**Title**
Progress Monitoring at Grade vs Instructional Level



**Author**
Zhu, Yiwen

**Publication Date**
2011

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

# Progress Monitoring at Grade vs. Instructional Level

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Arts

in

Education

by

Yiwen Zhu

June 2011

Thesis Committee:
Dr. Michael L. Vanderwood, Chairperson
Dr. Sara Castro-Olivo
Dr. Gregory J. Palardy

The Thesis of Yiwen Zhu is approved:

_________________________________

_________________________________

_________________________________

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE THESIS

Progress Monitoring at Grade vs. Instructional Level

by

Yiwen Zhu

Doctor of Philosophy, Graduate Program in Education
University of California, Riverside, June 2011
Dr. Michael L. Vanderwood, Chairperson

This study compared the use of grade versus instructional level material for progress monitoring struggling readers by examining the materials' psychometric characteristics and the data's influence on teacher expectations and instructional decision-making. Students were progress monitored with both sets of probes for 6 – 8 weeks. Each set of data was analyzed for sensitivity to growth and predictive validity and evaluated by classroom teachers. Results found no statistical difference in sensitivity to growth or predictive validity ($p > .05$). Both data based decision rules based on trend analysis and teacher responses to questionnaires indicated that grade level data was more likely to portray student progress as inadequate and the current intervention as needing modification and/or increase in intensity. Limitations of the study and directions for future research are discussed.

**Table of Contents**

Progress Monitoring at Grade vs. Instructional Level

Research over the last three decades clearly identifies the importance of monitoring student progress and how engaging in progress monitoring can lead to enhanced student outcomes. By obtaining an accurate understanding of how a student is responding, educators can make targeted and immediate changes to promote better performance. In fact, this was the original purpose for which Curriculum Based Measurement (CBM) was developed (Stecker, Fuchs, and Fuchs, 2005). In 1986, Lynn and Doug Fuchs completed a meta-analysis that found a .70 effect size for improvement in student outcomes when teachers used progress monitoring data to inform instructional decisions.

One reason monitoring student progress is a valuable practice is because having objectives when teaching and goal setting lead to better outcomes (Fuchs, Fuchs, and Deno, 1985). Not only does research support that establishing goals improves outcomes, studies have shown that higher teacher expectations for students is related to better academic outcomes (Brookover et al., 1979; Brophy and Good, 1970; Rosenthal and Jacobson, 1968 as cited in Beady and Hansell, 1981). This “Pygmalion effect” in the classroom has been found in Rosenthal and Jacobson's Pygmalion experiment in 1968 and other subsequent studies (Rosenthal, 2002). In a randomized experimental study, Rosenthal and Jacobson told the teachers of one group of students that the students’ scores on a nonverbal IQ test predicted large gains in academic achievement in the next eight months. In reality, this was not the case. After 8 months, post-tests using the same IQ test found that students in the experimental group as a whole scored higher than those in the control group (Rosenthal and Jacobson, 1966, 1968, 1992 as cited in Rosenthal, 2002). Other researchers also found a similar effect of teacher expectations on students' performance on IQ tests (Radenbush, 1984, 1994; Smith, 1980 as citied in Rosenthal, 2002). A study conducted by Zanna and colleagues (1975) with design similar to the Pygmalion experiment found teacher expectations to affect performance on standardized tests in English and math. The English and math teachers' expectations were manipulated using made-up reports of student potential based on the Princeton Academic Potential Inventory. Students for whom teachers were induced to have higher expectations improved more on post-tests (Zanna et al., 1975). Finally, studies of teacher expectations for students of different races, such as studies of the white-black achievement gap, have suggested that the gap may be due in part to differences in teacher expectations for the students (Beady and Hansell, 1981).

**Academic Goal Setting Research**

Goal setting has been studied in a number of ways. An early and notable study by Fuchs, Fuchs, and Deno in 1985 examined the relationship of goal ambitiousness (using IEP goals) to student achievement with 58 special education students. Student achievement was measured using the Passage Reading Test (Fuchs et al. 1984) and the Stanford Diagnostic Reading Test. Teachers set specific goals, which included the number of words per minute the student should read at a certain passage level by a certain number of weeks. Goal ambitiousness was determined by the ratio of the goal to a three-day baseline performance, and grouped into high, moderate, and low ambitiousness based on pre-designated cutoffs. Students were progress monitored twice each week. A multiple analysis of covariance found goal ambitiousness to be significantly and positively correlated to achievement outcomes (F = 4.72, $p < .001$). Interestingly, no interaction was found between goal ambitiousness and students' handicapping conditions, indicating ambitious goals may be used for students of diverse conditions. However, in this study, goal ambitiousness conditions were not randomly assigned but rather instead were designated post hoc.

Fuchs, Fuchs, and Hamlett in 1989 found that dynamic (continuously adjusted) and ambitious goal setting led to better outcomes than when progress monitoring was neglected. Using CBM math computation, they compared outcomes of using static annual goals, dynamic goals, and a control group with no progress monitoring. In the dynamic goal setting condition, teachers were instructed to raise goals whenever student growth suggested that a more ambitious goal may be reached. In static goal setting, teachers were not *instructed* to raise the goal when the student's progress was steeper than the aim-line, though they could choose to do so if they thought it was appropriate. The Concepts of Number subtest from the Stanford Achievement Test (1982) and the Math Computation Test (Fuchs, 1987) were used as achievement outcome measures. Also recorded were the number of goal changes, the number of intervention changes, and the final goal to median baseline score ratio. Results showed that students in the dynamic group had more goal changes and a higher final goal to baseline ratio than the static group. Achievement outcomes on the Math Computation Test of the dynamic group were also significantly higher than that of the control group, $F(4, 48) = 2.71$, $p < .05$. Achievement outcomes of the dynamic group directly compared with the static group however, did not show a significant difference. However, a limitation was the static goal condition did not necessarily have static goals, as the teachers were allowed though not instructed to increase goals when appropriate.

