UC Riverside
UC Riverside Electronic Theses and Dissertations

Title
Literacy Progress Monitoring: Efficiency Versus Stability

Permalink
https://escholarship.org/uc/item/9f7719gx

Author
Garcia, Melissa Jeanne

Publication Date
2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Literacy Progress Monitoring: Efficiency Versus Stability

Master of Arts

in

Education

by

Melissa Jeanne Garcia

September 2012

Masters Committee:
Dr. Mike Vanderwood, Chairperson
Dr. Gregory Palardy
Dr. Cathleen Geraghty

Copyright by
Melissa Jeanne Garcia
2012

The Thesis of Melissa Jeanne Garcia is approved:

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE THESIS

Literacy Progress Monitoring:
Efficiency Versus Stability

by

Melissa Jeanne Garcia

Masters of Arts, Graduate Program in Education
University of California, Riverside, September 2012
Dr. Mike Vanderwood, Chairperson

Growth rates generated from a single probe per measurement occasion versus three probes and taking the median were compared by examining the scores psychometric characteristics. Students who were struggling in reading were monitored with DIBELS Next progress monitoring passages during the academic year. Data were collected from six elementary schools by the primary researcher yielding seven weeks of data for 219 students. A paired-samples t-test indicated a statistical difference between the single probe and median probe model growth rates ($p < .01$). However, predictive validity analysis indicated both were significant predictors of future reading performance ($p < .05$). There were also minimal differences between the standard error of the estimates and R-squared values produced. Based on the results generated, it is suggested that a single probe progress monitoring approach would not compromise the predictive validity and utility for making important educational decisions (i.e. guide instructional decision-making).

Table of Contents

List of Tables

Literacy Progress Monitoring: Efficiency vs. Stability

The early identification of students in need of academic assistance increases the likelihood of educational success (Jenkins & O'Connor, 2002; Torgesen, 2002). In addition, prevention efforts are more cost efficient, beneficial, and effective than delayed services. These and other notions support implementation of early screening and intervention for prevention of academic difficulties. However, there are still questions regarding how best to monitor student progress during these prevention efforts.

A proposed system of identification and prevention of academic difficulties is Response to Intervention (RtI), or a Multitiered Systems Approach. RtI is "the practice of (1) providing high-quality instruction and intervention that match students' needs and (2) using students' learning rate over time and level of performance to make important education decisions" (Buffum, Mattos, & Weber, 2009, p.14). RtI is considered a comprehensive, systematic approach to addressing learning problems for all students because it uses differentiated and intensified assessment and instruction (Wixson, 2011). Most Multitiered Systems models are composed of three tiers: in Tier I all students in a given class, school, or district receive some differentiated instruction through the core curriculum and 80% are expected to benefit. Each student's rate of growth is monitored using standard triannual universal screening procedures across the academic year (fall, winter, spring). In Tier II, those who do not make sufficient progress in Tier I instruction receive more targeted and intensive instruction. Progress is continually monitored on a biweekly basis. An additional 15% of students are expected to succeed with this supplemental instruction/intervention. The students who do not demonstrate progress within Tier II interventions are considered for more targeted and intensive interventions. Responsiveness to these interventions is typically measured on a weekly basis. The critical components of each tier include the universal screening of all students, selection and implementation of evidence-based interventions, and lastly the use of frequent progress monitoring with appropriate measures in order to assess progress towards goals (Reschly & Bergstrom, 2009).

Progress Monitoring

One critical component that has received considerable empirical support is the area of progress monitoring. Progress monitoring measures allow for databased decision making using high-quality, direct measures of performance. These measures include a system of brief assessments that are given frequently, at least once a month, to determine if a student is making satisfactory progress through the curriculum and is likely to meet long-term goals (Stecker, Fuchs, & Fuchs, 2008). The data generated from each assessment period are plotted on a graph and a line of best fit is superimposed on the data to indicate the student's rate of improvement. Progress monitoring scores provide teachers and practitioners with information regarding a student's level of performance and his or her rate of academic improvement. If yearly goals are set, a student's initial performance (i.e., baseline) can be connected to the long-term goal in order to indicate the rate of improvement that is expected. A teacher or practitioner can then compare the student's actual rate of improvement to his or her projected rate of improvement according to empirically derived benchmarks or national performance. These data determine if the student is responding appropriately to the instructional program or intervention provided. It is important that progress monitoring measures have the following characteristics: (a) be quick to administer, (b) have adequate reliability and validity, (c) be representative of what the student is learning, (d) aid in intervention development, and (e) be sensitive to academic performance so intervention effectiveness can be evaluated (Elliot & Fuchs, 1997).

