)	(0.73)
Grade 3, γ_{21}	–	–	–	−5.95*	−6.01*
				(1.04)	(1.04)
Grade 5, γ_{22}	–	–	–	6.55*	6.44*
				(1.06)	(1.06)
Random effects					
σ^2	367.82	199.82	195.18	161.71	151.73
τ, b_0	1663.36*	1392.77*	1221.67*	1211.49*	1333.36*
τ, b_1	–	–	0.28	16.48	–
τ, b_2	–	–	–	–	37.68*
Fit indicators					
Deviance statistic	50973.52	48770.79	48733.61	48204.87	48217.21
Level-1 pseudo-R^2	–	0.46	0.47	0.56	0.59
Deviance statistic	50973.52	48770.79	48733.61	48204.87	48217.21
AIC	50977.52	48774.79	48741.61	48212.87	48225.21
BIC	50989.14	48786.41	48763.40	48232.56	48247.00

Standard errors are in parentheses

* $p < 0.001$

[a] Where γ_{10} and γ_{20} represented linear parameters

[b] Where γ_{10} and γ_{20} represented fall-winter change and winter-spring change parameters, respectively

df = 1,012), but the winter-spring change parameter, π_{2i}, in the second model was significant ($\chi^2 = 1981.12$, *df* = 1,712). Because the chi-square test of reduction in deviance was unavailable here, we appealed to other indicators of fit. Given that there were significant differences between students in winter-spring change in ORF but not between students in fall-winter change, parsimony would argue for dropping the random effect that was not statistically significant. In addition, the second model reduced variance at level-1 more effectively, and thus the second model was chosen as the better fitting model.

Across-type model comparisons

Comparing the linear growth to the discontinuous growth model, the discontinuous model produced markedly lower deviance, AIC, and BIC values than did the linear model, clearly ruling out the linear model as the best fit for the data.

To make a clearer determination of the best fitting model, observed and predicted values for the data were compared graphically for each of the three grades studied (see Figs. 1, 2, and 3). The best-fitting model for each of the model types—the second linear model and the second discontinuous model—was used. For each grade level, the graphical representation of the data confirms that the linear growth model does not fit the data nearly as well as the discontinuous model. In answer to our first research question, the discontinuous growth model fit the grade 3–5 within-year CBM oral reading fluency measures better than a linear growth model. The second discontinuous model that specified the intercept (π_{0i}) and the winter-spring change parameter (π_{2i}) as random was used to model between-student growth differences.[5]

Comparing the discontinuous model to a curvilinear model

As an additional analysis, we also modeled ORF growth using a curvilinear model, and compared the results to those of the linear and discontinuous model. The quadratic growth model was employed at level-1 to represent curvilinear growth across the three occasions, with *month* representing the month of testing, and an additional term, month_{ti}^2, representing quadratic time, or months-squared (see Appendix 1 for curvilinear HLM equations). Similar to the discontinuous model, the linear and quadratic terms of the curvilinear growth model could not both be estimated in the same model so two quadratic models were compared.[6] The results of the two curvilinear models yielded somewhat mixed results, but the first model (i.e., with random linear growth slope and fixed quadratic slope) was determined to be the preferred model. Because the models had the same number of parameters, a significance test could not be conducted comparing the deviance statistics, but the first model had lower deviance, AIC, and BIC values. Both models explained about 60% of the level-1 variance in ORF.

[5] The same analyses were conducted separately for each grade level (3, 4, and 5), and using the same fit indicators, the discontinuous growth model in which the winter-spring change parameter varied was chosen as the preferred model for each grade level independently.

[6] In the first quadratic model, the intercept term (π_{0i}) and the linear growth rate parameter (π_{1i}) included random coefficients and the quadratic growth parameter (π_{2i}) was fixed. In the second quadratic model, the intercept term (π_{0i}) and the quadratic growth rate parameter (π_{2i}) included random coefficients and the linear growth parameter (π_{1i}) was fixed.

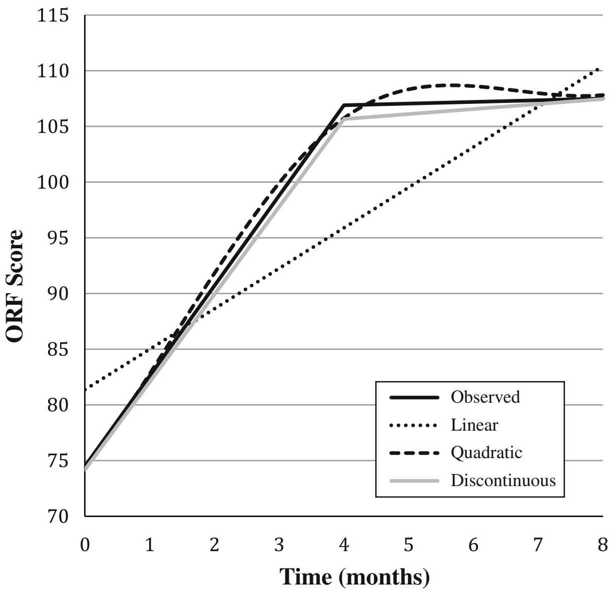


Fig. 1 Predicted trajectories for three parameterizations of growth compared to observed growth in grade 3

