Teacher Decision-Making and Within-Year Growth in Math

Jessica L. Saven

Daniel Anderson

Joseph F. T. Nese

Julie Alonzo

Gerald Tindal

University of Oregon

Note: Funds for this data set used to generate this report come from a federal grant awarded to the UO from the Institute of Education Sciences, U.S. Department of Education: Developing Middle School Mathematics Progress Monitoring Measures (R324A100026 funded from June 2010 - June 2014). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA (April 27- May 1, 2013).

Additional information can be requested by contacting Jessica Saven at jsaven@uoregon.edu.

Abstract

We descriptively examine teachers' decisions to progress monitor students in math based on fall easyCBM® benchmark quintiles. Hierarchical linear modeling (HLM) is also used to investigate the growth of students during the school year in three mathematical domains using progress monitoring measures. We model the growth of students in three mathematical domains simultaneously, allowing for a nuanced evaluation of student growth, including the examination of the correlations among growth for the three mathematical domains. Participants included 26,537 sixth grade students from 24 states. Descriptive findings illustrate that students were progress monitored with little score-based regularity. Implications for teacher decision-making and progress monitoring students in mathematics are discussed.

Teacher Decision-Making and Within-Year Growth in Math

Response to intervention (RTI) is an instructional framework aimed at systematically identifying and monitoring the progress of students performing below expectations. The RTI model has gained traction in recent years as a school-wide improvement effort (Deno et al., 2009), and was endorsed in the most recent revision of the Individuals with Disabilities Education Act (IDEA, 2004). Previous research has demonstrated that the positive effect of RTI on individual student achievement may be quite large (Burns, Appleton, & Stehouwer, 2005). Under the basic RTI model, all students are administered benchmark screening assessments, typically seasonally (i.e., fall, winter, spring; Christ, Silberglitt, Yeo, & Cormier, 2010). Students performing below some educator-defined normative cut-point are then provided an academic intervention and their progress is monitored with regular administration of curriculum-based measurement (CBM) probes. Teacher's instructional decisions can then potentially be based on a set of data that inform both (a) students' relative standing at any single point in time [benchmark screenings] and, (b) the effect of intervention practices on the growth of students performing below expectations [progress monitoring probes].

The purpose of this study is twofold. First, we use a large extant dataset from a formative assessment system designed for use within RTI to descriptively examine teachers' decisions to progress monitor students. The extant database is useful because it helps us better understand how decisions are being made in the field, rather than how they theoretically should be made. Despite the abundance of research surrounding RTI (see Burns et al., 2005; Deno et al., 2009; Gersten & Dimino, 2006; Johnson, Mellard, Fuchs, & McKnight, 2006), surprisingly little has focused on *how* these decisions are actually made in the field by teachers. The validity of RTI rests upon these decisions, and research by Tindal and colleagues suggested they may not always function as might be expected (Tindal, Alonzo, Nese, & Saez, 2012). Our descriptive analyses are intended to add to the emerging literature.

The second purpose of the study is to investigate the growth of sixth grade students who were assigned progress-monitoring probes in math. In an extensive review of the literature on progress monitoring in math, Foegen, Jiban, and Deno (2007) concluded that “considerable research is needed to further establish these [progress monitoring math] measures” (p. 132). Thus, a secondary purpose of this paper is to evaluate the growth of students observed through short (16 item) progress monitoring mathematics measures. We model the growth of students in three mathematical domains simultaneously, allowing for a nuanced evaluation of student growth. Modeling the growth of students across domains also allows us to examine the degree to which growth in one mathematical domain correlates with growth in another mathematical domain. Because all analyses were conducted with an extant dataset, however, the measures that students were administered were contingent upon teachers’ decisions.

Decision Making within RTI

The majority of research on decisions made within RTI to date has focused on the utility of the measures used, rather than the actual decisions being made. For example, VanDerHeyden (2011) discusses the importance of classification accuracy analyses to determine the adequacy of potential decisions made. Using letter sounds and oral reading fluency measures, she found that many measures often did not classify students at a higher than chance rate using four separate criteria. In an earlier study VanDerHeyden and Burns (2008) investigated the technical adequacy of mathematics CBMs, concluding that “the degree to which accurate decision making can be based on various scores in mathematics assessment will depend on the evolution of a model for knowing what to measure, on what occasions or at what intervals, to accurately reflect particular goals" (p. 222). In other words, if the measurement process does not result in a set of data that accurately and reliably indicate the underlying construct, then the validity of any potential decision could be called into question.

While the validity of decisions depends, in part, upon the accuracy of the measures used, it also depends upon the decision-making framework applied. As Burns, Scholin, Kosciolek, and Livingston (2010) noted, "Evaluating the psychometric properties of assessment data has to occur within the context of the decisions being made" (p. 104). Using a single subject design, the authors empirically investigated the validity of two commonly applied decision making frameworks within RTI– dual discrepancy (D. Fuchs, Fuchs, McMaster, & Al Otaiba, 2003) and aim or goal lines (Deno, 1986). The authors used data from 30 second grade students and found that 40% of classification decisions changed based on the model applied. The validity of decisions within RTI thus depends highly on both the technical characteristics of the measures used *and* the decision-making framework applied.

Another potential source of variability in the types of decisions made, but less discussed in the literature, is the amount of training educators (teachers, principals, and district-wide RTI adoption leaders) have on the systems being implemented. Such training includes, but is not limited to; the measurement system implemented, the adopted decision-making framework, the reporting scale, school- or district-wide policies, and the resources available to provide students with interventions. In a series of papers from a conference symposium, Tindal et al. (2012) used extant data to show a surprising array of decisions being made, including students being progress-monitored on a regular basis despite performing well above the 50^{th} percentile of normative performance.

Mathematics Progress Monitoring

Relative to reading, math progress-monitoring and RTI in general is in its infancy (Gersten et al., 2012). Further, the majority of research that has been conducted in the area of math has been aimed at the primary grade levels (e.g. L. Fuchs & Fuchs, 2001; L. Fuchs, Fuchs, & Compton, 2012; Gersten et al., 2012). The focus of research on the primary grades has been largely due to the need for early intervention. However, more research is needed in the middle school grades to explore what represents typical levels of achievement, how students are progress-monitored progress over time, and how growth varies by mathematical domain – all of which are contingent upon the instruction provided to the student, which is dependent upon teacher decision-making within RTI.

In a literature review, Foegen et al. (2007) identified 32 studies related to CBM math progress monitoring, of which 6 were related to middle school mathematics, and only 1 explored student growth. More recent review by Lembke, Hampton, and Beyers (2012) suggests the situation has not greatly changed, finding only three studies reporting mathematics growth for students in grades 6, and no studies in grades 7 or above.