Conte and Hintz (2000) studied methods of goal setting by comparing two graphical approaches. One method was to draw a diagonal line from the baseline point to the long-term goal (termed dynamic goal line in this study). This method provides weekly (or daily) performance goals throughout the intervention period. In an alternative method, a horizontal goal line is drawn across the graph, indicating only the long-term goal but not short-term goals throughout the intervention (termed static goal line in this study). Students in the group using dynamic goals made more growth on CBM oral reading fluency probes than both the static and control groups, with an effect size of .47 compared to the control group. It is important to note that in this study, students in each group were shown the graph before and after each time they were given a CBM oral reading fluency measure. Hence, the effect of goal awareness and the two methods of goal setting on students are confounded with the effect on teachers (Conte and Hintz, 2000).

**Use of Progress Monitoring to Improve Instruction**

It is important to note that merely collecting progress data does not lead to improved performance; the data must be utilized in instructional decision-making. Many studies have shown that when teachers do not make instructional changes when data indicates change is appropriate (poor implementation fidelity of data based decision making), their students do not make gains in performance (e.g. Tindal et al., 1981; Skiba et al., 1982; King et al., 1983; as cited in Stecker et al., 2005). On the other hand, when teachers are trained to make changes according to progress monitoring data and provided with periodical feedback, their students made significant gains compared to a control group (e.g. Fuchs et al., 1984; Jones and Krouse, 1988 as cited in Stecker et al., 2005). These studies indicate that teachers make few instructional modifications when they are not guided to do so. In most studies that found a positive effect of CBM, teachers received guidance from researchers on how to examine CBM data (Stecker et al., 2005).

An early study by Fuchs, Fuchs, and Stecker in 1989 examined the effects of CBM on instructional planning. Teachers in the CBM group progress monitored students at instructional level. Once every week, need for instructional change was discussed based on the data and ideas for reading interventions were provided. Teachers in the comparison group observed students during lessons and scored worksheets to monitor progress. Post treatment, teachers responded to a questionnaire that asked what the students' goals were, what information they used to evaluate if students were on track to reach their goals, how they decided if a changed need to be made, and the number of instructional modifications made. Results showed that teachers in the CBM group defined goals in more behavioral terms, made more instructional modifications, and were more objective in their instructional decision- making. Interestingly, teachers who used CBM progress-monitoring data were more conservative in estimating progress toward goals and less optimistic about goal attainment than teachers who did not. While students' in the CBM group performed higher, many did not reach their goals. The authors state that this suggests that conservative estimates may be better for instructional planning (Fuchs et al., 1989).

In 1991, Fuchs, Fuchs, Hamlett, and Stecker systematically examined the mediating effect of consultation on using CBM to progress monitoring math performance. Teachers who received instructional consultation with CBM described more changes in instructional strategy, tried new strategies, and motivated students more. In 1992, Fuchs, Fuchs, Hamlett, and Ferguson compared the use of a computerized expert system with CBM to guide decision-making with the use of CBM progress monitoring only, and found some additional gains for students in the expert system group. Recently, Ball (2009) conducted a study in which a group of kindergarten teachers were given feedback on students' fall, winter, and spring benchmark performance on the Dynamic Indicators of Basic Early Literacy Skills (*DIBELS*). Compared to students whose teachers did not receive feedback, students whose teachers received feedback made significantly greater gains on three of four *DIBELS* probes from fall to spring. Despite greater improvement, only about 50% of the students in the feedback group met end of the year benchmark goals, suggesting that the effect of feedback in this study was small (Ball, 2009). This may be explained by the survey results of teachers' perceptions of the usefulness of feedback; most teachers said feedback had limited to moderate influence on their instruction. This further supports that even when teachers are provided progress monitoring data and feedback, there are limited benefits if teachers are not motivated to use the information in instructional decision-making.

**Progress Monitoring at Grade versus Instructional Level**

According to Dunn and Eckert (2002), there are three things to consider when progress monitoring students: the instructional level of the student, a short-term or long-term goal, and the difficulty level of progress monitoring materials. *Best Practices in School Psychology* and Fuchs have recommended progress monitoring with material at instructional level (Shapiro, 2008; Fuchs and Oxall, 2007). Ed Shapiro in *Best Practices in School Psychology* argued that "progress monitoring at the level of individual students plays a major role in deciding when a student needs to be moved to a different level of instruction" and that "goals [should] be established at the level of the individual student so that the outcomes of his or her progress can be measured against appropriate expectations" (Shapiro, 2008, p. 143). Shapiro in 1996 also asserted that goal level material should be material that a student can be expected to master by the end of the progress monitoring period (as cited in Conte and Hintz, 2000). Additionally, Lynn Fuchs instructed users to find the student's instructional level when using CBM (Fuchs and Oxall, 2007). One set of instructional level guidelines used today was developed by Fuchs, Fuchs, and Deno in 1982, according to their research that indicated students in grades 1 and 2 should read 40-60 words per minute and students in grades 3 through 5 should read 70-100 words per minute to benefit from classroom instruction (as cited Sibley, Biwer, Hesch, 2001).

On the other hand, progress monitoring at grade level, even for struggling readers, has some intuitive appeal, particularly for keeping expectations high since grade level materials are more difficult than instructional level materials. With grade level probes, there may be more room to grow, goals can be set that are relative to grade level benchmarks, and student performance may be better compared to same grade peers. As suggested by the research reviewed above, high expectations and ambitious goal setting may encourage educators to more quickly adjust instructional practices and lead to better achievement. In addition, the larger gap between student performance and the grade level expectation may stress the student's need for support and motivate teachers to change instruction sooner. However, if a goal is too difficult to obtain, an intervention will more likely appear unsuccessful; on the other hand, if it is too easy to obtain, an intervention will more likely appear successful (Fuchs et al., 1985).