We found the values of the curvilinear model to be quite similar to those of the discontinuous model (Tables 5 and 7). In addition, Figs. 1, 2, 3 show the relative similarity between the fixed effects of the two models. Although the predictions from these two different models result in near-identical predictions at each time point, their contrasting underlying assumptions are made most readily apparent in Fig. 1. Specifically, the discontinuous model assumes linear growth between time points, while the quadratic model of course does not. The similarly good fit of the discontinuous and quadratic models suggests that with more time points, CBM ORF benchmark growth may best represented as curvilinear, but we offer caveats to this speculation in the discussion.

Baseline model for describing growth at level-1

The results of the discontinuous model—with the intercept (π_{0i}) and the winter-spring change parameter (π_{2i}) specified as random—are described here. This final model revealed that mean ORF rates differed by grade level, and also that change in ORF over time differed between grades. Grade 3 students demonstrated rapid growth early in the year and little to no growth later in the year. Their average reading fluency score at fall was about 74 cwpm and their winter ORF scores increased to about 106 cwpm, but their average spring ORF scores barely changed, staying at about 108 cwpm. In contrast, among grade 4 students, growth was less rapid initially and more rapid later. Their average reading fluency scores at fall was estimated at about 99 cwpm, the average winter ORF score increased to about 123 cwpm, and the average spring ORF score increased again to about 131 cwpm.

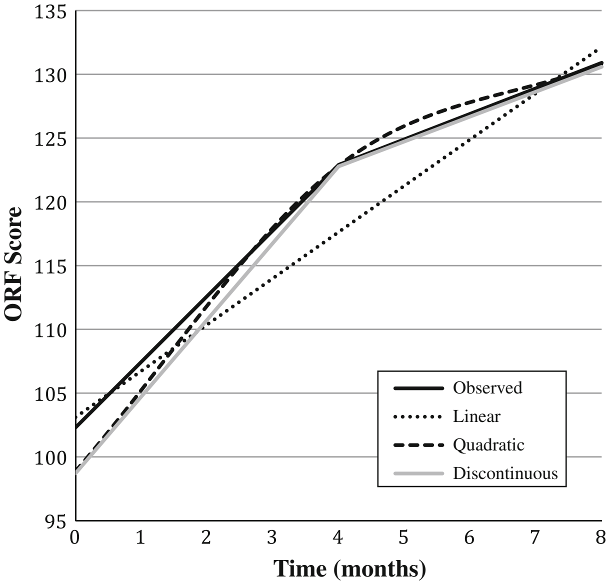


Fig. 2 Predicted trajectories for three parameterizations of growth compared to observed growth in grade 4

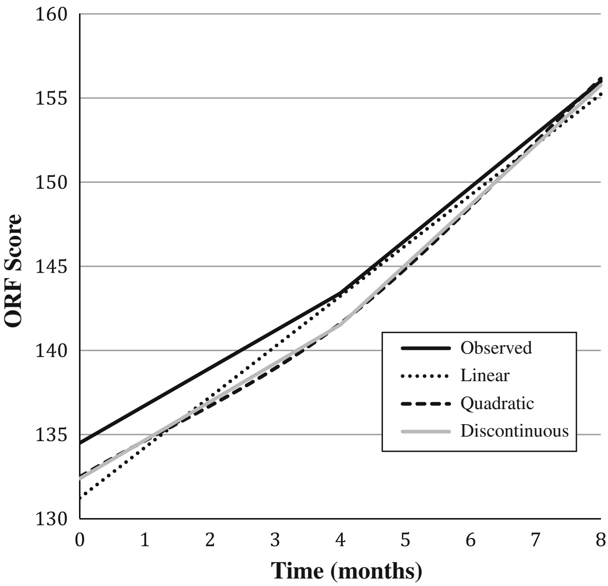


Fig. 3 Predicted trajectories for three parameterizations of growth compared to observed growth in grade 5

Finally, for grade 5 students the pattern of faster growth earlier in the year is reversed, such that their ORF increases less initially than it does later in the year. The average reading fluency score at fall for grade 5 students was estimated at about 132 cwpm, the average winter ORF score increased to about 142 cwpm, while the average spring ORF score increased to about 156 cwpm. In short, on average, ORF scores for students in grades 3 and 4 increased more sharply from fall to winter than from winter to spring, whereas for grade 5 students, ORF scores increased at a more even rate over the academic year.