It seems clear that more research is needed in mathematics progress monitoring – particularly related to the growth that students make. In what follows, we present our methods for descriptively examining an extant dataset relative to teachers' decision to progress-monitor students. Given the work of Tindal et al. (2012), who also used extant data, we expect to find patterns that do not necessarily coincide with theoretical models for RTI (e.g. Deno et al., 2009). We then explore the growth of students who were progress-monitored during the school year by mathematical domain. The breakdown of growth by mathematical domain may lead to new insights into how middle school students progress over the course of a school year.

Methods

easyCBM® Assessment System

An extant sixth grade dataset from the 2011-2012 easyCBM® mathematics tests was used in this study. Data included scores on the fall benchmark assessment and 30 possible progress monitoring assessments. The progress monitoring measures consisted of 10 test forms in each of the three National Council of Teachers of Mathematics (NCTM) focal point standards: *Number and Operations (NO)*, in which students develop their understanding of and fluency with multiplication and division of fractions and decimals; *Algebra (ALG)*, in which students write, interpret, and use mathematical expressions and equations; and *Number Operations and Ratios (NOR),* in which students connect ratio and rate to multiplication and division (NCTM, 2006). All test forms contained 16 items, and were developed to be of equivalent difficulty using a Rasch model (Lai, Alonzo, & Tindal, 2009). The fall benchmark measure consisted of one 45-item form, with approximately one-third of all items aligned to each focal point. In a study by Nese et al., (2010) benchmark internal reliability estimates ranged from .73 to .83, and split-half reliabilities ranged from .84 to .89.

Participants

Our original sample consisted of 26,537 students from 111 districts in 24 states who took either the fall benchmark measure or at least one sixth grade progress-monitoring measure. Submitting demographic information by teachers was voluntary, and was thus sparse. Given this limitation, of those students with indicated demographic information: 42% were female (12% missing); 5% had a disability (67% missing); 4% were classified as English language learners (66% missing); 2% were Asian, 5% were African American, 13% were Latino, 4% were

American Indian or Alaska Native, 2% were Multi-ethnic, 0.3% were Native Hawaiian or Pacific Islander, and 33% were White (40% missing).

The analytic sample used to explore whether students' fall benchmark focal point subscores were predictors of students' progress monitoring intercepts and slopes included only students who had both a valid fall benchmark score for each focal point and at least one occasion of progress monitoring at the sixth grade level; restricting the sample to 4,563 students. This drastic drop in number of students was expected, as theoretically only students who perform below expectations receive progress monitoring probes. As mentioned previously, the optional entry of demographics led to sparse information about these students. A total of 61 districts from 11 different states were represented in the analysis. The sample was approximately 38% female (19.2% missing), and 7% of students had a disability (72% missing). The ethnic breakdown of students was as follows: 2% Asian, 7% African American, 16% Latino, 3% American Indian or Alaska Native, 3% multi-ethnic, 0.2% Native Hawaiian or Pacific Islander, 25% White, and 44% of the sample had missing ethnicity information. Students designated as English language learners made up 5% of the sample (69% missing).

Analyses

Descriptive. To examine the patterns in which sixth grade students were assigned progress monitoring measures, descriptive tables were created summarizing student information by fall benchmark quintiles based on total benchmark score using the easyCBM® reported norms (http://www.easyCBM.com). The quintiles themselves range as follows: quintile one, twentieth percentile and below (score of less than 25); quintile two, 21^{st} to 38^{th} percentile (score of 26 to 28); quintile three, 44^{th} to 60^{th} percentile (score of 29 to 32); quintile four, 61^{st} to 80^{th} percentile (score of 33 to 36); and quintile five, 81^{st} percentile and above (score of 37 to 45). Quintile two did not end at the 40th percentile and quintile three did not begin at the 41st percentile because a fall benchmark score of 28 was equivalent to the 38th percentile, and a fall benchmark score of 29 equivalent to the 44th percentile, necessitating that quintiles two and three be broken at that point.

Table 1 includes the total sample of all sixth grade students who took either the fall benchmark measure or at least one sixth grade progress-monitoring measure (n = 26,537), and displays the number of students who were progress-monitored above, below, and on grade level. The table is further broken down into two totals: students with a valid fall benchmark score, and all students who took any type of measure. Tables 2 and 3 include only students whose scores were used for the HLM analysis, and were consolidated into two areas by fall benchmark quintile: (a) focal points in which students were progress-monitored (Table 2); and (b) total number of focal points in which students were progress-monitored (Table 3).

Hierarchical linear growth model. One HLM model was used to examine student growth in mathematics progress monitoring in which all three focal points served as the univariate outcome, with time (level-1) nested within student (level-2). This approach allowed us to better control for type-1 error, provided the analysis with increased power, and allowed for correlation among slopes for all focal points. For modeling slope, the linear and quadratic time variables for each focal point were individually centered. That is, each student's first time point was coded 0 with all subsequent time points coded in the fractional months (representing the days occurring between assessments) elapsed from the first administration, regardless of when progress monitoring began.

The overall model intercept was represented by a reference measure, NO; two dummy vectors representing ALG and NOR were also entered in the model to indicate when the outcome measured ALG or NOR as opposed to NO. These dummy vectors represented the difference in the intercept for ALG or NOR from NO. Each measure type then had its own clock variable coded to represent both a linear and quadratic slope. When the dummy vectors indicated the outcome was NO, the ALG and NOR time variables were coded as 0. Similarly, when the dummy vectors indicated the outcome was NOR, the ALG and NO time variables were coded 0, and when the dummy-vectors indicated the outcome was ALG, the NO and NOR time variables were coded 0. This coding scheme resulted in a model where all terms dropped from the model except those relevant to the specific domain indicated by the dummy vectors, and the growth of students on each measure type could be fit concurrently.

All HLM analyses were conducted with the HLM 7 software (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011), and full information maximum likelihood was used for estimation, by which parameter estimates are considered to be unbiased under the MAR condition (Baraldi & Enders, 2010). .

The full unconditional model (Model 1) for the analysis was defined as

$$\left\{\begin{aligned} PMscore_{ti} = {} & \pi_{0i} + \pi_{1i}(D\_ALG_{ti}) + \pi_{2i}(D\_NOR_{ti}) + \pi_{3i}(LinearNO_{ti}) \\ & + \pi_{4i}(LinearALG_{ti}) + \pi_{5i}(LinearNOR_{ti}) + \pi_{6i}(QuadraticNO_{ti}) \\ & + \pi_{7i}(QuadraticALG_{ti}) + \pi_{8i}(QuadraticNOR_{ti}) + e_{ti} \\ \pi_{0i} &= \beta_{00} + r_{0i} \\ \pi_{1i} &= \beta_{10} + r_{1i} \\ \pi_{2i} &= \beta_{20} + r_{2i} \\ \pi_{3i} &= \beta_{30} + r_{31} \\ \pi_{4i} &= \beta_{40} + r_{4i} \\ \pi_{5i} &= \beta_{50} + r_{5i} \\ \pi_{6i} &= \beta_{60} + r_{6i} \\ \pi_{7i} &= \beta_{70} + r_{7i} \\ \pi_{8i} &= \beta_{80} + r_{8i} \end{aligned}\right.$$

where

$PMscore_{ti}$ = progress monitoring score for student i at time t,

D_ALG_{ti} *and* D_NOR_{ti} = dummy variables that represent the difference between the intercept (NO at time zero) and these respective focal points,

$LinearNO_{ti}$, $LinearALG_{ti}$, and $LinearNOR_{ti}$ = linear growth slopes for NO, ALG, and NOR respectively, and

$QuadraticNO_{ti}$, $QuadraticALG_{ti}$, and $QuadraticNOR_{ti}$ = quadratic, curvilinear growth slopes for NO, ALG, and NOR, respectively.