When utilizing progress monitoring measures there are two decision rules that help teachers set ambitious goals and determine when intervention modifications are necessary. One approach is the trend-line rule whereby the student's current rate of performance (expressed by a trend line) is compared with the projected rate of progress (i.e., goal line). If the trend of student performance is steeper than the goal line, the goal should be raised. If the trend of student performance is less steep than the goal line, then an instructional or intervention modification is needed. The trend-line rule requires at least four weeks of intervention and at least eight data points. If at least three weeks of intervention have occurred and at least six data points have been collected, the three-point data rule can be utilized. According to this rule, an instructional change should be made if three consecutive data points are below the goal line. If three consecutive data points are above the goal line, the student's goal should be raised. If the data points are neither above nor below the goal line, the student's instructional program should be continued as is and his or her progress monitored. Research has suggested that when teachers use decision rules to guide instruction student outcomes are enhanced (Fuchs, Fuchs, & Hamlett, 1989a; Fuchs, Fuchs, & Hamlett, 1989b

In their seminal meta-analysis, Fuchs and Fuchs (1986) examined the effects of formative evaluation procedures (i.e., progress monitoring) on student achievement. They described systematic formative evaluation as the preferential approach to developing individualized educational programs due to its inductive nature, which avoids reliance on initial diagnoses of learner characteristics. Results provided evidence that when teachers use progress monitoring to inform instructional decisions, there is a significant improvement in student outcomes. Subsequent research has yielded similar results (Fuchs, Deno, & Mirkin, 1984; Fuchs, Fuchs, Hamlett, & Stecker, 1991; Jones & Krouse, 1988).

Limitations with Current Research

At present, while progress monitoring has been established as a valid and reliable tool, there is considerable variation in the way these assessment measures are administered (Mellard, McKnight, & Woods, 2009; Wesson, King, & Deno, 1984). These variations have led researchers to focus efforts on establishing best practices within a multitiered systems framework. One such area of practice that continues to be debated in the field is the number of fluency-based probes to use per progress monitoring occasion.

Within the schools, it is common practice to use one fluency probe when progress monitoring students to assess the impact of Tier II interventions. Yet, examination of typical response indicates a significant amount of variation from probe to probe (Jenkins, Graff, & Miglioretti, 2009). Although a trend line analysis can eliminate the variation, teachers may interpret the wide variation in scores as limiting the quality of the data. As a result, they may make inaccurate changes to instruction. Furthermore, in the most recent Dynamic Indicators of Basic Early Literacy Next (DIBELS Next) technical report, the authors suggest, "for individual progress monitoring, we recommend assessing with one passage on each progress monitoring occasion and making decisions about progress based on the moving median of the three most recent passages" (Powell-Smith, Good, & Atkins, 2010, p. 39). This report suggests practitioners can administer one assessment probe during each progress monitoring session. Yet, in this same manual, the authors recommend obtaining more information (i.e., administer more alternate forms) for students with variable progress monitoring trend lines. These two recommendations could create a conundrum for a practitioner attempting to measure student response to intervention, because more data points utilizing a single probe will need to be administered to generate a more stable trend line. An alternative solution to this issue has been proposed, whereby more accurate estimates of student growth will be obtained by administering three probes in one progress monitoring session and taking the median. The challenge is to determine whether the additional time devoted to progress monitoring does in fact generate more stable and accurate results.

Despite this problem, there is minimal research within the field regarding which practice is best. While oral reading fluency measures are considered more efficient than other methods, teachers regard once-or twice-a-week progress monitoring as impractical (Wesson, King, & Deno, 1984). In addition, opposing researchers suggest using measures of growth to predict future reading performance do not make contributions to prediction that are independent of measures of achievement status (Schnastneider et al., 2008).

One study, conducted by Jenkins, Graff, and Miglioretti (2009), did address the question of how many scores per measurement occasion are appropriate for valid estimates of reading growth. The authors explored the validity of growth estimates derived from more and less frequent schedules of progress monitoring (i.e., monitoring every week, every two weeks, every three weeks, every four weeks, and every nine weeks) and the number of scores to establish baseline. Within these varying measurement schedules, students were measured on two to four passages in a sitting. Doing so allowed for comparison between growth slopes using one or more curriculum based measurement (CBM) scores at each monitoring point. In addition, Jenkins and colleagues computed slopes using only one score per monitoring occasion. Participants included 41 students ranging from fourth through eighth grade. Standard CBM passages for grades first through sixth developed at Vanderbilt University were used, and various studies have demonstrated adequate validity relative to oral reading skills (Compton et al., 2006). Slopes produced from five monitoring schedules, one passage every week, two passages every two weeks, three passages every three weeks, four passages every four weeks, and first/last weeks only, were compared. Linear regression was used to estimate the average growth in words read correctly (WRC) per week (slope). The researchers calculated 15 slopes for each student. One slope, the standard for which other slopes were compared, was computed using all 29 WRC scores and is considered the "true slope." The overall result of reducing the number of progress monitoring passages to one per measurement occasion was to increase slope estimates. This indicated that when only one progress monitoring passage is used, growth rates were significantly inflated. This inflated growth rate could lead to teachers delaying a change in instruction or intervention. Furthermore, results indicated that using fewer progress monitoring scores per measurement occasion significantly degraded concurrent validity for all but the weekly measurement scheme.