Explaining between-student differences in growth

Table 6 displays the summary of student-level predictors of differences in reading fluency growth. The mean fall reading fluency score was about 88, 111, and 145 cwpm for students in grades 3, 4, and 5 respectively, who were male, not receiving free or reduced price lunch, not in special education, and not receiving LEP services. Across grade levels, females began the year reading about four more cwpm than males (Coefficient = 3.99, $SE = 1.50$, $p < 0.001$); students eligible for FRL began the year reading about 11 cwpm less than students not eligible for FRL (Coefficient = −11.39, $SE = 1.62$, $p < 0.001$); students receiving special education services began the year reading about 37 cwpm less than general education students (Coefficient = −36.53, $SE = 2.08$, $p < 0.001$); and students receiving LEP services began the year reading about 20 cwpm less than non-LEP students (Coefficient = −20.05, $SE = 3.17$, $p < 0.001$). After controlling for these student characteristics, there were no significant differences in fall oral reading fluency scores for ethnic minority and ethnic non-minority students.

The mean winter reading fluency score, after controlling for student characteristics, was about 120, 136, and 154 cwpm for students in grades 3, 4, and 5, respectively. The only significant slope predictor indicated that students receiving special education services demonstrated slower growth from fall to winter by about 4 cwpm than general education students (Coefficient = −3.96, $SE = 1.32$, $p < 0.01$).

The mean spring reading fluency score, after controlling for student characteristics, was about 122, 144, and 169 cwpm for students in grades 3, 4, and 5, respectively. No student-level demographic variables significantly explained differences in the winter to spring change in ORF scores.

In addition to reading growth estimates, we examined the variance in ORF explained by our model. Our results indicated that the winter-spring error variance parameter significantly varied between students, but no student characteristic variable significantly explained this variance. Although the level-2 student variables explained 24% of the overall variance in ORF scores between students, the final model indicated significant variance remained in both the intercept and the winter-spring slope parameters.

As a final note, we examined interaction effects between the covariates (sex, FRL, ethnic minority status, SpEd status, and LEP status) and grade-level to determine whether there was an effect of student characteristics between grades at different points in time. This analysis led to somewhat different results, but we

Table 6 Final model of growth in reading fluency

Fixed effects	Coefficient	SE	*df*	*p* value
Intercept, γ_{00}	111.45	1.78	2,457	0.000
Grade 3, γ_{01}	−23.32	1.75	2,457	0.000
Grade 5, γ_{02}	33.39	1.88	2,457	0.000
Female, γ_{03}	3.99	1.50	2,457	0.008
FRL, γ_{04}	−11.39	1.62	2,457	0.000
Ethnic minority, γ_{05}	0.15	1.88	2,457	0.937
SpEd, γ_{06}	−36.53	2.08	2,457	0.000
LEP, γ_{07}	−20.05	3.17	2,457	0.000
Fall-Winter change, γ_{10}	24.45	1.26	5,163	0.000
Grade 3, γ_{11}	7.53	1.21	5,163	0.000
Grade 5, γ_{12}	−14.89	1.29	5,163	0.000
Female, γ_{13}	−0.37	1.03	5,163	0.717
FRL, γ_{14}	−0.13	1.10	5,163	0.903
Ethnic minority, γ_{15}	2.67	1.39	5,163	0.054
SpEd, γ_{16}	−3.96	1.32	5,163	0.003
LEP, γ_{17}	−1.15	2.29	5,163	0.616
Winter-Spring change, γ_{20}	8.18	1.00	2,457	0.000
Grade 3, γ_{21}	−6.02	1.04	2,457	0.000
Grade 5, γ_{22}	6.43	1.05	2,457	0.000
Female, γ_{23}	−0.28	0.85	2,457	0.742
FRL, γ_{24}	1.27	0.93	2,457	0.172
Ethnic minority, γ_{25}	−2.04	1.13	2,457	0.072
SpEd, γ_{26}	−1.86	1.00	2,457	0.063
LEP, γ_{27}	−0.30	2.10	2,457	0.888

Random effects	Variance	*df*	Chi-square	*p*-value
Intercept, r_{0i}	1000.14	1707	19212.78	0.000
Winter-Spring slope, r_{2i}	39.37	1707	1981.60	0.000
Level-1, e_{ti}	150.30	–	–	–

deemed them of questionable relevance given the small sample sizes within grade for smaller demographic groups. Thus, we chose to be more conservative and report the main effects model. A description of this interaction model can be found in Appendix 2.

Discussion

The purposes of the study were to examine whether a linear or discontinuous growth model best fit the data on ORF benchmark measures for students in grades 3–5, and to examine growth rate variance between students. Contrary to common assumptions in the field based on early works (Deno et al., 2001; Fuchs et al., 1993), results here indicated that a linear growth model did not describe well change in students' oral reading fluency and that a discontinuous growth model was more appropriate.

Results suggest more growth occurs in the fall than in the spring, at least for students in grades 3 and 4, and more growth occurs for students in earlier grades. More growth in the fall than in the spring is consistent with prior research that has reported monotonic but decreasing gains in ORF across the school year (Ardoin & Christ 2008; Christ, et al., 2010; Good & Kaminski, 2003; Hasbrouck & Tindal, 2006). More growth in the earlier grades than in the later grades is also consistent with prior research (Ardoin & Christ, 2008; Christ et al., 2010; Deno et al., 2001; Fuchs et al., 1993; Good & Kaminski, 2003; Hasbrouck & Tindal, 2006; Pearson, 2008) and developmental reading theory, which indicates that greater growth on assessments directly testing decoding and fluency skills should be evident at earlier reading stages in the earlier grades. For grade 5 students, growth was actually faster in spring than in fall. It may be that reading fluency growth is greater overall for students in lower grades, with greater gains made in the fall. But as students progress in grade and skill, growth may become linear or the trend may even reverse with less growth in fall and more in spring for older students. More research is needed to understand the trend of reading skills growth across grades.