For the *a priori* conditional growth model, we added a series of dummy-coded vectors to distinguish students who started progress monitoring in any sixth grade focal point in the winter (*WINTSTART*) and spring (*SPRSTART*), such that students who started in the fall were the reference group. These seasonal control variables allowed us to assess whether or not students' growth in math progress monitoring score was influenced by the time of year the student began taking progress monitoring probes. These seasonal controls were not broken down by focal point, as we theorized that being progress monitored in any focal point had the potential to influence growth in other focal points, as they are all constructs within mathematics.

Additionally, we entered focal point subscores from the fall 2011 easyCBM math benchmark as predictor variables. The addition of these variables (*FallNO, FallALG, & FallNOR)*, which were all entered grand-mean centered, allowed us to assess the degree to which benchmark score focal point subscores predicted the intercept and growth of students on the same progress monitoring focal point.

Our final *a priori* conditional model (Model 3) including both covariates is shown below:

$$
\begin{cases}
PMscore_{ti} = \pi_{0i} + \pi_{1i}(D\_ALG_{ti}) + \pi_{2i}(D\_NOR_{ti}) + \pi_{3i}(LinearNO_{ti}) \\
\quad +\pi_{4i}(LinearALG_{ti}) + \pi_{5i}(LinearNOR_{ti}) + \pi_{6i}(QuadraticNO_{ti}) \\
\quad + \pi_{7i}(QuadraticALG_{ti}) + \pi_{8i}(QuadraticNOR_{ti}) + e_{ti} \\
\pi_{0i} = \beta_{00} + \beta_{01}(FallNO_i) + \beta_{02}(WINTSTART_i) + \beta_{02}(SPRSTART_i) + r_{0i} \\
\pi_{1i} = \beta_{10} + \beta_{11}(FallALG_i) + \beta_{12}(WINTSTART_i) + \beta_{12}(SPRSTART_i) + r_{1i} \\
\pi_{2i} = \beta_{20} + \beta_{21}(FallNOR_i) + \beta_{22}(WINTSTART_i) + \beta_{23}(SPRSTART_i) + r_{2i} \\
\pi_{3i} = \beta_{30} + \beta_{31}(FallNO_i) + \beta_{32}(WINTSTART_i) + \beta_{33}(SPRSTART_i) + r_{3i} \\
\pi_{4i} = \beta_{40} + \beta_{41}(FallALG_i) + \beta_{42}(WINTSTART_i) + \beta_{43}(SPRSTART_i) + r_{4i} \\
\pi_{5i} = \beta_{50} + \beta_{51}(FallNOR_i) + \beta_{52}(WINTSTART_i) + \beta_{53}(SPRSTART_i) + r_{5i} \\
\pi_{6i} = \beta_{60} + \beta_{61}(FallNO_i) + \beta_{62}(WINTSTART_i) + \beta_{63}(SPRSTART_i) + r_{6i} \\
\pi_{7i} = \beta_{70} + \beta_{71}(FallALG_i) + \beta_{72}(WINTSTART_i) + \beta_{73}(SPRSTART_i) + r_{7i} \\
\pi_{8i} = \beta_{80} + \beta_{81}(FallNOR_i) + \beta_{82}(WINTSTART_i) + \beta_{83}(SPRSTART_i) + r_{8i}
\end{cases}
$$

The r_{0i} term represented the between-student variability in the intercept, which in this case was our reference measure, fall NO. The r_{1i} and r_{2i} terms represented the between-student variability in the ALG and NOR intercepts, respectively. The r_{3i} through r_{5i} terms represented the between-student variability in linear growth rate for each focal point, while the r_{6i} through r_{8i} terms the between-student variability in the quadratic growth rate for each focal point.

Results

Descriptive Analysis. Descriptive information illustrated trends that were divergent from theory about how teachers select students for progress monitoring in some cases. The most straightforward results came from the 21,402 students who had a valid benchmark score. Of these students, 5,041 were assigned to progress monitoring. Among these 5,041 students, 92% of students were progress-monitored at least once on grade level. Less than 1% of the students were assessed at least once above grade level, and 20% were assessed below grade level at least once. About 41% of the students from benchmark quintile one (low performing students; score of less than 25) were assigned to progress monitoring at any grade level. More specifically, 31% of students from quintile one received at least one measure on grade level, less than 1% above grade level, and 13% below grade level (see Table 1). In contrast, 18% of students from quintile three (average performing students; score between 29 and 32) received progress monitoring; 18% of the quintile was progress-monitored on grade, less than 1% of the quintile above grade, and 2% of the quintile below grade. Finally, 14% of students from quintile five (high performing students, score above 36) were assigned to progress monitoring; 13% of that quintile was progress-monitored on grade level, and less than 1% of the quintile both above and below grade level. By and large, students with lower fall benchmark scores received progress-monitoring probes at a higher frequency than their higher-scoring peers.

Table 2 illustrates that 70% of students who had both a valid benchmark score and were progress-monitored received at least one measure in NO, 87% in ALG, and 49% in NOR. While the percentages vary within approximately 11 percentage points across fall benchmark quintiles for NO (66% to 77%) and ALG, (81% to 89%), the focal points show opposite trends: students in lower quintiles were progress-monitored less often in NO than those in higher quintiles, with the reverse being the true for ALG. Additionally, only 65% of students in the first quintile were progress-monitored in NOR, whereas 81% of students in quintile five were progress-monitored at least once in this area.

The students assigned progress monitoring on grade level with a valid benchmark score were also broken down by quantity of focal points they were administered (Table 3). Overall, 29% of students received progress monitoring in one focal point, 14% in two focal points 14%, and 57% in all three focal points. Students in benchmark quintile one were more often progress-monitored in all three focal points (52%) than their higher-achieving peers, and received progress monitoring in one focal point more often than in two (31% and 17%, respectively). Students in the higher quintiles showed similar trends, with quintile three progress-monitored in three focal points 57% of the time, and quintile five progress-monitored in three focal points 67% of the time. Moving upward through the quintiles, the amount of progress monitoring in one or two focal points decreases, and progress monitoring in three focal points increases.