Purpose of Study

Due to the lack of research in this area, the purpose of this study elaborated on Jenkins, Graff, and Miglioretti's work and focused solely on the question of the number of probes to give per measurement occasion. The research questions addressed in this study are:

1. Is there a significant difference in growth rates produced using three oral reading progress monitoring probes and taking the median versus one?
2. Is there a difference between the amount of variance accounted for when using one oral reading fluency progress monitoring probe versus three on outcome variables (California Standards Test English Language Arts).
3. For individual cases, is there increased variability when using one oral reading probe versus three and taking the median.
4. Is there a significant difference in growth rates produced using three oral reading progress monitoring probes and taking the median versus one when looking at passage level?
5. Is there a difference between the amount of variance accounted for when using one oral reading fluency progress monitoring probe versus three on outcome variables (CST ELA) when looking at passage level?

Method

Participants and Setting

Participants included 219 students receiving Tier 1 and Tier 2 reading interventions from six schools in an urban Southern California school district. The sample was composed of 77 second graders (35%), 45 third graders (21%), 53 fourth graders (24%), and 44 fifth graders (20%). There were 100 females (46%) and 119 (54%) males. The overall percentage of students receiving free and reduced lunch for the district was 48%, with 41.4% at the elementary level. The majority of participating students were Hispanic/Latino (53%). The rest of the sample included Caucasian (45%), Asian American/Pacific Islander (2%), and African American (1%). The sample included 28% English Language Learners. All of the participating students received interventions such as, Peer Assisted Learning Strategies (24%), Language! (20%), Rewards (16%), Voyager Passport (10%), Push in Reading (7%), and Differentiated Instruction Time/In-Class Intervention (5%).

Materials

Dynamic Indicators of Basic Early Literacy Skills Next (DIBELS Next). As part of the school district assessment procedures, standard Dynamic Indicators of Basic Early Literacy Skills Next passages from grades one to five were used to assess student's accuracy and fluency. DIBELS Oral Reading Fluency (DORF) test-retest reliabilities ranged from .91 to .97 for elementary students and alternate form reliability for the different passages .90 to .97 (Powell-Smith, Good, & Atkins, 2010). All of these reliability coefficients fall within range of the decision-making standards for individuals (Salvia, Ysseldyke, & Bolt, 2010). Validity coefficients ranged from moderate to strong with other DIBELS scores, GRADE Total Test, and the Standard 4th Grade Reading

Passage used in the NAEP 2002 Special Study of Oral Reading. Criterion-related validity with the external criterion GRADE Total Test ranged from .48 to .77 for the beginning of the year, .47 to .80 for the middle of the year, and .40 to .75 for the end of the year. Concurrent validity coefficients with DIBELS Next Retell and Daze, and the Standard 4th Grade Reading Passage range from .44 to .76, .70 to .78, and .83 to .97. Predicative validity coefficients for Oral Reading Fluency with Daze measures ranged from .40 to .79.

Inter-rater reliability. Inter-rater reliability data were collected and calculated between two administrators of the DORF probes. The number of agreements (in scoring students' accuracy on individual words) was divided by the total number of words in the passage, and aggregated across 12 passages. A total of six observations were made. The inter-rater reliability was 97%.

Procedure

Participating students were progress monitored once or biweekly throughout the school year by teachers, school psychologists, and intervention specialists. As per the district's procedures, weekly or biweekly (depending on Tier 2 or Tier 3 status) the students were administered three DIBELS Next Oral Reading Fluency probes per measurement occasion and the median taken. The participants were administered instructional and grade level probes. Following empirical guidelines (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993), instructional level was established by testing back to the highest level in which the student could read 70-100 wpm on third through fifth grad probes or 40-60 wpm on first and second grade probes. On grade level probes, if a student could not read at least 10 words, testing back occurred until they could meet this standard. This aligns with DIBELS administration instructions stating that testing should be discontinued if a student cannot read at least 10 words correct on the first page (Good & Kaminski, 2002). The primary researcher collected the progress monitoring booklets after the completion of the academic school year, yielding seven weeks of progress monitoring data for each student. Each of the seven weeks of progress monitoring data included in the analyses included three DORF scores. The median was taken from the three scores included in each of the seven weeks of progress monitoring data. The single passage chosen was the first data point in each week. This procedure aligned with previous research (Jenkins, Graff, & Miglioretti, 2009). For each student, single probe data and median data were graphed separately and a linear least squares regression line was fitted to represent the trend line. Based on the trend line, the slope (i.e. growth rate) was determined for both. Data were verified prior to analysis.