Across the three grades, males and students receiving FRL or LEP services began the school year with significantly lower reading fluency scores than their counterparts but had reading growth similar to their counterparts across the year. This would still mean, however, that these groups finished the year with lower reading fluency scores than did their counterparts. Although the gap does not widen, it does not narrow either.

Students receiving special education services began the year reading at significantly slower rates than general education students and had significantly less growth during the year; thus, they finished the year even further behind. Deno et al. (2001) and Christ et al. (2010) also found substantially lower growth rates for students with learning disabilities or receiving special education than for students in general education. Deno et al. (2001) reported weekly growth rates for students with learning disabilities at 0.58 across students in grades 3–5. Christ et al. (2010) estimated special education growth fall rates of 1.09, 0.99, and 0.90, and special education spring growth rates of 0.89, 0.67, and 0.67, for students in grades 3–5, respectively. Our weekly fall-winter growth rates[7] for students in special education were 1.56 for grade 3, 1.14 for grade 4, and 0.31 for grade 5, and our winter-spring growth rates were 0.02 for grade 3, 0.35 for grade 4, and 0.71 for grade 5. Implications are difficult to infer from the paper by Deno et al. (2001), particularly because not enough information is available about the regression models; however,

[7] We use the term "weekly growth rates" to refer to linear trends within the discontinuous model. Weekly growth scores were calculated by dividing the fixed effects for fall-winter and winter-spring coefficients by the number of weeks between benchmark assessments (approximately 18 weeks).

our results appear similar to those estimated by Christ et al. (2010). This is an important beginning to the effort to establish reading fluency growth estimates and expectations for student groups for the purpose of informing instructional practice and intervention.

The level-2 student variables explained about a quarter of the overall variance in ORF scores between students, but the final model indicated significant variance remained in both the intercept and the winter-spring slope parameters. Future modeling of similar data should at least report these values (Christ et al., 2010; Silberglitt & Hintze; 2007) but moreover should attempt to meaningfully explain this variance at the classroom and school level. We explored a three-level model with schools ($n = 20$) modeled at level-3, modeling the same parameters and interactions as the two-level model. Although the level three variance component was significant, we found the intraclass correlation (ICC) was 0.007, so that less than 1% of the variance in ORF scores could be attributed to between-school variance. In addition, the fixed effects were very similar for the two- and three-levels, with the largest difference between coefficients for the initial grade 4 ORF status (fall intercept, controlling for student characteristics) being 0.85 cwpm, and the difference between the other 53 fixed effect pairs ranging between −0.41 and 0.33 cwpm. There were no significant fixed effects in the three-level model that were not significant in the two-level model, and only one fixed effect that was significant in the two-level model that was not significant in the three-level model (the sex by grade 5 interaction for winter-spring fixed effects differed by 0.07, and p-values differed by 0.003). Thus, we presented the two-level model here for the sake of simplicity. It remains the charge of future research to attempt to meaningfully explain variance at the classroom and school level. Potential explanatory variables of interest include information about classroom instruction, the nature and degree of academic intervention, or information about school resources, as some variance may be attributable to the classroom and school variables rather than being attributable to the students themselves.

Our study adds to the sparse research on CBM reading growth rates, but more research needs to be done for students receiving various interventions and special education services, as observations of growth rates may play a key role in decision-making for interventions and special education evaluation referrals. Research of this kind should pay particular attention to the special education label of the student and the type and degree of instruction they receive.

Limitations

The primary limitation of this study is the reliability of the reading fluency measure for studying growth. This reliability can be affected by a number of things, including variability in passage difficulty within grade levels, or the administration practices of assessing reading fluency. To address issues of passage/form equivalency, the ORF measures used in this study were written according to word count and grade-level guidelines and designed to be of equivalent difficulty. Alternate-form reliability coefficients during field testing of ORF passages were above .88, suggesting that changes in scores reflected changes in reading proficiency rather than changes in passage difficulty. In addition, the pattern of reading fluency growth across grade levels aligns with research that shows decreasing within-year fluency growth across grade levels, and decreasing fluency growth between grade levels.

An additional limitation of this study is that the sample was geographically homogenous and somewhat ethnically homogenous, which limits the generalizability of the results. The findings of our study should not be extended to a heterogeneous population; rather they are qualified to a unique population, a quality which it shares with similar studies.