Unconditional HLM model results. For the full unconditional model (Model 1), the fixed effects were all significant, but the random effects for the NOR intercept and NO linear and quadratic slopes were not ($p > .50$). Fixing the random effects for the NO linear and quadratic slopes led to a model in which the random effect for the NOR intercept was significant ($p < .05$). Hypothesis testing showed that this revised model fit the data significantly better than the full unconditional model, $\chi^2(17) = 114.68$, $p < .001$. However, the NO quadratic fixed effect was no longer significant ($p > .50$), and was thus removed to form the final unconditional model (see Model 2, Appendix 1).

The winter and spring start dummy vectors were then added to the model at the student level to control for the season in which students began progress monitoring. The seasonal control variables were only significant in a limited number of instances, although the overall model including these covariates explained significantly more variance than the final unconditional model, $\chi^2(16) = 146.22$, $p < .001$. The winter start variable was a non-significant predictor of the ALG intercept, the NO and ALG linear slopes, and both quadratic slopes. The spring start variable was a non-significant predictor of the NO intercept, the NO and NOR linear slopes, and the NOR quadratic slope. These variables were thus removed from the model for the sake of parsimony.

Students' focal point subscores from the fall benchmark test were then added to the model as student-level predictors to explore the magnitude of the relation between the screening instrument and students' intercept/growth in progress monitoring. The model including these subscores fit the data significantly better than the final seasonal control model, $\chi^2(8) = 2031.25$, *p* < .001. The subscores were universally significant predictors of the intercepts, but were not significant predictors of the ALG linear or quadratic slopes or the NOR quadratic slope (*p* > .05). Both variables were thus removed from the model, yielding the final conditional model (see Model 4, Appendix 1).

Final model results. Results of our full model testing are presented in tables 4 (fixed effects) and 5 (random effects). Our final model indicated that the average beginning achievement (i.e., intercept) for students in fall for NO was 11.08, with a standard deviation for the variance component of 2.00. In fall, students scored, on average, 1.97 points lower in ALG, and 0.77 points lower in NOR, with standard deviations for the variance components of 1.48 and 1.25, respectively. In fall, students gained, on average, 0.19 points per month in NO, with no significant between-student variance. On average, students gained 0.09 points per month in ALG in the fall, with a standard deviation for the variance component of 0.67; and 0.41 points per month in NOR, with a standard deviation for the variance component of 0.44. Additionally, students progress-monitored in ALG and NOR had a curvilinear aspect to their growth rates. Students progress-monitored in fall in ALG showed a very slight acceleration in growth (0.02 problems per month squared, *SD* for the variance component of 0.09), whereas students progress-monitored in NOR showed a very slight deceleration in growth rate (-0.03 problems per month squared, *SD* for the variance component of 0.05). Figure 1 displays the math progress monitoring growth in NO, ALG, and NOR as it varied when monitoring began in fall (F), winter (W), and spring (S).

Overall, the relations among the variables varied by the time the student entered progress monitoring. Students who began progress monitoring in the winter had a slightly lower NO intercept (-0.35) and a slightly higher NOR intercept (0.56) than students who began progress monitoring in the fall, but fall and winter ALG intercepts were not significantly different and were thus constrained (Figure 1). These students did not differ in their NO growth, but exhibited, on average, roughly half the linear NOR growth (-0.20 points per month) of students who began in the fall. Students who began progress monitoring in the spring had a slightly higher ALG intercept (0.56) and NOR intercept (0.91) than students who began in the fall (Figure 1). These students had a substantially lower rate of linear growth (-0.77 points per month) for ALG, but had a higher rate of acceleration in growth (0.36 points per month). Students who began progress monitoring in the spring did not differ significantly from students who began in the fall for NO or NOR.

Students' fall NO benchmark subscore was a significant predictor of their NO progress monitoring intercept. For every 1-point increase in the subscore for NO on the fall benchmark, students' initial score in NO progress monitoring increased by 0.46 points (Figure 2a). The fall benchmark subscore for NO had a significant negative association with the linear growth rate for NO in progress monitoring, with every 1-point increase in fall benchmark subscore corresponding with 0.01 fewer points gained per month. The fall benchmark subscore for ALG was also a significant predictor of the ALG progress monitoring intercept (Figure 2d). For every 1-point increase on the fall benchmark in ALG, students scored, on average, 0.19 points higher initially on the ALG progress monitoring probes. There were no significant differences in the linear or curvilinear slopes for ALG based on fall benchmark subscores. NOR fall benchmark subscores were also a significant predictor of the NOR progress monitoring intercept. For every 1-point increase in fall benchmark score for NOR, students scored, on average, 0.25 points higher on their initial NOR progress monitoring measure (Figure 2g). The NOR linear growth rate decreased slightly for every one point increase in fall benchmark subscore (0.02 problems per month). Overall, students who scored higher in each focal point for the fall benchmark scored higher initially on the respective progress monitoring measures in a given focal point, but exhibited slightly less growth per month than students with lower benchmark scores in those areas.

Correlations. The full unconditional model, discussed previously, was used to evaluate the degree of relation among growth slopes in three mathematical domains. The model suggested that the strongest correlation between linear growth rates was for NO and ALG (.57). The measures that students performed highest on initially (NO & NOR) also had the least correlated linear growth slopes (.33). The correlation between the curvilinear growth rates of NO and ALG was moderate (.68), while the correlations between the quadratic slopes of NO and NOR and ALG and NOR were weak (.07 and .23, respectively). Although the curvilinear growth rates of NO and ALG were most highly correlated, the quadratic for NO was later removed from the model due to nonsignificance.

Discussion

This study descriptively examined an extant dataset relative to teachers' decision to progress-monitor students, and explored the growth of students who were progress-monitored within the school year by three mathematical domains. The goals were (a) to illustrate trends in teachers' decisions to assign students to progress monitoring in ways that either aligned with or diverged from theoretical RTI models (e.g., Deno et al., 2009), and (b) to gain new insights into how middle school students progress in mathematics during the course of a school year.

At first glance, teacher decision-making relative to progress monitoring appears reasonable: overall, 24% of students who took the fall benchmark test were progress-monitored at least once. This percentage aligns well with recommendations from Fuchs and Fuchs (2006), who suggest that all students performing below the normative 25th percentile be progress monitored. However, of the students scoring in quintile one on the fall benchmark, only 41% of students received progress monitoring probes at any grade level, and only 31% received progress monitoring probes at a sixth grade level. This would seem to imply that many students whose performance suggested were most in need of mathematics interventions did not receive them. At the same time, nearly 14% of students performing in quintile five received progress monitoring probes, nearly all of which were administered on grade-level. Because we were limited to extant data, we can only speculate as to why these patterns emerged, but they do appear to diverge from theoretical models of RTI implementation, suggesting more research is needed.

It is possible that many students who performed in quintile one and did not receive progress monitoring did, in fact, receive an intervention. The effect of the intervention may simply have not been documented with easyCBM math probes. Teachers may have used other sources of information, including professional judgment and/or other measurement systems to gauge the effectiveness of the intervention and monitor these students' progress over time. Yet, even if we accept that not all students in our sample who received an intervention also received easyCBM progress monitoring, the overall percentages still appear low for quintile one. Ideally, the vast majority of students whose fall benchmark performance placed them in the quintile one of normative performance would receive an intervention followed by progress monitoring. It thus remains concerning that nearly two-thirds (59%) of students scoring in quintile one did not receive progress monitoring measures.