Progress monitoring at grade level is now the method recommended by the National RTI Center and several leading researchers, including Lynn and Doug Fuchs (Fuchs and Fuchs, 2007). Yet, at this stage, comparison of the impact of progress monitoring at grade level versus instructional level has not occurred in a highly structured study. Progress monitoring at grade level can begin to be justified if using grade level assessment tools has comparable predictive validity and sensitivity to growth and leads to similar or better student outcomes. If it is the case that assessment at the student's instructional level is of "critical importance," then the difficulty of assessment materials should affect sensitivity and predictive validity (Fuchs and Deno, 1992, p. 233). The following studies that have examined these criteria for progress monitoring materials at various difficulty levels.

**Sensitivity to Growth**

In 1979, Mirkin and Deno found that material at the independent level (50 – 70 wpm) and the instructional level (33 – 60 wpm) were more sensitive to measuring student growth than material at the frustration level (10 – 30 wpm), with rates of improvement being 1.00, 1.03, and .48 respectively (as cited in Fuchs and Deno, 1992). This study suggests that when choosing progress monitoring material, the level of material should not be in the frustration range. Unlike modern progress monitoring practices, students in this study were progress monitored daily for 18 days using passages from “Power Builder Kits” (Mirkin and Deno, 1979). In 1984, Fuchs, Tinal, and Deno progress monitored 20 general and special education students with word lists from different curriculum levels, with each student receiving every level. Significant differences were found between curriculum levels, but this study used word lists rather than CBM reading passages (as cited in Shinn, Gleeson, and Tindal, 1989).

On the other hand, Shinn and colleagues (1989) found no differences in sensitivity to growth as a function of difficulty level of materials. In their study, one group of students in grades 3 to 8 were assigned to be monitored with material one level below and one level above instructional level. Another group was assigned to material two levels above and four levels above instructional level. Overall, four difficulty levels were compared for differences in rates of improvement. Results found no significant differences in average slopes for one level below instructional level (4.3 words per week) and one level above (3.7 words per week). For the group progress monitored at two and four levels above instruction, the difference in average slopes was also not statistically significant. An interesting finding was that while average slopes were not significantly different, three students' slopes were positive when progress monitored at one level above but negative when monitored at one level below. Two students had positive slopes at two levels above but negative slopes at four levels above. These results may reflect that one level below had a floor effect, whereas four levels above was too difficult. They also suggest that difficulty level of progress monitoring may result in different conclusions about student progress. It should be noted that in this study, passages from the Ginn 360 curriculum (Clymer, Blanton, Johnson, and Lapp, 1973 as cited in Shinn et al., 1989) were used, not current measures such as AIMSweb or DIBELS. The Ginn 360 did not have equivalency of passages, so the range of difficulty within a level may have been as much as the range of difficulty that across levels. Furthermore, though four levels of difficulty were compared, none of those were specifically instructional level or grade level (Shinn et al., 1989).

Hintz, Daly, and Shapiro (1998) found differences in growth rate across difficulty levels for younger students. Eighty students in grades 1 through 4, mostly in general education, were progress monitored with CBM at grade level (which in this study was assumed to be the same as instructional level for those in general education) and one year above grade level (or goal level). Positive growth slopes were found for both grade level and goal level. While there was no difference in slopes for grades 3 and 4, for grades 1 and 2, growth rates were higher on grade level probes than goal level probes. This finding suggests difficulty level may have a larger impact for younger readings, perhaps due to a floor effect. A limitation of this study is that it was unknown how the students in general education differed in their actual reading levels, since their actual reading levels were not measured. Hence, the difficulty level of each set of probes may have been different across students. Additionally, most of students (88%) were in general education, so results may be less applicable to students in Tier 2 and 3 interventions Hintz et al., 1998).

To further the study by Hintz et al. (1998), Dunn and Eckert (2002) also compared the sensitivity to growth of instructional level material (also grade level in this study) to challenging level material (one year above grade level). Participants were 20 general education students reading at second grade level. They found that when instructional level materials were used, the rate of improvement was .65 words per week. When challenging materials were used, it was .92. This difference was not statistically significant however. Also, no significant differences were found in accuracy. In the studies conducted by Hintz et al. (1998) and Dunn and Eckert (2002), the gap between instructional and challenging level was only one year. Results may have been different had the gap been multiple years. Also, only one CBM probe at each level (instead of three) was administered at each progress monitoring time, and students had a substantial amount of variability in oral reading fluency performance from time to time, suggesting possible inaccuracy in measurement (Hintz et al., 1998; Dunn and Eckert, 2002).

**Validity**

A study by Fuchs and Deno (1992) was the first to examine criterion validity of passages of various difficulty levels. The study examined two curricula, consisting of 10 levels of Ginn 720 (1976) passages and nine levels of Scott Foresman (1976) passages. The passage comprehension subtest of the Woodcock Reading Mastery Tests (WRMT) (1973) was used as the criterion measure of reading comprehension. Analysis showed no differences in criterion validity for any of the grade levels from grades 1 through 6 (average correlation .91). The study also compared student growth rates using the different levels. Although growth rates were robust and linear at all levels, there was a negative moderate to strong correlation between grade level and slope, suggesting sensitivity to growth decreases as difficulty increases. It is important to note that this study did not conduct actual progress monitoring with the probes. Instead, all Ginn and Foresman probes of varying difficulty level were administered in one 45 – 60 minute session with each student. Growth rate was calculated using a common third grade passage across students in different grade levels (Fuchs and Deno, 1992).

Wayman et al. (2007) concluded based on a literature synthesis of CBM research including the studies discussed above, that the technical adequacy (predictive validity) of CBM is sufficiently robust across passages of varying difficulty levels. However, sensitivity to growth may be affected if material is two or more levels above the student's instructional level, particularly for young readers (Wayman et al., 2007).