Implications

The results of this study have important implications for research and theory on reading fluency. The discontinuous model with the randomly varying winter-spring slope was chosen as the growth model that best fit the current data. An interesting finding, however, was the remarkable similarity of the curvilinear and discontinuous growth models in both fit indices and predicted values (and that both were more appropriate for the data than the linear model). This finding was related to the decreased ORF growth in winter-spring compared to fall-winter, which was most obvious for grade 3 students, but also marked for grade 4 students. In contrast, grade 5 students showed more nearly linear growth, but with slightly larger gains in fluency between winter and spring than between fall and winter. As is characteristic of so many studies of oral reading fluency, the current study only incorporated three time points. While this frequency of assessment is consistent with general formative assessment practice, it does limit the extent to which we can draw sound conclusions about the superiority of alternative theories of reading fluency development. Because we could not estimate all random parameters of the quadratic and discontinuous models, it may well be that the optimal model would include all random effects. Nevertheless, this research is important to schools precisely because it follows the framework of assessment used in practice rather than what is preferred by researchers and methodologists. Schools are currently using estimations of student risk-status and making high-stakes eligibility decisions based on the results of these formative assessments, and therefore need to know more about expected within-year growth. Ignoring the potential of non-linear growth models because of a limited number of available random parameters is a disservice to those practitioners who rely on accurate benchmarks for growth data. Based on our results, the assumption of constant within-year linear ORF growth across grade levels does not hold. Because we found uneven growth by season and grade, if we provide schools benchmarks based on the linear assumption (i.e., one "weekly growth" rate for schools to use day-to-day), then those benchmarks may not be as useful for judging adequate student progress.

Thus, the results of this and other studies (Ardoin & Christ, 2008; Christ et al., 2010; Hasbrouck & Tindal, 2006) suggest that researchers need to explore models beyond those representing linear growth. Although the discontinuous model was chosen over the linear model for our data, without research utilizing more than three observations within a school year, it may be impossible to make a definitive modeling decision. Future research that includes four or more benchmark assessment periods should be employed to investigate more definitively whether growth in reading rate is truly curvilinear or discontinuous in nature and begin to explore theoretical reasons behind the growth trajectory. If a quadratic model is determined to be the best to explain within-year reading growth, future research may also explore developmental or educational explanations of such growth trajectories.

These departures from linear growth hold important implications for reading instruction and intervention. They suggest that expecting constant growth rates across the school year is an unreasonable expectation, particularly for students in earlier grades. They also suggest that for students in grades 3 and 4 who do not demonstrate normative growth in the first half of the school year appropriate interventions may, in fact, enable them to catch up to their peers during the typically slower growth second half of the year. In contrast, grade 5 students who do not demonstrate expected progress during the first half of the year have twice as much ground to cover to catch up to their peers.

Practitioners working within an RTI framework can also benefit from these findings. Similar to recent research, our results indicated significantly different growth slopes between special education and general education students (Deno et al., 2001; Christ et al., 2010; Silberglitt & Hintze, 2007). Having more accurate expectations for fluency growth can help teachers design instruction and implement interventions. Realistic and ambitious goals can be set for low-achieving students, expectations for students' response to the reading intervention can be better understood, and growth can be meaningfully interpreted and qualified. In practice, however, the challenge remains getting classroom teachers to accurately interpret data to link instruction and intervention to the assessment results.

In addition, Deno et al. (2001) found that although the typical growth rate of students with learning disabilities was substantially lower than that of students in the general education sample, evidence from five previous studies with "demonstrably effective reading conditions" (p. 515) demonstrated that the growth rate of students in special education with effective teaching practices approached an equal growth rate to general education students. This suggests that a reasonably ambitious growth standard for special education students may be what is typical for general education students. The authors also noted that general education students receiving effective reading practices might also have exhibited a boost in their fluency growth. More research needs to be done examining the efficacy of reading interventions for special education students as well as for low achieving regular education students in an effort to increase students' within-year reading growth in comparison with the appropriate reference group expectations.

Conclusion

The current study contributes to our building understanding of how oral reading fluency changes over time and what individual characteristics contribute to differences in those changes. First and foremost, growth was not found to be linear. Whether future research more clearly demonstrates that growth is curvilinear or truly discontinuous (varying in slope from time point to time point), this study adds to the evidence that expectations for change in reading rates should not use a linear model. In addition, the current results indicate that grade level makes a great deal of difference in growth rates and in the shape of growth over time. Finally, individual student characteristics, such as gender or free lunch status, explain differences in how students begin the school year, but with the exception of special education status, not in how they improve over time. To better understand differences in students' reading rates and change in those rates, we may need to turn our attention to classroom and school characteristics.

Acknowledgments This research is supported in part by grant number R305B080004 from the Institute of Education Sciences, U.S. Department of Education. Opinions expressed do not necessarily reflect the opinions or policies of the U.S. Department of Education.