Our findings also echo those of Tindal et al. (2012), in that a surprising number of students received progress monitoring measures despite initially scoring above the 50th percentile on the fall mathematics benchmark assessment. However, we do not know whether these and other observed decisions to progress monitor students were made by teachers operating within any form of RTI framework, or by those using progress monitoring measures for other purposes. This limitation impacts the validity of generalizations made from such extant data. Future investigation of mathematics progress monitoring practices, and subsequent student growth, within schools and districts may illuminate many differences in patterns of how students are assigned to progress monitoring. Descriptive statistics illustrated that while teachers generally used progress monitoring as it was theoretically intended for quintiles two through four, teachers in some cases misused the system when assigning students from quintiles one and five to progress monitoring. Whether this is intentional or accidental is unknown.

Table 3 illustrates that the percentage of students assigned to progress monitoring in all three focal points increased as fall benchmark quintile increased. This finding was unexpected, as we assumed that students who performed lower on the fall benchmark would have more need of intervention and subsequent progress monitoring in multiple domains of math, relative to students who performed higher on the fall benchmark. While we can only speculate on why we observed this relation, it is likely that teachers were assigning students to progress monitoring for reasons other than to monitor their response to a specific intervention. The finding also perhaps points to a need for studies of RTI in which teachers are observed and/or provided incentives to document their specific decision-making practices, so we can begin to understand why these patterns occur.

A concern for the limited amount of research regarding progress monitoring of middle school students in mathematics (Foegen et al., 2007; Gersten et al., 2012; Lembke et al., 2012) reinvigorated by the observed patterns of progress monitoring practices in our data led us to examine student progress monitoring growth in three mathematical domains using HLM. The goal of this portion of the study was to explore what “typical” achievement was for a sixth grade student who received progress monitoring probes. This analysis was limited: theoretically, students assigned to progress monitoring should also receive some form of mathematics intervention. We have no way of knowing which students did, in fact, receive such interventions, which may reduce the generalizability of our findings

Low performing students were most often progress-monitored in ALG, whereas higher performing students were more often progress-monitored in NO and NOR. We are unsure of why these patterns exist; however, it merits further investigation as it is possible that students with varying mathematical skills may have significantly different needs and growth patterns. Future research should model student growth and the effects of progress monitoring based on their achievement level using methods such as dividing the sample into groups based on normative achievement (e.g., perhaps quintiles) before exploring student growth. These efforts would help to describe the patterns of growth for students receiving progress monitoring in mathematics.

As schools can vary greatly in how they implement RTI, future research should explore the growth of students by school and perhaps district, perhaps by including one or both of these additional levels to our HLM model. If information on the school or district RTI practices or policies was also obtained then difference between schools in overall student growth could be evaluated relative to these implementation efforts. Such insight would help inform RTI implementation for mathematics, as well as more accurately represent the growth of students. Our study represents an initial step in this direction.

The findings from this study were congruent with those of Tindal et al. (2012), suggesting that progress monitoring patterns in extant data diverge from theoretical models for

RTI (e.g. Deno et al., 2009). Information regarding how middle school students who are progress monitored grow during the course of a school year can enhance teacher decision-making up to a point: when differences in growth are described in terms of fractions of a point increase per month, as our model illustrated, it may not necessarily translate to information educators can use to better serve the needs of their students. While this study addressed some of the gaps in extant literature regarding sixth grade mathematics progress monitoring and subsequent student growth, additional research is necessary to better understand student progress monitoring growth in a way that is more useful for teachers.

Examining actual progress monitoring practices of educators in which specific decisions are documented and tracked over the course of a school year is necessary in order to later map them with student growth. The individuals in this extant data were assumed to be progress monitored according to RTI conventions for the growth model, but descriptive statistics clearly indicate that was not the case for all students. A system of refinement is called for in order to classify which students actually meet criteria for being progress-monitored, including documented interventions delivered to such students, and measures of student progress taken at accepted intervals.

Investigating how RTI is implemented in classrooms and schools can be very costly work. However, as evident in this study, there is great variety in teacher decision-making when it comes to assigning students to progress monitoring, as well as what mathematical domains students are progress monitored in. We need to better understand teacher decision-making in order to assist teachers in making more appropriate decisions about progress-monitoring practices and to make progress monitoring within an RTI system more effective. Without this information, modeling the growth of progress-monitored students precisely cannot be guaranteed.

References

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48,* 5-37. DOI: 10.1016/j.jsp.2009.10.001

Burns, M., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23*, 381-394. doi: 10.1177/073428290502300406

Burns, M., Scholin, S., Kosciolek, S., & Livingston, J. (2010). Reliability of decision-making frameworks for response to intervention for reading. *Journal of Psychoeducational Assessment, 28*(2), 102-114.

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based measurement of oral reading (CBM-R): An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review, 39*, 447-462.

Deno, S. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review, 15*, 358-374.

Deno, S., Reschly, A., Lembke, E., Magnusson, D., Callender, S., Windram, H., et al. (2009). Developing a school-wide progress-monitoring system. *Psychology in the Schools, 46*, 44-55. doi: 10.1002/pits.20353

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *The Journal of Special Education*, *41*, 121-139.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why and how valid is it? *Reading Research Quarterly, 41*, 93-99. doi: 10.1598/RRQ.41.1.4

Fuchs, D., Fuchs, L., McMaster, K. N., & Al Otaiba, S. (2003). 2003. In H. L. Swanson, K. R. Harris & S. Graham (Eds.), *Handbook of Learning Disabilities*. New York, NY: Guilford Press.

Fuchs, L., & Fuchs, D. (2001). Principles for the prevention and intervention of mathematics difficulties. *Learning Disabilities Research & Practice, 16*, 85-95. doi: 10.1111/0938-8982.00010

Fuchs, L., Fuchs, D., & Compton, D. (2012). The early prevention of mathematics difficulty: Its power and limitations. *Journal of Learning Disabilities, 45*(3). doi: DOI: 10.1177/0022219412442167

Gersten, R., Clarke, B., Jordan, N. C., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for the primary grades: Beginnings of a research base. *Exceptional Children, 78*(4), 423-445.

Gersten, R., & Dimino, J. (2006). RTI (response to intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly, 41*(1), 99-108.

Individuals with Disabilities Education Act, 20 U.S.C., Pub. L. No. 108-446 § 1400 *et seq.* (2004).

Johnson, E., Mellard, D. F., Fuchs, D., & McKnight, M. A. (2006). *Responsiveness to intervention: How to do it*. Lawrence, KS: National Research Center on Learning Disabilities.