Results

Differences Between Growth Rates

First a descriptives statistics table was generated to provide an initial description of the data. The data indicated a minor departure from a normal distribution, which analysis of skewness and kurtosis confirmed (see Table 1). However, linear models tend to be robust against these types of violations (Raykov & Marcoulides, 2008). To answer research question one, whether there is a significant difference between the growth slopes generated from this procedure, a pairwise t-test was utilized. The Pearson r correlation between growth rates was found to be significant in a positive direction (r (217)=. 80, p<. 001). The paired t-test was significant (t (218)=-3.49, p<. 001). The growth rates produced from single passages (μ=. 58, *SD*=1.0) versus the median (μ=. 73, *SD*=. 89) were significantly different. The median model generated a higher mean growth rate and a smaller standard deviation, indicating less variability than the single model. To gain a better understanding of why the median score was higher than the single probe approach, the single probe approach and median probe approach were compared to using two probes and taking the average. The growth rates produced from single passages (μ=. 58, *SD*=1.0) versus two and taking the average (μ=. 82, *SD*=. 91) were significantly different (t (218)=-6.70, p<. 001). Similarly, there was a significant difference t (218)=3.59, p<. 001) between the median probe approach (μ=. 73, *SD*=. 89) versus two and taking the average (μ=. 82, *SD*=. 91). While utilizing two passages and taking the average yielded a higher growth rate than the single and median approach, the median probe approach still had the smallest standard deviation, indicating the least variability.

The Pearson r correlation between CST ELA scores and the median baseline score was found to be significant in a positive direction (r (216) =. 33, p < . 001), as was the single baseline score (r (216) =. 31, p <. 001). This indicated that as CST ELA scores increased, baseline scores also increased. Both the median growth and single growth rates were not significantly correlated with the CST ELA (r (216) =. 10, p > .05). The median slope model was found to be significantly negatively correlated with the median baseline score (r (216) = -.23, p <. 001). As median oral reading fluency scores increased, median growth rates decreased. Findings were similar for the single growth rate model and single baseline score (r (216) = -.34, p <. 001).

Predictive Validity

Predictive validity was examined using multiple regression analyses, with growth rates and baseline scores as the predictor variables and CST ELA scores as the dependent variable (see Table 3). Previous research has found that growth rates alone do not add unique information to prediction of future reading skills (Schnastneider, Wagner, & Crawford, 2008). Therefore, an oral reading fluency baseline score was added as a second predictor. The single probe model was a significant predictor of CST ELA scores, F (2,213) = 16.14, $p < .001$, and explains 12% of the variance in CST ELA scores. The baseline point was a significant predictor, $t = 5.55$, $p < .001$, as was the single passage growth rate, $t = 3.02$, $p < .01$. The alternative model for median probes was also a significant predictor of CST ELA scores, F (2,213) =17.59, $p < .001$. The model explained 13% of variance in CST ELA scores. Both the baseline point, $t = 5.70$, $p < .001$, and median growth rate, $t = 2.89$, $p < .01$, were significant predictors of CST ELA scores.

To examine if the predictive validity coefficients for the single probe model versus the median probe model were significantly different a Fisher's Z test was used (Fisher, 1921). The correlation coefficients were converted to z-scores using the formula $Z_r = [\ln(1+r)-1n(1-r)]/2$. The standard error of the estimate, s_{zr}, was calculated using the formula $1/\sqrt{(n-3)}$. The null hypothesis of no significant differences between the correlations (H_0: $Z_{r1}=Z_{r2}$) was tested using the formula $Z = (Z_{r1}-Z_{r2})/ s_{zr}$. The coefficients for the single probe model and median probe model were found not to be statistically different, $z = .24$, $p < .05$.

To determine if there was a significant difference between the amounts of variance accounted for when using a single probe model versus a median model on the outcome variable (CST ELA scores) the R-Squared values for both models were compared. As discussed previously, the single probe model accounted for 12% of the total variation in CST ELA scores and the median model accounted for 13%. The change in R-squared value for the single model was .034 and .037 for the median model. There was minimal difference between the values, suggesting there is not a significant difference between the approaches in terms of amount of variance accounted.