Appendix 1: Curvilinear growth model

$$Y_{ti} = \pi_{0i} + \pi_{1i}\text{month}_{ti} + \pi_{2i}\text{month}_{ti}^2 + e_{ti} \tag{8}$$

$$\pi_{0i} = \beta_{00} + \beta_{01}\text{Third}_i + \beta_{02}\text{Fifth}_i + r_{0i} \tag{9}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}\text{Third}_i + \beta_{12}\text{Fifth}_i(+r_{1i}) \tag{10}$$

$$\pi_{2i} = \beta_{20} + \beta_{21}\text{Third}_i + \beta_{22}\text{Fifth}_i(+r_{2i}) \tag{11}$$

For the curvilinear growth model, the linear and quadratic terms could not both be estimated in the same model so two quadratic models were compared. In the first quadratic model, the intercept term (π_{0i}) and the linear growth rate parameter (π_{1i}) included random coefficients, and the quadratic growth parameter (π_{2i}) was fixed. In the second quadratic model, the intercept term (π_{0i}) and the quadratic growth rate parameter (π_{2i}) included random coefficients, and the linear growth parameter (π_{1i}) was fixed. The results of the curvilinear models yielded somewhat mixed results, but the first model (i.e., with random linear growth slope and fixed quadratic slope) was determined to be the preferred model because it had lower deviance, AIC, and BIC values. Both curvilinear models explained about 60% of the level-1 variance in ORF (Table 7).

Table 7 Level-1 parameters and fit indicators of the curvilinear growth model

	Null model	Curvilinear model[a]	
		π_{0i} & π_{1i} variance	π_{0i} & π_{2i} variance
Fixed effects			
Intercept, γ_{00}	125.36*	98.87*	98.75*
	(0.87)	(1.35)	(1.35)
Grade 3, γ_{01}		−24.65*	−24.54*
		(1.90)	(1.90)
Grade 5, γ_{02}		33.61*	33.67*
		(2.10)	(2.10)
1st Growth parameter, γ_{10}	–	7.99*	8.04*
		(0.36)	(0.37)
Grade 3, γ_{11}	–	3.57*	3.53*
		(0.51)	(0.52)
Grade 5, γ_{12}	–	−6.39*	−6.40*
		(0.54)	(0.54)
2nd Growth parameter, γ_{20}	–	−0.50*	−0.51*
		(0.04)	(0.04)
Grade 3, γ_{21}	–	−0.42*	−0.42*
		(0.06)	(0.06)
Grade 5, γ_{22}	–	0.67*	0.67*
		(0.06)	(0.06)
Random effects			
σ^2	367.82	148.15	146.25
τ, b_0	1663.36*	1253.74*	1298.57*
τ, b_1	–	1.26*	–
τ, b_2	–	–	0.02*
Fit indicators			
Deviance statistic	50973.52	48220.62	48229.74
Level-1 pseudo-R^2	–	0.60	0.60
Deviance statistic	50973.52	48220.62	48229.74
AIC	50977.52	48228.62	48237.74
BIC	50989.14	48250.42	48259.54

Standard errors are in parentheses

* $p < 0.001$

[a] Where γ_{10} and γ_{20} represented quadratic parameters

Appendix 2: Interaction effects model

In grade 3, female and SpEd students have significantly higher and lower slopes, respectively, than male and non-SpEd students. The only other significant effect in grade 3 is that the slope from winter to spring for LEP students is clearly negative, while all the others are either positive or stable (Table 8).

Table 8 Model of growth in reading fluency including interaction terms

Fixed effects	Coefficient	SE	*df*	*p*-value
Intercept, γ_{00}	114.12	2.53	2,447	0.000
Grade 3, γ_{01}	−28.19	3.41	2,447	0.000
Grade 5, γ_{02}	31.77	3.67	2,447	0.000
Female, γ_{03}	−2.75	2.41	2,447	0.255
FRL, γ_{04}	−10.79	2.64	2,447	0.000
Ethnic minority, γ_{05}	3.02	3.06	2,447	0.323
SpEd, γ_{06}	−38.77	3.21	2,447	0.000
LEP, γ_{07}	−21.81	5.95	2,447	0.000
Sex-Grade 3	9.33	3.48	2,447	0.008
Sex-Grade 5	9.37	3.67	2,447	0.011
FRL-Grade 3	−3.92	3.79	2,447	0.301
FRL-Grade 5	−0.32	3.95	2,447	0.936
Ethnic minority-Grade 3	−0.33	4.36	2,447	0.940
Ethnic minority-Grade 5	−7.25	4.66	2,447	0.120
SpEd-Grade 3	13.81	4.57	2,447	0.003
SpEd-Grade 5	−6.42	5.14	2,447	0.212
LEP-Grade 3	6.80	7.78	2,447	0.382
LEP-Grade 5	0.13	7.96	2,447	0.987
Fall-Winter change, γ_{10}	23.04	1.79	5,133	0.000
Third, γ_{11}	12.67	2.37	5,133	0.000
Fifth, γ_{12}	−17.02	2.50	5,133	0.000
Female, γ_{13}	2.44	1.72	5,133	0.156
FRL, γ_{14}	2.77	1.88	5,133	0.141
Ethnic minority, γ_{15}	0.37	2.26	5,133	0.871
SpEd, γ_{16}	−8.32	2.06	5,133	0.000
LEP, γ_{17}	−4.70	2.94	5,133	0.109
Sex-Grade 3	−4.30	2.38	5,133	0.070
Sex-Grade 5	−2.26	2.56	5,133	0.378
FRL-Grade 3	−4.85	2.55	5,133	0.057
FRL-Grade 5	−2.51	2.69	5,133	0.351
Ethnic minority-Grade 3	0.35	3.16	5,133	0.913
Ethnic minority-Grade 5	6.03	3.39	5,133	0.075
SpEd-Grade 3	−3.58	2.77	5,133	0.196
SpEd-Grade 5	16.74	3.10	5,133	0.000
LEP-Grade 3	4.80	4.92	5,133	0.330
LEP-Grade 5	3.55	4.98	5,133	0.475
Winter-Spring change, γ_{20}	8.56	1.40	2,447	0.000
Third, γ_{21}	−7.12	2.03	2,447	0.001
Fifth, γ_{22}	6.19	1.88	2,447	0.001
Female, γ_{23}	2.05	1.44	2,447	0.156
FRL, γ_{24}	−0.57	1.56	2,447	0.716