Lai, C. F., Alonzo, J., & Tindal, G. (2009). *The development of K-8 progress monitoring measures in mathematics for use with the 2% and general education populations: Grade 6* (Technical Report No. 0907). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Lembke, E., Hampton, D., & Beyers, S. J. (2012). Response to intervention in mathematics: Critical elements. *Psychology in the Schools*, *49*, 257-272. doi: 10.1002/pits.21596

National Council of Teachers of Mathematics. (2006). Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence. Reston, VA: Author.

Nese, J. F. T., Lai, C. F., Anderson, D., Jamgochian, E.M., Kamata, A., Sáez, L., Park, B. J., Alonzo, J., & Tindal, G. (2010). *Technical Adequacy of the easyCBM Mathematics Measures: Grades 3-8, 2009-2010 Version* (Technical Report No. 1007). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

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

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). HLM 7: Hierarchical linear and nonlinear modeling. Chicago, IL: Scientific Software International.

Tindal, G., Alonzo, J., Nese, J. F. T., & Saez, L. (2012). *Validating progress monitoring in the context of RTI.* Paper presented at the Pacific Coast Research Conference (PCRC), Coronado, CA.

VanDerHeyden, A. (2011). Technical adequacy of response to intervention decisions *Council fo Exceptional Children, 77*(3), 335-350.

VanDerHeyden, A., & Burns, M. (2008). Examination of the utility of various measures of mathematics proficiency. *Assessment for Effective Intervention*, *33*, 215-224. doi: 10.1177/1534508407313482

Table 1

Grade Level Progress Monitored at by Fall Benchmark Quintile

Fall benchmark quintile	Total Students	Students progress monitored (#)	Students progress monitored (%)	Progress monitored on grade (%)	Progress monitored above grade (%)	Progress monitored below grade (%)
*	5135	1347	26.23	20.18	< 1%	8.22
1	5375	2006	41.04	31.33	< 1%	12.97
2	4162	1122	26.96	25.49	< 1%	4.81
3	4492	813	18.10	17.48	< 1%	1.85
4	3580	584	16.31	16.15	< 1%	0.81
5	3793	516	13.60	13.29	< 1%	0.29
Total with valid benchmark	21402	5041	23.55	21.55	< 1%	4.73
All students	26537	6378	24.00	21.28	< 1%	7.69

**These students did not take the fall benchmark*

Table 2

Focal Point Progress Monitored (PM) at by Fall Benchmark Quintile

Fall benchmark quintile	Total Students	% Students	PM in NO (%)	PM in ALG (%)	PM in NOR (%)
1	1636	35.85	66.26	89.85	65.10
2	1060	23.23	67.74	88.58	69.81
3	785	17.20	72.48	83.44	72.99
4	578	12.67	77.16	81.31	78.37
5	504	11.05	77.38	83.93	81.35
Total	4563	100	70.28	86.72	71.03

Table 3

Number of Focal Points (FP) Progress Monitored (PM) in by Fall Benchmark Quintile

Fall benchmark quintile	Total Students	% Students	PM in 1FP (%)	PM in 2FP (%)	PM in 3FP (%)
1	1636	35.85	30.75	17.30	51.96
2	1060	23.23	30.28	13.30	56.42
3	785	17.20	28.28	14.52	57.20
4	578	12.67	27.16	8.82	64.01
5	504	11.05	24.80	7.74	67.46
Total	4563	100	29.10	13.76	57.13

Table 4

Fixed Effects Estimates for all HLM Analyses

	Models			
	Full Unconditional	Final Unconditional	Seasonal Control	Final Model
Intercept (β_{00})	10.99***	11.02***	11.16***	11.08***
Level 1 (time)				
Int diff for ALG (β_{10})	-1.88***	-1.92***	-1.96***	-1.97***
Int diff for NOR (β_{20})	-0.49***	-0.51***	-0.72***	-0.77***
Linear NO (β_{30})	0.25***	0.19***	0.18***	0.19***
Linear ALG (β_{40})	0.08**	0.09**	0.08**	0.09**
Linear NOR (β_{50})	0.33***	0.32***	0.39***	0.41***
Quadratic NO (β_{60})	-0.01*			
Quadratic ALG (β_{70})	0.02***	0.02***	0.02**	0.02***
Quadratic NOR (β_{80})	-0.02***	-0.02**	-0.02***	-0.03***
Level 2 (student)				
NO Wint start ($int\ NO;\ \beta_{02}$)			-0.75***	-0.35***
ALG Spr start ($int\ ALG;\ \beta_{12}$)			0.48***	0.56***
ALG Spr start ($linear\ ALG;\ \beta_{41}$)			-0.73*	-0.77*
ALG Spr start ($quadratic\ ALG;\ \beta_{71}$)			0.33*	0.36*
NOR Wint start ($int\ NOR;\ \beta_{22}$)			0.50***	0.66***
NOR Wint Start ($linear\ NOR;\ \beta_{52}$)			-0.18***	-0.20***
NOR Spr Start ($int\ NOR;\ \beta_{23}$)			0.78***	0.91***
Fall NO score (int NO; β_{01})				0.46***
Fall NO score ($linear$ NO; β_{31})				-0.01**
Fall ALG score ($int\ ALG;\ \beta_{11}$)				0.19***
Fall NOR score ($int\ NOR;\ \beta_{21}$)				0.25***
Fall NOR score ($linear\ NOR;\ \beta_{51}$)				-0.02***

Note. Reference group for Wint start and Spr start is students who started progress monitoring in the Fall.
*p < .05, **p < .01, ***p < .001.

Table 5

Random Effects Estimates for all HLM Analyses

	Models			
	Full Unconditional	Final Unconditional	Seasonal Control	Final Model
Between-student NO (r_0)	6.07***	5.63***	5.49***	4.02***
Between-student ALG (r_1)	1.39**	1.58***	1.56***	2.20***
Between-student NOR (r_2)	1.65	1.42**	1.31**	1.56**
Between-student linear NO (r_3)	0.14			
Between-student linear ALG (r_4)	0.43***	0.40***	0.40***	0.44***
Between-student linear NOR (r_5)	0.19*	0.20**	0.19***	0.19*
Between-student quadratic NO (r_6)	0.00****			
Between-student quadratic ALG (r_7)	0.01***	0.01***	0.01***	0.01***
Between-student quadratic NOR (r_8)	0.00****	0.00****	0.00****	0.00****
Within student (e_{ti})	3.60	3.67	3.67	3.67

*p < .05, **p < .01, ***p < .001 ****value is less than .01 and p < .01.