**Purpose of the Study**

This study seeks to contribute to existing literature in the following ways. First, the literature is in need of more recent studies. Most studies were conducted prior to 2002 with less rigorous methodological design and older assessment materials such as the Ginn and Foresman reading passages. Available now for progress monitoring are AIMSweb, DIBELS, and the newest DIBELS Next. Second, this study will specifically compare instructional to grade level ORF progress monitoring material for struggling readers in reading interventions. The past studies reviewed compared one level below and one level above instructional level, or at grade level and one level above grade level. Past studies were also often conducted with general education students, so instructional and grade level was assumed to be the same and the comparison material was one grade level above. The research questions addressed in this study are:

1. What is the sensitivity to growth of instructional compared to grade level material?

2. What is the predictive validity of instructional and grade level probes for performance on winter benchmarks?

3. To what extent does instructional and grade level progress monitoring data agree on whether or not the student making adequate progress and if instructional changes should be made to the intervention?

4. How do teacher expectations, perceptions of intervention effectiveness, and perception of student performance relative to peers differ when teachers are shown each set of data?

## Method

### Participants and Setting

Progress monitoring data were collected for 24 fourth and fifth grade students participating in Tier 2 and Tier 3 reading interventions. The reading interventions used at the schools were *Voyager* and *Language!* The students were from two elementary schools in a large school district in southern California. School 1 is a newer school with students of higher SES compared to school 2. 29% of the sample was English language learners. 54% of the sample was Hispanic/Latino and 46% was Caucasian. 48% of the sample was female and 58% was male. The students' teachers, five from school 1 and three from school 2, participated in the teacher component of the study.

### Measures

*Oral Reading Fluency (ORF).* The *DIBELS Next* ORF probes were used for progress monitoring and benchmarking. ORF has been found to correlate well with measures of general reading achievement (Dynamic Measurement Group, 2008; Shinn et al., 1992) and to impact instructional decision-making (Cusumano, 2007). Furthermore, DIBELS *Next* ORF has high alternate form reliability coefficients of .83 to .95 for a single probe and .92 to .98 when multiple probes are used (Dynamic Measurement Group, 2010).

*DIBELS Next Maze (DAZE).* The DAZE is a measure of reading comprehension used as a second winter benchmark measure in addition to ORF. Alternate form reliabilities ranged from .66 to .81 while three form alternate form reliabilities range from .85 - .93 (Dynamic Measurement Group, 2010).

*Teacher Questionnaires.* Statements pertaining to teachers' expectations, perceptions of student performance and need, and perceptions of the students' performance relative to peers were presented with a five point Likert scale (1=Strongly Disagree, 2=Disagree, 3=Not Sure, 4=Agree, 5=Strongly Agree. Two or three questions targeting each area were included for a total of 10 questions. A search of the literature did not find a questionnaire that could be adapted for this study, so original questions were generated with the help of feedback from colleagues. The questionnaire is displayed in Table 1.

**Inter-rater Reliability**

Inter-rater reliability was calculated between the two administrators of the ORF 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. The inter-rater reliability was 97%.

**Procedures**

Each student was progress monitored in the fall, once per week for six to eight weeks, with both instructional level and grade level probes. Instructional level was found by testing back to the highest level in which the student could read 70-100 wpm on third through fifth grade probes or 40-60 wpm on first and second grade probes, according to the Fuchs' guidelines (as cited in Sibley, Biwer, Hesch, 2001). On grade level probes, if a student could not read at least 10 words, testing back occurred until they could meet this standard. This is based on the DIBELS administration instructions stating that testing should be discontinued if a student cannot read at least 10 words correct on the first passage (Good and Kaminski, 2002). Three instructional level probes and three grade level probes were administered each time and the median score for each set of probes was recorded. The presentation of instructional or grade level probes first was counterbalanced to account for fatigue.

The instructional and grade level progress monitoring data were graphed separately, with the date on the horizontal axis and words read per minute on the vertical axis. A linear least squares regression line was fitted to represent the trend line. A goal line was also displayed on each graph. For instructional level progress monitoring, goals were set using Fuchs' realistic rates of improvement (ROIs) at instructional level. They are: 1.5 words per week for grade 2, 1.0 words per week for grade 3, .9 words per week for grade 4, and .5 words per week for grade 5 (Fuchs et al., 1993). The ROI multiplied by the number of weeks of instruction was added to the first data point (baseline) to obtain the goal. For grade level progress monitoring, goals were set using grade level

DIBELS middle of the year benchmarks for being not at risk (in the some risk range).

At the beginning of the winter school quarter, the progress monitoring data were shared with the students' classroom teachers. Each student was assigned an identification number so the student's name would not appear on the graphs, in order to minimize non-data based judgment. Each graph was labeled with the student's grade and level of ORF probes. Teachers were provided with explanation regarding how the goals were set at instructional level and at grade level. For each school, the instructional and grade level graphs of all the students were mixed together and randomly divided among the teachers at that school. Therefore, the graphs each teacher evaluated may or may not have belonged to his or her students. Along with the graphs, teachers were given questionnaires to complete for each graph. Lastly, students' winter benchmarks were obtained for the outcome measures. Three grade level DIBELS Next ORF probes and the DAZE were given as part of the universal winter benchmarking process.

## Results

### Sensitivity to Growth

Table 2 presents each student's instructional and grade levels, the ROI at each level, and winter benchmarks for each student. 67% of the sample had a one year gap between instructional and grad level. 29% had a two year gap, and 4% had a three year gap between instructional and grade level. The means and standard deviations of instructional and grade level ROIs were $\mu$ = .65, *SD* = 1.75 and $\mu$ = .83, *SD* = 1.25 respectively. A one way repeated measures analysis of variance was used to compare the rates of improvement (ROIs) for the two sets of probes. A one way repeated measures analysis is appropriate for the within-subject design of this study. There was one independent variable (difficulty level), and each subject participated in both instructional and grade level conditions. Unlike analysis of variance between groups, repeated measures analysis does not assume independence of observations across levels, which by nature of the design did not exist since one set of instructional level data and one set of grade level data belonged to the same student (Lomax, 2001). ROIs were obtained from the slope of least squares regression of the progress data. The difference in mean ROIs was not statistically significant, $F$ (1, 22) = .314, $p > .05$.