Standard Error of the Estimate

The standard error of the estimate (SEE) was analyzed to determine if one model generated a more accurate measure of the predictors (i.e., has a smaller SEE). Standard error of the estimate is the standard deviation of the prediction errors. The SEE value for the single probe model was 26.4 and the median model was 26.6, indicating minimal differences between the model's SEE values.

Passage Level Differences

To determine if there was a significant difference between the growth slopes generated and passage levels (i.e., 2^{nd} and 3^{rd} grade), pairwise t-tests were utilized. The paired t-test was significant for second and third grade passage levels (t (31)= -2.29, $p <$ .05; t (59)= -4.35, $p < .001$; see Table 5). The growth rates produced from single passages versus the median were significantly different for second and third grade passage levels. The remaining passage levels indicated no significant differences between the single growth rate model and the median model.

Passage Level Predictive Validity

The predictive validity of passage levels was examined to answer question five using multiple regression analyses, with growth rates and baseline scores as the predictor variables and CST ELA scores as the dependent variable (see Table 4). The single probe model was a significant predictor of CST ELA scores for first grade level passages, F (2,75) = 10.49, $p < .001$, and second grade level passages, F (2,29) = 4.38, $p < .05$. First grade passages explained 20% of the variance in CST ELA scores and second grade explained 18% of the variance. The baseline point for first grade was a significant predictor, $t = 2.74$, $p < .01$, as was the single passage growth rate, $t = 2.03$, $p < .05$. For second grade, the baseline point was a significant predictor, $t = 4.07$, $p < .01$, as was the growth rate, $t = 2.33$, $p < .05$. The alternative median probe model was a significant predictor of CST ELA scores with second (F (2,29) = 4.93, $p < .01$) and third grade passage levels (F (2,57) = 3.12, $p < .05$). The grade passage model explained 20% of variance in CST ELA scores and the third grade model explained 7% of the variance. The baseline point, $t = 2.67$, $p < .01$, and median growth rate, $t = 2.32$, $p < .05$, were significant predictors of CST ELA scores. For third grade, the baseline point was not a significant predictor, $t = 1.81$, $p > .05$, however, the growth rate was significant, $t = 2.12$, $p < .05$.

Discussion

At present there is considerable variation in the way progress monitoring measures are utilized within schools (Mellard, McKnight, & Woods, 2009; Wesson, King, & Deno, 1984). One such area of variation is the number of passages used per measurement occasion. While DIBELS Next and AIMSweb promote using a single passage approach when progress monitoring, an examination of typical response indicates a significant amount of variation from probe to probe (Christ, 2006; Jenkins, Graff, & Miglioretti, 2009). To address this variation, the median approach or administering three passages and taking the median score, has been proposed. Proponents of this approach suggest this method will generate more accurate estimates of student growth than a single passage approach. While this issue is in need of addressing, to date, only one study has explored difference between the two approaches.

The purpose of this study was to elaborate on the work of Jenkins and colleagues (2009) and to determine whether there was a difference between utilizing a single passage approach to progress monitoring versus a median passage approach. Previously, Jenkins, Graff, and Miglioretti found that a single passage approach significantly inflated students' growth rates. Results from the present study indicated that a single passage progress monitoring approach would not compromise predictive validity and utility for making important educational decisions (i.e., guide instructional decision-making). Findings did indicate there was a significant difference in the growth rates produced and the standard deviation generated for the median model indicated less variability. However, both significantly predicted CST ELA scores and there was minimal difference between the amount of variance accounted for in CST ELA scores (R-squared values). Correlation matrices for both median growth rates and single growth rates were similar and while the growth rates were not found to be significantly correlated with CST scores, a correlation of .10 between growth rates and future reading performance is aligned with previous research (Schnastneider, Wagner, & Crawford, 2008). There were some differences between models by passage levels, whereby only first and second grade passage levels for the single model significantly predicted CST ELA scores and second and third for the median model. These results indicated the growth rates are predicting slightly differently on CST ELAs depending on grade level passages. There was also a significant difference found between the growth rates produced for second and third grade passage levels.

Higher SEE Values

Another area to examine when using curriculum based reading measures in practice is the stability of progress monitoring outcomes and CBM-R slopes based on the standard error of the estimate (SEE; Christ, 2006). Christ estimates that SEE results in a standard error of the slope (SEb) larger than expected, making the interpretation of individual students' rates of growth using CBM-R progress monitoring procedures difficult. Research has suggested that longer progress monitoring durations can reduce the magnitude of SEb and a smaller magnitude of SEE is likely to occur when measurement conditions are optimal (Christ, 2006). If both of these conditions are met, the SEE might fall in one of the optimal levels (2-6). However, if either (not both) the measurement conditions or the CBM-R probes are not well controlled, the SEE may fall in the moderate levels (8-12). If neither condition is met, SEE is likely to fall within the poor levels (14-18). With regards to the two most popular curriculum based measures, Ardoin and Christ (2009) found that AIMSweb yielded a mean SEE value of 11 and DIBELS a value of 15. These mean values were generated from twice weekly progress monitoring schedules over ten weeks. The authors suggested the larger SEE means are due to the inconsistency of passage difficulty within progress monitoring sets.