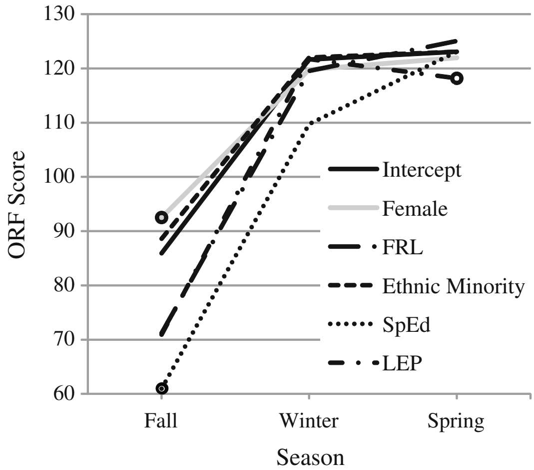
Table 8 continued

Fixed effects	Coefficient	SE	*df*	*p*-value
Ethnic minority, γ_{25}	−4.79	1.93	2,447	0.013
SpEd, γ_{26}	−2.53	1.69	2,447	0.136
LEP, γ_{27}	7.98	3.51	2,447	0.023
Sex-Grade 3	−3.15	2.06	2,447	0.127
Sex-Grade 5	−4.06	2.07	2,447	0.049
FRL-Grade 3	2.56	2.25	2,447	0.256
FRL-Grade 5	3.25	2.25	2,447	0.148
Ethnic minority-Grade 3	4.81	2.63	2,447	0.067
Ethnic minority-Grade 5	4.08	2.84	2,447	0.151
SpEd-Grade 3	2.53	2.41	2,447	0.296
SpEd-Grade 5	−0.04	2.44	2,447	0.988
LEP-Grade 3	−12.88	5.22	2,447	0.014
LEP-Grade 5	−11.57	4.82	2,447	0.017

Random effects	Variance	*df*	Chi-square	*p*-value
Intercept, r_{0i}	1001.66	1697	20030.50	0.000
Winter-Spring slope, r_{2i}	47.91	1697	2042.65	0.000
Level-1, e_{ti}	143.38	–	–	–

In grade 4, SpEd, LEP and FRL students had lower intercepts than non-SpEd, non-LEP, and non-FRL students respectively. The increase from fall to winter is significantly steeper (larger) for SpEd students than non-SpEd students. The increase from winter to spring for LEP students and ethnic minority students is respectively steeper than non-LEP students and flatter for non-ethnic minority students.

Fig. 4 Predicted growth trajectories for all grade 3 student subgroups based on the interaction model. The *circles* indicate statistically significant effects



Finally, in grade 5, female students had higher intercepts than male students. From fall to winter only SPED students demonstrated significantly different (steeper) slopes than their non-SPED peers. And from winter to spring, female and LEP students demonstrated significantly steeper slopes than their male and non-LEP peers.

The vagaries of these effects are difficult to interpret and may be meaningless. Figures 4, 5, 6 show that several intercepts and slopes look nearly as different from the main trajectory (e.g., male, non-SpEd, non-LEP), but due to small sample sizes, some are significant and others are not. Thus, it seems SpEd students clearly have different intercepts and slopes, and all subgroups except ethnic minorities had different intercepts, which were our results for the simpler main effects model.