**Predictive Validity**

Predictive validity was examined by conducting multiple regressions, with growth rates and baseline scores as the predictor variables and winter ORF or MAZE separately as the dependent variables. The grade level model was a significant predictor of winter benchmarks, $F$ (2, 21) = 17.28, $p < .001$, and explain 62% of the variance in winter ORF scores. The baseline point was a significant predictor, $t = 5.63$, $p < .001$, while ROI was not, $t = 1.50$, $p > .05$. The corresponding model for instructional level was also significant, $F$ (2, 21) = 25.456, $p < .001$. This model explained 71% of the variance in winter ORF scores. Again, the baseline point was a significant predictor of winter ORF, $t = 7.11$, p < .001, while ROI was not, $t = 1.92$, $p > .05$. Neither the grade nor instructional level model was significant for predicting DAZE scores, $F$ (2, 21) = .686, $p > .05$ and $F$ (2, 21) = .783, $p > .05$ respectively. Winter ORF and DAZE scores were not significantly correlated in this sample, $r = .245$, $p = > .05$. Tables 3 and 4 present the regression results and correlations.

Fisher's Z test was used to examine if the predictive validity (correlation coefficients) of each model for winter ORF was statistically different. The correlation coefficients were converted to z-scores using the formula $Z_r = [\ln(1+r) - \ln(1-r]/2$. The standard error of 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 correlations for the instructional and grade level models were found to be not statistically different, $z = .663$, $p > .05$.

**Agreement Between Data Sets**

To examine differences in conclusions made by using grade or instructional level data regarding whether or not the student is making adequate progress in response to the intervention provided and if instructional changes are warranted, data based decision rules were applied. The slope of the trend line was compared to that of the goal line for each graph. If the slope of the trend line exceeds that of the goal line, the data indicates the student is making progress and the goal should be raised. If the slope of the trend line is less than that of the goal line, the data indicates the student is not making adequate progress and instructional changes should be made to the intervention (Fuchs and Fuchs, 2007). Using this trend analysis, instructional and grade level data agreed for 67% of the students and disagreed for 33% of the students. For all the students for whom the two sets of data disagreed, the instructional level data indicated the intervention was effective whereas grade level suggested instructional changes were needed. The trend analysis conclusions for each student are presented in Table 2.

The extent to which instructional and grade level data agreed on whether or not the student was making adequate progress was also examined using a second approach. Students' trend line slopes were compared to the Fuchs realistic ROIs corresponding to the grade level of DIBELS probes (.9 words per week for grade 4, .5 words per week for grade) (Fuchs et al., 1993). If instructional and grade level trend line slopes were both greater or less than the realistic ROI at each respective level, the data agreed that the student was making adequate progress according to this method. For disagreements, the cases in which instructional data showed adequate progress but grade level did not, as well as the cases in which instructional data did not show adequate progress but grade level did, were identified. The latter cases are of particular interest and indicate that for some students, grade level may pick up growth when instructional level does not. Analysis of the cases found that instructional and grade level material disagreed for five students. Of these five, four had slopes greater than the Fuchs realistic ROI at grade level but not at instructional level.

**Teacher Questionnaire**

Means and frequencies of "agree" and "strongly agree" responses were recorded for each of the 10 questions. Table 1 displays the mean responses (1=Strongly Disagree, 2=Disagree, 3=Not Sure, 4=Agree, 5=Strongly Agree) and frequency counts of "agree" or "strongly agree" responses for each set of graphs by question. About twice as many teachers agreed the current instruction and intervention appears effective when looking at instructional level graphs compared to when looking at grade level graphs. Teachers were also more likely to agree that the student needs a lot of instructional support when looking at grade level graphs. For seven of the 24 students, teachers agreed the student should be referred to special education when looking at grade level graphs whereas for only one student, the teacher agreed the student should be referred for special education when looking at the instructional level graph. When looking at grade level data, teachers were somewhat more likely to agree they are confident that the data accurately represents the student's progress and that the data shows how the student is performing relative to peers. About one-third of the teachers responded that knowledge of which student the graph belonged to would affect their interpretation of the data.

For each question, a number of observations can be made by examining differences in responses for instructional and grade level graphs belonging to the same student. There were eight out of 24 cases in which teacher agreed ("4" or "5") that the current instruction and intervention was effective when looking at the instructional level graph but not when looking at the grade level graph. For seven cases, teachers agreed intervention should be modified or increased in intensity when looking at grade level graphs but not for corresponding instructional level graphs. There were seven cases in which the teacher disagreed ("1" or "2") that the student should be referred for special education when looking at instructional level graphs but agreed when looking at grade level graphs. For one student, the responses were the opposite; the teacher agreed he/she should be referred based on instructional level but not for grade level graphs. There were eight cases in which teachers agreed instruction should target more basic skills when looking at grade level but not instructional level graphs. However, there were also five cases in which teachers agreed instruction should target more basic skills when looking at instructional level graphs but not for corresponding grade level graphs. There were five cases in which the teacher agreed the student will catch up to peers in the class when looking at grade level but not instructional level graphs, but there were four cases in which responses were the opposite. These observations are displayed in table 5.