With regard to the present study, there was a minimal difference in the standard error of the estimate values produced (26.4 versus 26.6). However, the data produced a larger SEE than found in previous research when using DIBELS (Ardoin & Christ, 2009). The SEE values were also larger than the optimal, moderate, and poor levels suggested by Christ (2006) when evaluating passage quality. A possible explanation for this discrepancy is that both previous studies collected ten weeks of progress monitoring data and two measurement occasions per week. By comparison, the data in the present study is based on seven weeks of progress monitoring and one measurement occasion per week or biweekly. Christ (2006) suggested that longer progress monitoring durations reduces the magnitude of SEE, a limitation of the present research with regard to SEE values. As discussed previously, the SEE also depends on measurement conditions and the degree of standardization. Due to the data being secondary, limited information is available regarding each examiners degree of standardization and the measurement conditions, which could have caused a higher SEE value than found in previous research.

Comparisons to Previous Research

It was unexpected that there was a minimal difference between the single passage model and median passage model approach. Based on prior research (Jenkins, Graff, & Miglioretti, 2009) and the notion that a median approach would generate more stable and accurate growth rates, it was hypothesized that the median model predictive validity would be stronger. As discussed previously, Jenkins and colleagues (2009) found that reducing the number of progress monitoring passages to one per measurement occasion negatively affected validity. The present study did not find such results, instead a single passage approach to progress monitoring versus a median approach did not degrade validity. However, methodological variations from the Jenkins, Graff, and Miglioretti study should be noted. Jenkins and colleagues compared growth slopes generated from various numbers of CBM scores to an estimate of a "true slope," which included all 29 collected CBM scores. These methodological differences could explain the discrepancy in results. It was also expected that there would be a significant difference between the growth rates with the median model yielding less variability, however, upon further analysis, these results were surprising in that both models had similar predictive validity results. A possible explanation for this result is that despite varying growth rates, both models are comparable in predicting CST ELA scores.

Despite the unexpected results, these findings are significant for practitioners. As educators move towards using a Response to Intervention (RtI) approach, the above findings demonstrate that a more time efficient approach to progress monitoring (single passage per measurement occasion) can be used without compromising predictive validity. In addition, since the main goal of progress monitoring is to guide instructional decision making and to increase the current instruction and intervention (Stecker, Fuchs, & Fuchs, 2008), results indicated the single passage approach would be as equally effective in depicting student growth.

Another interesting finding was the similarity of growth rates produced in the present study and those found by Deno and colleagues (2001) for special education students. For both the single passage model and median model the growth rates were comparable across most grade passage levels. There were some exceptions, such as within the single passage model, third and fifth grade were much lower than those found by Deno. In addition, the second grade passage level growth rate for the median model was higher in the present study. Otherwise, most were highly similar. For instance, the median growth rate for first grade passage levels was .88 and Deno and colleagues found a growth rate of .83. The single growth rate for fourth grade passage levels in the present study was .51, whereas Deno's study yielded a growth rate of .58. These similarities increase the validity of the present study's results in that a significantly smaller sample size was utilized, however, comparable results were still produced. Important to note is that the growth rates generated in the present study were based on passage level rather than a student's grade level. This varies from Deno's work and may limit the association between studies.

Predictors of Future Reading Performance

A possible critique of the above results is based upon prior research from Schnastneider and colleagues (2008) whom suggest measures of growth do not add to prediction of future reading performance. While the single and median growth rates added little additional variance to the overall models, both were significant predictors at the .01 level. In addition, as discussed previously, the main purpose of progress monitoring is to guide instructional decision-making and to increase the current instruction and intervention (Stecker, Fuchs, & Fuchs, 2008), not to predict outcomes on state tests.