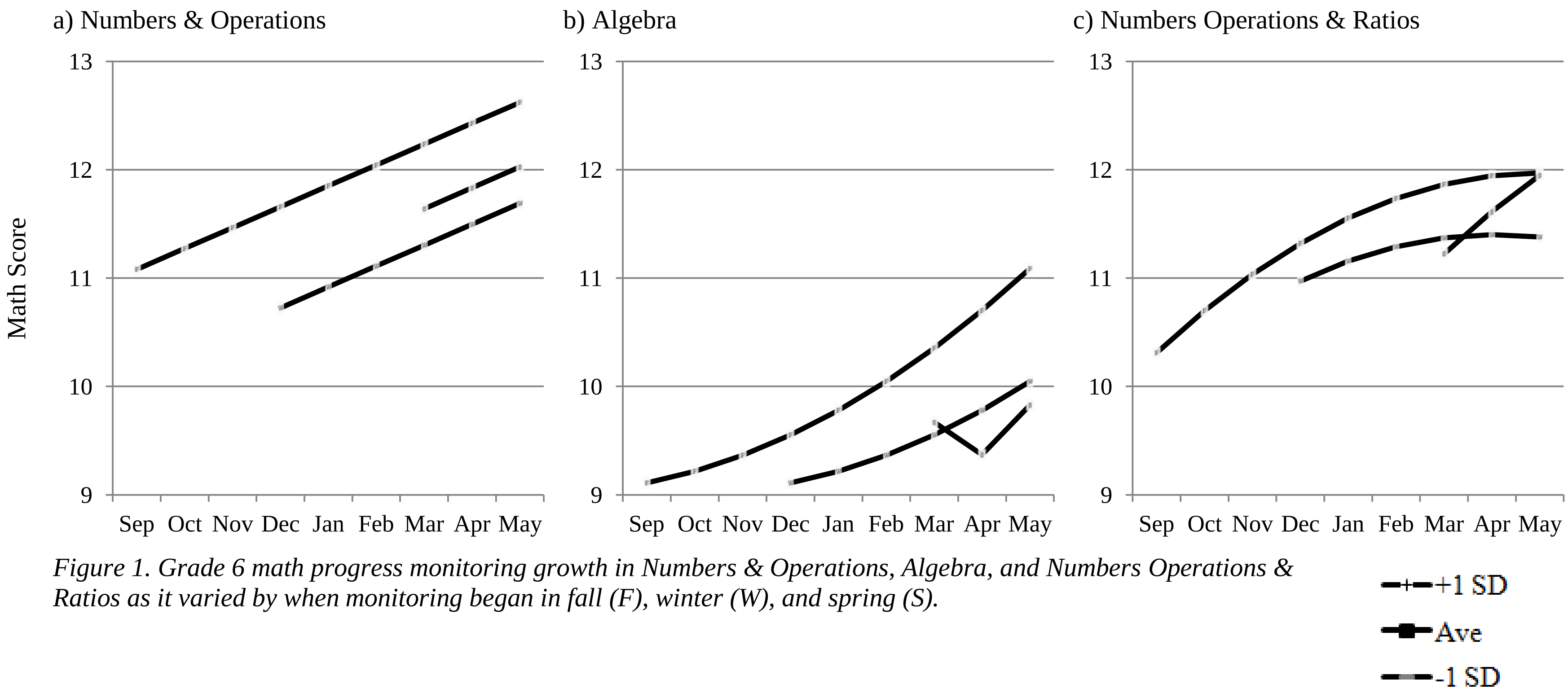

Figure 1. Grade 6 math progress monitoring growth in Numbers & Operations, Algebra, and Numbers Operations & Ratios as it varied by when monitoring began in fall (F), winter (W), and spring (S).

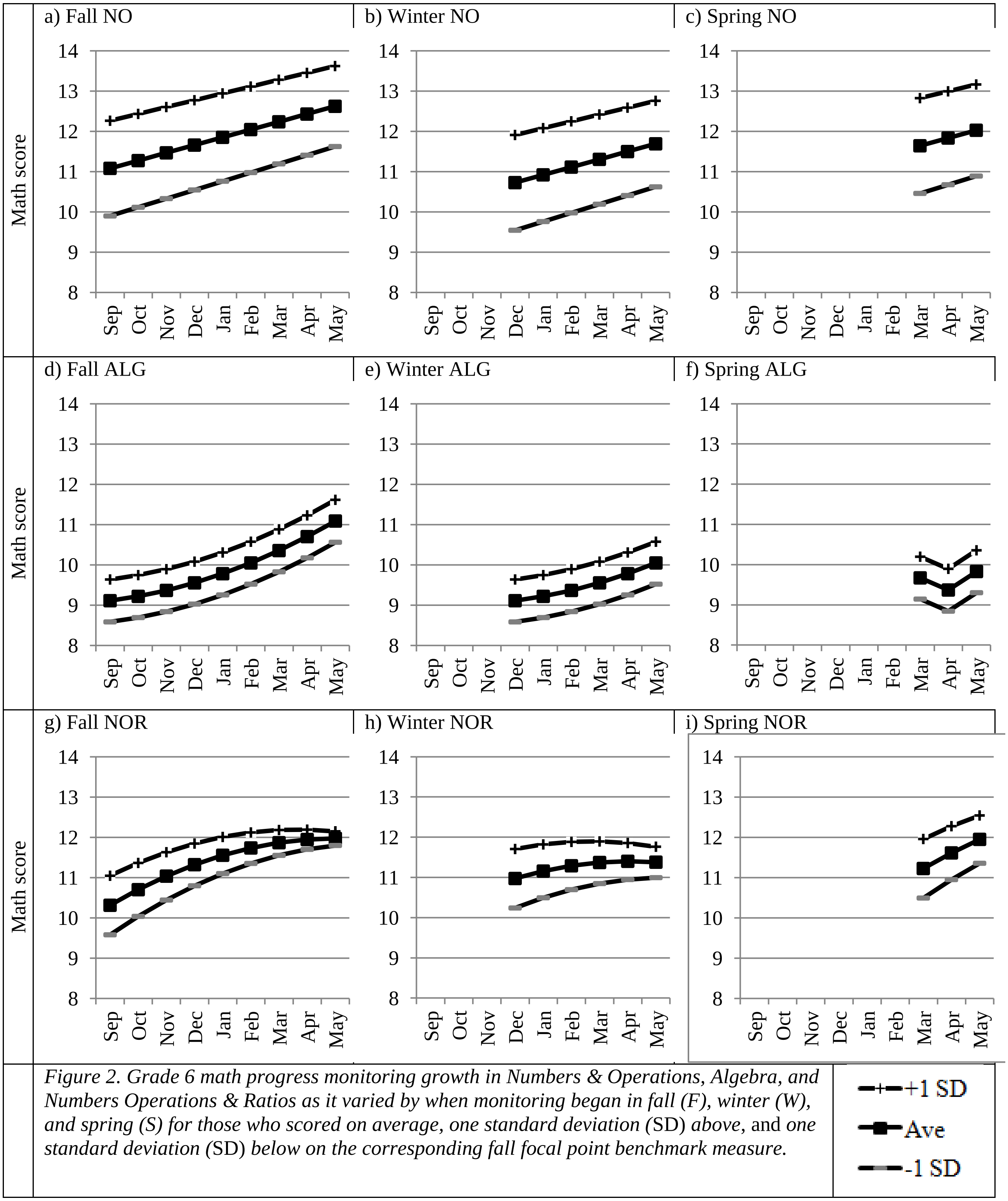

Figure 2. *Grade 6 math progress monitoring growth in Numbers & Operations, Algebra, and Numbers Operations & Ratios as it varied by when monitoring began in fall (F), winter (W), and spring (S) for those who scored on average, one standard deviation* (SD) *above*, and *one standard deviation* (SD) *below on the corresponding fall focal point benchmark measure.*

Appendix 1: Equations

The equations below provide an overview of our *a priori* unconditional and conditional models, as well as the final models developed through the model building process.