# Discussion

## Sensitivity to Growth and Predictive Validity

This study examined if there are significant differences between instructional and grade level progress monitoring for struggling readers by comparing psychometric characteristics, instructional decisions suggested by each set of data, and the influence of each set of data on teachers' perceptions and instructional decisions. Results of this study found no statistical difference in overall sensitivity to growth or predictive validity for grade and instructional level DIBELS Next ORF progress monitoring probes. Comparison of students' trend line slopes to realistic ROIs did show however, that grade level data can show progress greater than realistic ROIs when instructional level data do not. This was the case for 17% of the students, all of whom were fifth graders at fourth grade instructional level. It may be that for students who are only one level behind grade level and performing at the high end of that instructional level, grade level material may be more sensitive to growth. Shinn et al. (1989) had similarly found that three students' slopes were positive when they were progress monitored at one level above instructional level but negative when monitored at one level below. Regarding predictive validity, it was unexpected that it was not higher for grade level compared to instructional level. The hypothesis was that predictive validity for winter benchmarks would be significant for both sets of probes, but stronger for grade level, as benchmarks are always measured at grade level and therefore should be more highly correlated.

**Why Consider Progress Monitoring at Grade Level**

If grade level and instructional level progress monitoring are comparable in sensitivity to growth and predictive validity, perhaps schools need not maintain use of instructional level when grade level progress monitoring may have additional advantages. In monitoring student progress with material below grade level and comparing that progress to goals set below grade level, it is unclear what that data means. What does it mean for a fifth grade student to be making progress or not making progress at a third grade level? That student is neither a third grade student nor in a third grade classroom receiving third grade instruction. He or she is a fifth grader in a fifth grade classroom receiving primarily fifth grade instruction. As supported by this study, grade level progress monitoring may allow students' progress to be compared to more *appropriate*, grade level expectations without sacrificing psychometric accuracy. As quoted above, Shapiro stated that "progress [should] be measured against appropriate expectations" (Shapiro, 2008, p. 143). It can be argued that grade level expectations are more appropriate for a student in that grade. Again, existing research supports that higher expectations and more ambitious goal setting leads to better outcomes.

**Grade versus Instructional Level Data for Instructional Decision Making**

Using the same data based decision rules for each set of data, the same decision was made for 71% of the students regarding whether or not the student was making adequate progress and if instructional changes were needed. For a substantial proportion of students however, the two sets of data differently portrayed student progress. For 29% of the students, their instructional level data indicated the intervention was effective and goals should be raised whereas their grade level data indicated progress was not adequate and changes to the intervention should be made. These results suggest that when the two sets of data disagree, instructional level data are more likely to portray the intervention as effective whereas grade level data are more likely to show that instructional changes are needed to promote student achievement.

While questionnaire results did not clearly show that grade level data yield higher teacher expectations, they did indicate that instructional level data are more likely to portray an intervention as satisfactory whereas grade level data displayed alongside grade level goals are more likely to portray the intervention as needing modification and/or greater intensity. For 1/3 of the students, teachers agreed current instruction and intervention was effective when looking at instructional but not grade level graphs. Similarly, for about 30% of the students, the teacher agreed intervention should be modified or increased in intensity when looking at grade but not instructional level graphs. For these students, their instructional level data indicated the current intervention should continue as it is whereas grade level data indicated modifications should be made or intervention should be intensified. This finding obtained from teacher judgment based on the data parallels results obtained by using solely data based decisions. In fact, the percentage of disagreement between the two data sets was 30% for both trend line analysis and teacher judgment based on the graphs.

Data showing progress at an instructional level far below the students' grade level may mask the large extent of student that would be more apparent by comparing to grade level expectations. Grade level progress monitoring and goal setting may urge teachers to provide more intense intervention. Referring back to the existing literature on progress monitoring, just obtaining progress data and quantifying student progress does not itself lead to improved outcomes. A main purpose of progress monitoring is to guide instructional decision making, to increase the effectiveness of current instruction and intervention (Stecker et al., 2005). With expectations higher than instructional level goals, teachers may continue to make instructional adjustments and push for student progress even when the student has already surpassed his/her instructional level goal.

**Progress Monitoring for Instruction versus Special Education Eligibility**

Finally, progress monitoring to inform instruction should be differentiated from another use of progress monitoring data, for measuring response to intervention in determining special education eligibility. The former is the focus of this study. As the teacher questionnaire found, for nearly one-third of the students, teachers agreed the student should be referred for special education when looking at grade level but not instructional level graphs. There was only one case in which the responses were opposite. This suggests that using grade level data may influence teachers to be more likely to refer a student to special education, which can be considered a disadvantage of using grade level. Furthermore, because grade level goals are high for struggling readers far behind grade level, grade level data will most likely yield a trend line that is relatively flat compared to the goal line, which may make it appear that these students are all candidates for special education. Therefore, if progress monitoring at grade level, practitioners should be mindful of this tendency and use grade level data conservatively in RTI for eligibility.

**Limitations**

The small sample size of this study limits statistical power to detect differences. Results with a larger sample size may have detected a difference in sensitivity between the two sets of probes and/or greater predictive validity for grade level probes. Given the sample size of 24, an alpha level of .05, and a recommended desired power of .80 (Cohen, 1988 as cited in Lipsey, 1990), the minimum effect size this study could have detected is approximately .80 (Lipsey, 1990). An effect size of .80 is considered large (Cohen, 1988 as cited in Lipsey, 1990), which is unlikely between instructional and grade level material considering previous studies found non-significant effect sizes (Dunn and Eckert, 2002; Fuchs and Deno, 1992; Shin et al., 1989).

Another potential limitation of this study is that grade level benchmarks were not used as goals for instructional level progress monitoring. The original rational for this was that a fifth grade student at instructional level three is not a third grader, so third grade benchmarks do not have the same meaning as fifth grade benchmarks for that student. However, because goals set using ROIs are typically lower than benchmarks goals, using ROIs for instructional level goal setting makes the gap between the student's progress and the goal line smaller. In other words, it makes the expectation lower at instructional level. Visually, this makes the instructional and grade level graphs more different in terms of student progress and may have contributed to the differences in teacher responses to the questionnaire for instructional and grade level graphs.