Limitations

There are several limitations associated with the present study. The use of secondary data in this study limited the researchers ability to monitor and control for standardization of administration of DIBELS Next passages to students. As discussed previously, the measurement conditions and level of standardization could have contributed to the higher standard error of the estimate value found. The stability of the trend line could have been less impacted had measurement conditions been optimal (i.e., extraneous variables well controlled). In addition, longer progress monitoring durations reduce the magnitude of SEE and increase the stability of trend lines. Had more progress monitoring data been collected, perhaps the SEE values would have been reduced. Difficulty level of passages could also influence the stability of trend lines. DIBELS DORF passages were developed according to the DMG Passage Difficulty Index (Cummings, Wallin, Good, & Kaminski, 2007) and each progress monitoring set includes triads (and one dyad) of slightly easier, middle, and slightly harder passages. Not all probes were administered in the same order, which could have influenced the stability of trend lines. Furthermore, the first passage was chosen for the single probe analysis, which could also influence the trend lines and limits the generalizability of results. Another limitation of the present study was the limited interrater data available due to the data being secondary. While these factors are limitations, within the district annual trainings are provided by school psychologists to those administering DIBELS Next at their school sites, as well as ongoing one-on-one consultation throughout the year. These trainings increase the likelihood of a standard and consistent practice among examiners. Lastly, due to district procedures, students were progress monitored at instructional level rather than grade level. This limitation affects the presents study's external validity and ability to make generalizations regarding using a single passage approach. A goal for future research would be to replicate the above results in a more controlled study, whereby participants are progress monitored at grade level.

References

Ardoin, S.P. & Christ, T.J. (2009). Curriculum-based measurement of oral reading: Standard errors associated with progress monitoring outcomes from DIBELS, AIMSweb, and an experimental passage set. School Psychology Review, 38(2), 266-283.

Buffum, A., Mattos, M., & Weber, C. (2009). Pyramid response to intervention: RTI, professional learning communities, and how to respond when kids don't learn. Bloomington, IN: Solution Tree.

Christ, T. J. (2006). Short term estimates of growth using curriculum-based measurement of oral reading fluency: Estimates of standard error of the slope to construct confidence intervals. *School Psychology Review, 35,* 128 –133.

Compton, D. L., Fuchs, L. S., Hamlett, C. L., Powell, S. L., Capizzi, A., & Morgan, P. (2006, February). *Are CBM passages comparable: Examining the effects of set, grade, child characteristics, and passage characteristics on oral reading fluency?* Paper presented at the Fourteenth Annual Meeting of the Pacific Coast Research Conference, Coronado, CA.

Cummings, K. D., Wallin, J., Good, R. H. III, & Kaminski, R. A. (2007). The DMG Passage

Difficulty Index [Formula and Computer Software]. Dynamic Measurement Group, Eugene, OR.

Elliot, S.N., & Fuchs, L.S. (1997). The utility of curriculum-based measurement and performance assessments as alternatives to traditional intelligence and achievement tests. School Psychology Review, 26, 224-233.

Fisher, R.A. (1921). "On the 'probable error' of a coefficient deduced from a small sample. Metron, 1, 3-32.

Fuchs, L.S., Deno, S.L., & Mirkin, P.K. (1984). Effects of frequent curriculum-based measurement and evaluation of pedagogy, student achievement, and student awareness of learning. American Educational Research Journal, 21, 449-460.

Fuchs, L.S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. Exceptional Children, 53(3), 199-208.

Fuchs, L.S., Fuchs, D., & Hamlett, C.L. (l989a). Effects of alternative goal structures within

curriculum-based measurement. Exceptional Children, 55, 429-438.

Fuchs, L.S., Fuchs, D., & Hamlett, C.L. (l989b). Effects of instrumental use of curriculum-based

measurement to enhance instructional programs. Remedial and Special Education, 10(2),

43-52. Implications for Practice.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., & Stecker, P.M. (1991). Effects of curriculum-based measurement on consultation on teacher planning and student achievement in mathematics operations. American Educational Research Journal, 28, 617-641.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*(1), 27-48.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). Dynamic Indicators of Basic Early Literacy Skills (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Good. R. H.. & Shinn. M. R. (1990). Forecasting accuracy of slope estimates for reading curriculum-based measurement: Empirical evidence. *Behavioral Assessment, 12.* 179-193.

Hintze, J. M., & Christ, T. J. (2004). An examination of variability as a function of passage variance in CBM progress monitoring. *School Psychology Review, 33,* 204–217.

Hintze, J. M., Daly, E. J., & Shapiro, E. S. (1998). An investigation of the effects of passage difficulty level on outcomes of oral reading fluency progress monitoring. *School Psychology Review, 27,* 433–445.

Jenkins, J.R., Graff, J.J., & Miglioretti, D.L. (2009). Estimating reading growth using intermittent CBM progress monitoring. Exceptional Children, 75(2), 151-163.

Jenkins, J. R., & O'Connor, R. E. (2002). Early identification and intervention for young children with reading/learning disabilities. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 99-149). Mahwah, NJ: Lawrence Erlbaum.