Model 1: Full *a priori* Unconditional Model

$$
\left\{
\begin{aligned}
& PMscore_{ti} = \pi_{0i} + \pi_{1i}(D\_ALG_{ti}) + \pi_{2i}(D\_NOR_{ti}) + \pi_{3i}(LinearNO_{ti}) \\
& \quad + \pi_{4i}(LinearALG_{ti}) + \pi_{5i}(LinearNOR_{ti}) + \pi_{6i}(QuadraticNO_{ti}) \\
& \quad + \pi_{7i}(QuadraticALG_{ti}) + \pi_{8i}(QuadraticNOR_{ti}) + e_{ti} \\
& \pi_{0i} = \beta_{00} + r_{0i} \\
& \pi_{1i} = \beta_{10} + r_{1i} \\
& \pi_{2i} = \beta_{20} + r_{2i} \\
& \pi_{3i} = \beta_{30} + r_{31} \\
& \pi_{4i} = \beta_{40} + r_{4i} \\
& \pi_{5i} = \beta_{50} + r_{5i} \\
& \pi_{6i} = \beta_{60} + r_{6i} \\
& \pi_{7i} = \beta_{70} + r_{7i} \\
& \pi_{8i} = \beta_{80} + r_{8i}
\end{aligned}
\right.
$$

Model 2: Final Unconditional Model

$$
\left\{
\begin{aligned}
& PMscore_{ti} = \pi_{0i} + \pi_{1i}(D\_ALG_{ti}) + \pi_{2i}(D\_NOR_{ti}) + \pi_{3i}(LinearNO_{ti}) \\
& \quad + \pi_{4i}(LinearALG_{ti}) + \pi_{5i}(LinearNOR_{ti}) + \pi_{6i}(QuadraticALG_{ti}) \\
& \quad + \pi_{7i}(QuadraticNOR_{ti}) + e_{ti} \\
& \pi_{0i} = \beta_{00} + r_{0i} \\
& \pi_{1i} = \beta_{10} + r_{1i} \\
& \pi_{2i} = \beta_{20} + r_{2i} \\
& \pi_{3i} = \beta_{30} \\
& \pi_{4i} = \beta_{40} + r_{4i} \\
& \pi_{5i} = \beta_{50} + r_{5i} \\
& \pi_{6i} = \beta_{60} + r_{6i} \\
& \pi_{7i} = \beta_{70} + r_{7i}
\end{aligned}
\right.
$$

Model 3: Final *a priori* Conditional Model Including Both Covariates:

$$
\begin{cases}
PMscore_{ti} = \pi_{0i} + \pi_{1i}(D\_ALG_{ti}) + \pi_{2i}(D\_NOR_{ti}) + \pi_{3i}(LinearNO_{ti}) \\
\quad +\pi_{4i}(LinearALG_{ti}) + \pi_{5i}(LinearNOR_{ti}) + \pi_{6i}(QuadraticNO_{ti}) \\
\quad + \pi_{7i}(QuadraticALG_{ti}) + \pi_{8i}(QuadraticNOR_{ti}) + e_{ti} \\
\pi_{0i} = \beta_{00} + \beta_{01}(FallNO_i) + \beta_{02}(WINTSTART_i) + \beta_{02}(SPRSTART_i) + r_{0i} \\
\pi_{1i} = \beta_{10} + \beta_{11}(FallALG_i) + \beta_{12}(WINTSTART_i) + \beta_{12}(SPRSTART_i) + r_{1i} \\
\pi_{2i} = \beta_{20} + \beta_{21}(FallNOR_i) + \beta_{22}(WINTSTART_i) + \beta_{23}(SPRSTART_i) + r_{2i} \\
\pi_{3i} = \beta_{30} + \beta_{31}(FallNO_i) + \beta_{32}(WINTSTART_i) + \beta_{33}(SPRSTART_i) + r_{3i} \\
\pi_{4i} = \beta_{40} + \beta_{41}(FallALG_i) + \beta_{42}(WINTSTART_i) + \beta_{43}(SPRSTART_i) + r_{4i} \\
\pi_{5i} = \beta_{50} + \beta_{51}(FallNOR_i) + \beta_{52}(WINTSTART_i) + \beta_{53}(SPRSTART_i) + r_{5i} \\
\pi_{6i} = \beta_{60} + \beta_{61}(FallNO_i) + \beta_{62}(WINTSTART_i) + \beta_{63}(SPRSTART_i) + r_{6i} \\
\pi_{7i} = \beta_{70} + \beta_{71}(FallALG_i) + \beta_{72}(WINTSTART_i) + \beta_{73}(SPRSTART_i) + r_{7i} \\
\pi_{8i} = \beta_{80} + \beta_{81}(FallNOR_i) + \beta_{82}(WINTSTART_i) + \beta_{83}(SPRSTART_i) + r_{8i}
\end{cases}
$$

Model 4: Final Conditional Model

$$
\begin{cases}
PMscore_{ti} = \pi_{0i} + \pi_{1i}(D\_ALG_{ti}) + \pi_{2i}(D\_NOR_{ti}) + \pi_{3i}(LinearNO_{ti}) \\
\quad +\pi_{4i}(LinearALG_{ti}) + \pi_{5i}(LinearNOR_{ti}) + \pi_{6i}(QuadraticALG_{ti}) \\
\quad + \pi_{7i}(QuadraticNOR_{ti}) + e_{ti} \\
\pi_{0i} = \beta_{00} + \beta_{01}(FallNO_i) + \beta_{02}(WINTSTART_i) + r_{0i} \\
\pi_{1i} = \beta_{10} + \beta_{11}(FallALG_i) + \beta_{12}(SPRSTART_i) r_{1i} \\
\pi_{2i} = \beta_{20} + \beta_{21}(FallNOR_i) + \beta_{22}(WINTSTART_i) + \beta_{23}(SPRSTART_i) + r_{2i} \\
\pi_{3i} = \beta_{30} + \beta_{31}(FallNO_i) \\
\pi_{4i} = \beta_{40} + \beta_{41}(SPRSTART_i) + r_{4i} \\
\pi_{5i} = \beta_{50} + \beta_{51}(FallNOR_i) + \beta_{52}(WINTSTART_i) + r_{5i} \\
\pi_{6i} = \beta_{60} + \beta_{61}(SPRSTART_i) + r_{6i} \\
\pi_{7i} = \beta_{70} + r_{7i}
\end{cases}
$$