Student identity was not shown on the graphs in this study to reduce confounding by non-data related information. However, in school-based practice, teachers typically view data knowing which student the data are for, which is likely to impact decisions. However, about 70% of responses from this sample of teachers indicated analysis of the data would not have been impacted by knowledge of which student the graph belonged to. Hence, the result of this study's questionnaire may be less applicable to schools in which teachers are not trained and encouraged to make data based decisions.

Another unexpected finding was that winter ORF was not significantly correlated with winter DAZE. This may be due to the small sample size or the low performance of many of the students. The reason cannot be explained by this data. Finally, from the questions teachers asked when completing the questionnaires, some of the questions on the questionnaire (particularly question 9) may be unclear.

**Future Directions**

Future research examining instructional versus grade level progress monitoring should be conducted with a larger sample. A larger sample would also allow researchers to examine an additional research question, whether sensitivity to growth, predictive validity, and influence on teacher expectations differ by the number of years between a student's instructional and grade level. It may be that grade level is more practical for students who are only one level behind, but less practical for students more than two levels behind. In addition, a better teacher questionnaire should be developed to examine if teacher expectations are influenced by each data set. If the recommendation to use grade level material is to become accepted by more schools, further research on questions like the ones in this study should be conducted to evaluate which advantages and disadvantages of each set of materials are more consequential for students.

References


Ball, C. & Gettinger, M., 2009. Monitoring children's growth in early literacy Skills: Effects of Feedback on Performance and Classroom Environments. *Education and Treatment of Children, 32* (2), 189-212.

Beady, C.H., & Hansell, S. (1981). Teacher Race and Expectations for Student Achievement. *American Educational Research Journal Summer, 18* (2), 191-206. Conte, K.L. & Hintz, J.M. (2000). The Effects of Performance Feedback and Goal Setting on Oral Reading Fluency Within Curriculum-Based Measurement. *Diagnotistic, 25* (2), pp. 85 – 98.

Dunn, E. K., & Eckert, T. L. (2002). Curriculum-based measurement in reading: A comparison of similar versus challenging material. *School Psychology Quarterly, 17,* 24–46.

Dynamic Measurement Group (2008). *DIBELS 6th Edition Technical Adequacy Information (Tech. Rep. No. 6)*. Eugene, OR: Author. Available: http://dibels.org/pubs.html.

Dynamic Measurement Group (2010). *DIBELS Next Technical Manual Supplement*. Retrieved from http://dibels.org/pubs.html.

Fuchs, L.S. (2002). Best Practices in Defining Student Goals and Outcomes. In A.Thomas & J.Grimes (Eds.), *Best Practices in School Psychology IV* (pp. 553-563). Washington, DC: National Association of School Psychologists.

Fuchs, L. S., & Deno, S. L. (1992). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58,* 232–243.

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. (2007). Using CBM for Progress Monitoring. *National Center on Response to Intervention.* Retrieved from http://www.rti4success.org.

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.

Fuchs, L., Fuchs, D., & Deno, S. (1982). Developing goals and objectives for educational programs. [Teaching guide]. In Shapiro, E. S. (1996). *Academic skills problems: Direct assessment and intervention.* New York: Guilford Press.

Fuchs, L. S., Fuchs, D., & Deno, S. L. (1985). Importance of goal ambitiousness and goal mastery to student achievement. *Exceptional Children, 52, 63-7l.*

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children, 55,* 429-438.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., & Ferguson, C. (1992). Effects of expert system consultation within curriculum based measurement using a reading maze task. *Exceptional Children, 58,* 436– 450.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., & Stecker, P.M. (1991). Effects of curriculum-based measurement and 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., et al. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22* (1), 27-48.

Fuchs, L.S., Fuchs, D., & Stecker, P.M. (1989). Effects of curriculum-based measurement on teachers' instructional planning. *Journal of Learning Disabilities, 22* (1), 51–59.

Fuchs, L.S. & Oxall, I., (2007). Progress Monitoring: What, Why, How, When, and Where [Powerpoint slides]. Retrieved from http://www.studentprogress.org/weblibrary.asp#cbm_class.

Fuchs, L.S. Tindal. G., & Deno, S.L. (1984). Methodological issues in curriculum-based reading assessment. *Diagnostique, 9,* 191-207.

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.

Hintze, J. M., & Shapiro, E. S. (1997). Curriculum-based measurement and literature based reading: Is curriculum-based measurement meeting the needs of changing reading curricula? *Journal of School Psychology, 35,* 351–375.

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.

Lomax, R.G. (2001). *Statistical Concepts: A Second Course for Education and the Behavioral Sciences.* Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Mirkin, P.K., & Deno, S.L. (1979). Formative evaluation in the classroom: An approach to improving instruction (Research Report No. 10). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Shapiro, E.S. (2008). Best Practices in Setting Progress Monitoring Goals for Academic Skill Improvement. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology V* (pp. 141 – 157). Bethesda, MD: National Association of School Psychologists.

Rosenthal, R. (2002). The Pygmalion Effect and Its Mediating Mechanisms. In Aronson, J. (Ed.), *Improving Academic Achievement: Impact of Psychological Factors on Education* (pp. 26 – 35). California: Elsevier Science (USA).

Sibley, D., Biwer, D., & Hesch, A. (2001). Establishing Curriculum-Based Measurement Oral Reading Fluency Performance Standards to Predict Success on Local and State Tests of Reading Achievement. Unpublished Data. Arlington Heights, IL: AHSD 25.

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

Shinn, M.R., Gleeson, M.M., and Tindal, G. (1989). Varying the difficulty of testing materials: Implications for curriculum-based measurement. *Journal of Special Education, 23*(2), 223-233.

Stecker, P.M. & Fuchs, L.S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research & Practice, 15* (3), 128-134.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42,* 795-819.

Tindal, G., Flick, D., & Cole, C. (1992). The effect of curriculum on inferences of reading performance and improvement. *Diagnostique, 18,* 69–84.