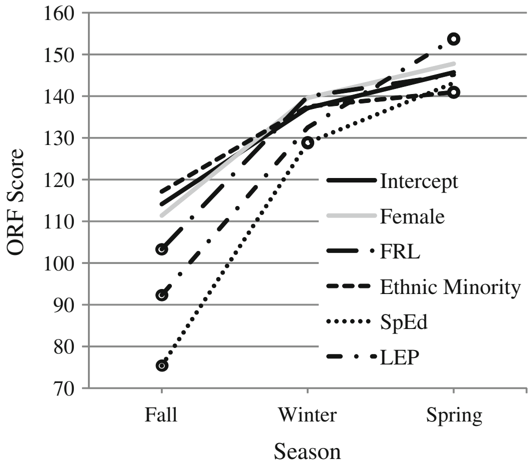


Fig. 5 Predicted growth trajectories for all grade 4 student subgroups based on the interaction model. The *circles* indicate statistically significant effects

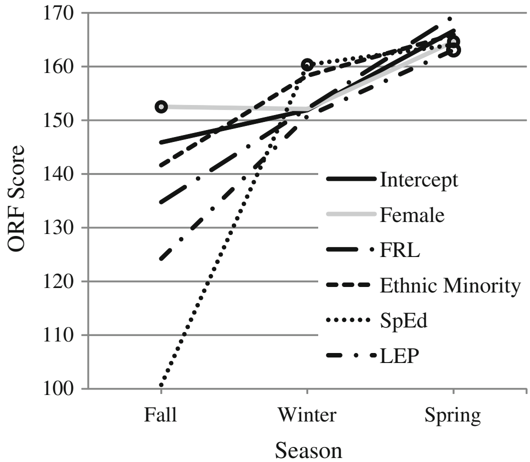


Fig. 6 Predicted growth trajectories for all grade 5 student subgroups based on the interaction model. The *circles* indicate statistically significant effects

References

Alonzo, J., Mariano, G. J., Nese, J. F., & Tindal, G. (2010). *Reliability of the easyCBM® reading assessments*. San Diego, CA: Paper presented at the Pacific Coast Research Conference.

Alonzo, J., & Tindal, G. (2007a). *The development of fifth-grade passage reading fluency measures for use in a progress monitoring assessment system (Technical Report # 43)*. Eugene, OR: Behavioral Research and Teaching.

Alonzo, J., & Tindal, G. (2007b). *The development of word and passage reading fluency measures in a progress monitoring assessment system (Technical Report No. 40)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Alonzo, J., Tindal, G., Ulmer, K., & Glasgow, A. (2006). *EasyCBM® online progress monitoring assessment system*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Ardoin, S., & Christ, T. (2008). Evaluating curriculum-based measurement slope estimates using data from triannual universal screenings. *School Psychology Review, 37*, 109–125.

Christ, T. J., Silberglitt, 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.

Cummings, K. D., Good, R. H., Powell-Smith, K. A., Smolkowski, K., Baker, S., Gau, J., & Atkins, T. (2008). *ROC done right: examining the decision utility of educational measures*. San Diego, CA: Symposium presented at the 16th Annual Pacific Coast Research Conference.

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

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

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

Fuchs, L. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33*, 188–192.

Fuchs, L., Fuchs, D., Hamlett, C., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Reveiw, 22*, 27–48.

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 & Practice, 18*(3), 157–171.

Good, R., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61–88). New York, NY: The Guilford Press.

Good, R. H., & Kaminski, R. A. (2003). *Dynamic indicators of basic early literacy skills* (6th Ed). Longmont, CO: Sopris West.

Graney, S., Missall, K., Martinez, R., & Bergstrom, M. (2009). A preliminary investigation of within-year growth patterns in reading and mathematics curriculum-based measures. *Journal of School Psychology, 47*, 121–142. doi:10.1016/j.jsp.2008.12.001.

Hasbrouck, J., & Tindal, G. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher, 59*, 636–644. doi:10.1598/RT.59.7.3.

Kaminski, R., Cummings, K. D., Powell-Smith, K. A., & Good, R. H. (2007). *Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) for formative assessment and evaluation. Unpublished manuscript, Dynamic Measurement Group, University of Oregon*. Oregon: Eugene.

Keller-Margulis, M., Shapiro, E., & Hintze, J. (2008). Long-term diagnostic accuracy of curriculum-based measures in reading and mathematics. *School Psychology Review, 37*, 374–390.

Kranzler, J. H., Brownell, M. T., & Miller, M. D. (1998). The construct validity of curriculum-based measurement of reading: An empirical test of a plausible rival hypothesis. *Journal of School Psychology, 36*, 399–415. doi:10.1016/S0022-4405(98)00018-1.

Pearson (2008). *AIMSweb progress monitoring and RTI system*. Retrieved May 12, 2010 from https://www.aimsweb.com/

Pearson (2011). *AIMSweb growth aggregate, Reading*(*R-CBM*). Retrieved March 9, 2010 from https://www.aimsweb.pearson.com/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed. ed.). Thousand Oaks, CA: Sage.

Shapiro, E., Keller, M., Santoro, L., & Hintze, J. (2006). Curriculum-based measures and performance on state assessment and standardized tests. *Journal of Psychoeducational Assessment, 24*, 19–35.

Shinn, M. (2007). Identifying students at risk, monitoring performance, and determining eligibility within response to intervention: Research on educational need and benefit from academic intervention. *School Psychology Review, 36*, 601–617.

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

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.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis*. New York: Oxford University.

Stage, S. (2001). Program evaluation using hierarchical linear modeling with curriculum-based measurement reading probes. *School Psychology Quarterly, 16*, 91–112. doi:10.1521/scpq.16.1.91.19159.