Jones, E.D., & Krouse, J.P. (1988). The effectiveness of databased instruction by student teachers in classrooms for pupils with mild handicaps. Teacher Education and Special Education, 11(1), 9-19.

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

Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2005). An investigation of the reliability and standard error of measurement of words read correctly per minute using curriculum-based measurement. *Journal of Psychoeducational Assessment, 23,* 326–338.

Powell-Smith, K. A., Good, R. H., & Atkins, T. (2010). *DIBELS Next Oral Reading Fluency Readability Study* (Technical Report No. 7). Eugene, OR: Dynamic Measurement Group.

Raykov, T. & Marcoulides, G.A. (2008). *An introduction to applied multivariate analysis*. New York: Routledge.

Reschly, D. J., & Bergstrom, M. K. (2009). Response to intervention: In T.B Gutkin & C.R Reynolds (Eds.), Handbook of school psychology (4th ed., p. 434-460). New York: Wiley.

Salvia, J., Ysseldyke, J., & Bolt, S. (2010). Assessment in special and inclusive education (11th Ed). Wadsworth Cengage Learning, Belmont: CA.

Schnastneider, C., Wagner, R.K., Crawford, E.C. (2008). The importance of measuring growth in response to intervention models: Testing a core assumption. Learning and Individual Differences, 18, 308-315.

Stecker, P.M., Fuchs, D., & Fuchs, L. (2008). Progress monitoring as essential practice within response to intervention. Rural Special Education Quarterly, 27(4), 10-17.

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School*

Psychology,40, 7-26.

Wesson, C.L., King, R.P., & Deno, S.L. (1984). Direct and frequent measurement: If it's so good for use, why don't we use it? Learning Disability Quarterly, 7, 45-48.

Wixson, K. (2011). A systematic view of RTI research: Introduction to the special issue. The Elementary School Journal, 111(4), 503-510.

Table 1

Model Predictor Descriptives

	n	Mean	*SD*	Skewness	Kurtosis
GRS	219	.58	1.00	.09	1.75
GRM	219	.73	.89	.54	2.21
BM	219	51.90	26.80	-.18	-1.19
BS	219	51.50	27.70	-.18	-1.13
CST	216	293.04	40.10	.51	.18

Note. GRS = Growth rate single; GRM = Growth rate median; BM = Baseline median score; BS = Baseline single score; CST = California Standards Test, English Language Arts; SD = Standard deviation.

Table 2

Model Descriptives by Passage Level

	Single Probe Model						Median Probe Model		
Passage Level	*n*	Mean BS	*SD*	Mean GRS	*SD*	Mean BM	*SD*	Mean GRM	*SD*
First	78	23.76	14.60	.92	.95	25.04	14.40	.88	.83
Second	32	49.53	12.30	.62	.84	48.00	11.50	.94	.90
Third	60	72.32	10.60	.20	1.14	71.30	10.90	.53	1.00
Fourth	36	76.08	15.80	.51	.89	77.36	13.70	.54	.72
Fifth	10	81.50	11.70	.19	.64	81.10	11.90	.67	.94

Note. Mean BS = mean single baseline score; SD = standard deviation; Mean GRS = mean growth rate single; Mean BM= mean median baseline score; Mean GRM = mean growth rate median.

Table 3

Predictive Validity Coefficients

Predictor	Single Probe Model	Median Probe Model
	CST ELA	CST ELA
Baseline Point	.54*	.54*
Growth Rate	7.72*	8.00*
R^2	.12	.13
F	16.14*	17.59*

Note. CST ELA = California Standards Test, English Language Arts
*p < .01, n= 216

Table 4

Grade Comparisons- Single Model vs. Median Model

Single Probe Model						Median Probe Model					
Passage Level	*n*	*B*	R^2	*F*	SEE	Passage Level	*n*	*B*	R^2	*F*	SEE
First	78	9.29**	.20	10.49*	33.00	First	78	7.92	.18	9.50*	33.40
Second	32	15.59**	.18	4.38**	33.60	Second	32	15.90**	.20	4.93*	33.10
Third	60	9.10	.04	2.31	35.70	Third	60	9.96**	.07	3.12**	35.30
Fourth	36	-7.94	.10	2.83	35.20	Fourth	36	-5.20	.09	2.72	35.30
Fifth	10	2.30	.14	.57	42.40	Fifth	10	10.96	.20	.86	41.00

* $p < .01$, ** $p < .05$

Table 5

Differences Between Passage Levels Growth Rates

Passage Level	*t*	Sig. (2-tailed)
First	.78	.44
Second	-2.23	.03**
Third	-4.35	.00*
Fourth	-.07	.95
Fifth	-2.14	.06

$^*p < .01$, $^{**}p < .05$