Zanna, M.P. et al. (1975). Pygmalion and Galatea: The Interactive Effect of Teacher and Student Expectancies. *Journal of Experimental Social Psychology, 11*, 279-287.

Table 1
Mean and Frequency of Teacher Responses by Question

| | Instructional Level | | Grade Level | |
|---|---|---|---|---|
| | Mean | Freq “4” or “5” | Mean | Freq “4” or “5” |
| 1. The current instruction / intervention appears effective for this student. | 3.04 | 9 | 2.29 | 4 |
| 2. The student needs a lot of instructional support. | 3.67 | 16 | 3.79 | 20 |
| 3. The student's current instructional (intervention) material should be increased in difficulty. | 2.17 | 4 | 2.21 | 1 |
| 4. The student will be able to catch up to others students in the class. | 2.79 | 4 | 2.71 | 4 |
| 5. The student should be referred for special education. | 2.33 | 1 | 3 | 7 |
| 6. I am confident that this is an accurate representation of the student's progress. | 3.42 | 11 | 3.67 | 16 |
| 7. This data shows how the student is performing compared to his/her peers. | 3.13 | 12 | 3.29 | 15 |
| 8. The current intervention should be modified and/or increased in intensity. | 3.25 | 14 | 3.63 | 17 |
| 9. Current instruction should be made more basic / target more basic skills. | 3.33 | 11 | 3.58 | 14 |
| 10. It would make a difference to me if I knew which student the graph belongs to. | 2.54 | 6 | 2.88 | 9 |

Table 2
*Instructional and Grade Level ROIs and Winter Benchmarks by Student*

| | | | ROI | | Adequate response using trendline analysis? | | Benchmark | |
|---|---|---|---|---|---|---|---|---|
| ID | Grade | Ins.Lev | Ins. | Grade | Ins. | Grade | ORF | MAZE |
| 1 | 4 | 2 | 3.15 | 0.88 | **Yes** | **No** | 44 | 10 |
| 2 | 4 | 3 | -1.68 | -1.79 | No | No | 67 | 8 |
| 3 | 4 | 3 | 3.01 | 1.55 | **Yes** | **No** | 70 | 7 |
| 4 | 4 | 3 | 1.57 | 2.17 | Yes | Yes | 66 | 7 |
| 5 | 4 | 3 | -1.18 | 0.62 | No | No | 90 | 11 |
| 6 | 4 | 3 | 0.32 | -2.79 | No | No | 53 | 11 |
| 7 | 4 | 3 | -0.21 | -0.29 | No | No | 62 | 1 |
| 8 | 5 | 4 | -3.29 | 0.87 | No | No | 60 | 7 |
| 9 | 5 | 4 | -0.68 | 0.64 | No | No | 81 | 15 |
| 10 | 5 | 3 | 2.96 | 2.1 | **Yes** | **No** | 87 | 14 |
| 11 | 5 | 4 | 1.6 | 1.95 | **Yes** | **No** | 88 | 5 |
| 12 | 5 | 3 | 0.04 | 0.27 | No | No | 41 | 6 |
| 13 | 4 | 3 | 2.95 | 2.75 | **Yes** | **No** | 83 | 10 |
| 14 | 4 | 1 | -0.23 | -0.08 | No | No | 24 | 7 |
| 15 | 4 | 2 | 1.5 | 0.57 | No | No | 39 | 6 |
| 16 | 4 | 3 | 1.23 | 1.67 | **Yes** | **No** | 49 | 7 |
| 17 | 4 | 2 | 1.9 | 1.64 | **Yes** | **No** | 31 | 8 |
| 18 | 5 | 3 | 0.26 | 0.19 | No | No | 51 | 12 |
| 19 | 5 | 4 | -1.94 | 1.38 | No | No | 74 | 11 |
| 20 | 5 | 4 | 0.72 | 0.61 | No | No | 90 | 9 |
| 21 | 4 | 3 | -1.39 | 0.29 | No | No | 71 | 7 |
| 22 | 4 | 3 | 2.64 | 1.89 | **Yes** | **No** | 58 | 12 |
| 23 | 4 | 2 | 1.25 | 1.14 | No | No | 38 | 10 |
| 24 | 4 | 3 | 1.17 | 1.71 | Yes | Yes | 61 | 13 |

Table 3
Predictive Validity Coefficients

| Predictor | Winter ORF | | Winter DAZE | |
|---|---|---|---|---|
| | Grade level Model *B* | Instructional Level Model *B* | Grade level Model *B* | Instructional Level Model *B* |
| Baseline Point | 1.02* | 1.24* | .05 | .06 |
| ROI | 3.13 | 2.68 | .30 | .38 |
| $R^2$ | .62 | .71 | .06 | .07 |
| *F* | 17.28* | 25.46* | .69 | .78 |

* p < .001

Table 4
*Correlations*

| | Ins. Base | Ins. ROI | Grade Base. | Grade ROI | ORF | DAZE |
|---|---|---|---|---|---|---|
| Ins. Base. | 1 | | | | | |
| Ins. ROI | -.35 | 1 | | | | |
| Grade Base. | .72** | -.27 | 1 | | | |
| Grade ROI | .12 | .50* | .04 | 1 | | |
| ORF | .81** | -.07 | .76** | .23 | 1 | |
| DAZE | .18 | .123 | .22 | .13 | .245 | 1 |

* $p < .05$, ** $p < .001$
ROI = Rate of improvement; ORF = Oral Reading Fluency; DAZE = DIBELS MAZE

Table 5
*Differences in Reponses for the Same Student*

| | Agreed for Instructional but not Grade level | Agreed for Grade but not Instructional level |
|---|---|---|
| The current instruction / intervention appears effective for this student. | 8 | - |
| The current intervention should be modified and/or increased in intensity. | - | 7 |
| The student should be referred for special education. | 1 | 7 |
| Current instruction should be made more basic / target more basic skills. | 5 | 8 |
| The student will be able to catch up to others students in the class. | 4 | 5 |