



**British Journal of Education, Society &
Behavioural Science**
4(4): 527-540, 2014

SCIECEDOMAIN *international*
www.sciencedomain.org



Effects of Progress Monitoring on Math Performance of At-Risk Students

Richard Lambert¹, Bob Algozzine^{1*} and Jennifer Mc Gee²

¹Department of Educational Leadership, UNC Charlotte, Charlotte, NC 28223, USA.

²College of Education, Appalachian State University, Boone, NC 28608, USA.

Authors' contributions

Authors RL and BA designed the study and wrote the first draft of the manuscript. Authors RL and JMG managed the data collection and analysis. All authors read and approved the final manuscript.

Original Research Article

Received 30th September 2013
Accepted 28th November 2013
Published 14th January 2014

ABSTRACT

Aims: In this research, we evaluated the effects of progress monitoring grounded in a commercially-available tool used to customize assignments and keep track of progress in mathematics for students in elementary school.

Study Design: We used a randomized controlled trial and multilevel analysis to test the effect of the treatment on the outcome measures while nesting students within their classroom.

Place and Duration of Study: Students in three elementary schools in the Midwestern region of the United States were in the study which took place across an academic year.

Methodology: We used two-level hierarchical linear models for our analyses because of the nested nature of our data. We compared outcomes across high- and low-implementation fidelity treatment group classrooms as well as across treatment and control classrooms.

Results: We found statistically significant treatment differences for monthly growth rate and elementary school fidelity of implementation effects were documented.

Conclusion: Professionals engaged in progress monitoring use a variety of measures to track student performance and to assist in instructional decision making when data indicate a need for change. We found that the use of a computer-based individualized mathematics assignment and progress monitoring program resulted in improvements in both curriculum-

*Corresponding author: Email: rfaigozz@uncc.edu;

based and standardized assessments. The effects of using the system were greater when level of implementation (i.e., intervention fidelity) was higher. The value of progress monitoring and the importance of measuring the relationship between fidelity of implementation and achievement outcomes that we found support prior research.

Keywords: Progress monitoring; curriculum-based measurement; mathematics.

1. INTRODUCTION

Progress monitoring has been defined as “a practice that helps teachers use student performance data to continually evaluate the effectiveness of their teaching and make more informed instructional decisions” [1] (p.81); it also provides a valuable window in efforts to judge levels of implementation fidelity [2,3,4] and their effects on achievement outcomes [5,6]. In academics, progress monitoring involves measuring a student’s performance regularly (weekly or monthly), comparing actual and expected rates of learning to determine change, and adjusting instruction to meet individual needs and support continuing growth and achievement [7]. Professionals engaged in progress monitoring use a variety of measures to track student performance [8,9,10,11] and to assist in instructional decision making when data indicate a need for re-teaching or other adjustment [12,13,14,15]. Mastery measurement and curriculum-based assessment are approaches to progress-monitoring with long-standing support.

Accelerated Math (AM) is a technology-enhanced tool used to customize assignments and monitor progress in mathematics for students in grades 1–12. The goal of AM is to generate high-quality data for teachers to use in making important educational decisions. Its computer-based assessments provide time efficiency in quick administration, valid and reliable results for at-risk students, rich data for informing instruction, ready access to data in online databases, and multi-function (e.g., screening, progress monitoring, and outcome) data in single assessments¹. The Accelerated Math (AM) software creates individualized assignments aligned with national guidelines, state standards, and/or local expectations, scores student work, and generates reports on student performance [16,17,18,19,20]. Recently listed by the National Center on Response to Intervention (NCRTI) as its first mathematics mastery measurement tool, the system can be used in conjunction with an existing mathematics curriculum to replace other forms of practice and aid teachers in using progress-monitoring data to differentiate instruction. It keeps track of individual students’ daily activities on personalized assignments and tests, provides immediate feedback to students and teachers through information generated from individual or class diagnostic reports, alerts teachers when students are having difficulty with certain mathematics assignments, and gives teachers the information they need to support and adjust instruction. According to a What Work Clearinghouse report [21], the “extent of evidence for *Accelerated Math* is medium to large for math achievement” and the support includes both quasi-experimental research and randomized control trials (p. 1). For example, Spicuzza and Ysseldyke [22] reported positive effects of this curriculum-based instructional management system during an urban summer school program. In a more comprehensive study, Spicuzza,

¹Additional descriptive and technical information is available from the publisher’s website (<http://www.renlearn.com/am/> and <http://www.renlearn.com/RTI/>, last reviewed December 2009), the U. S. Department of Education What Works Clearinghouse Intervention Report (http://ies.ed.gov/ncee/wwc/pdf/wwc_accelmath_093008.pdf, last reviewed April 2009), and refereed publications by Nunnery and Ross [25], Ysseldyke and Bolt [27] and Ysseldyke and Tardrew [28].

Ysseldyke, Lemkuil, Kosciolk, Boys, and Teelucksingh [23] reported statistically significantly greater achievement gains for students who participated in AM than for their peers who did not use the progress monitoring system and the effects were evident for high-, middle-, and low-performing students. They also found that participation in AM increased the amount of feedback experienced by students at all skill levels and improved the communication of thinking and learning strategies to students by teachers. To support their generally positive outcomes, they noted that “further research and replication studies examining the effect of participation with AM for students at different skill levels are needed before conclusions can be made about differential effects of AM across skill levels” (p. 537). In a related study, Ysseldyke, Spicuzza, Kosciolk, Teelucksingh, Boys, and Lemkuil [24] reported positive outcomes for students enrolled in classrooms using AM and the greatest effects were observed for students whose teachers implemented the intervention to the greatest degree. Nunnery and Ross [25] reported the effectiveness of AM for students in grades 6-8 on state-wide assessments and Burns, Kanive and De Grande [26] reported differential effects across groups of students with different levels of mathematics problems and performance. Ysseldyke and Bolt [27] randomly assigned classrooms to treatment and control conditions. When teachers implemented the program with fidelity and “when they used the data from the system to manage and differentiate instruction, students gained significantly more than those for whom implementation was limited or nil” (p.453). Interestingly, “failure to take into account intervention integrity would have made it look like continuous progress monitoring did not enhance math results” (p. 453). Similarly, Ysseldyke and Tardrew [28] found that program effects were related to the level of fidelity with which AM was implemented and suggested that outcomes for all students improve when practices are implemented with high fidelity.

In this research, we evaluated the effects of progress monitoring grounded in the use of AM with elementary school students. We addressed the following research questions (and hypotheses):

1. To what extent was the mathematics achievement growth rate similar for students participating in AM progress monitoring intervention and their peers participating in “business-as-usual” control conditions? (H_0 =There will be no difference in growth rate for intervention and control conditions; H_a =There will be a difference in growth rate for intervention and control conditions)
2. To what extent was the mathematics achievement similar for students participating in high and low fidelity implementation of AM progress monitoring intervention? (H_0 = There will be no difference in growth rate across high and low levels of implementation; H_a =There will be a difference in growth rate across high and low levels of implementation)

Our work represented a replication and extension of previous research using students in second through fifth grade who were at risk for school failure based on ethnic and other background characteristics.

2. METHOD

2.1 Participants

Students in three elementary schools in the Midwestern region of the United States were in the study. The sample included similar numbers of male and female students, with about

40% from minority ethnic backgrounds, 76% eligible for free or reduced-price lunch, and 18% receiving services in special education placements. The distributions for gender, $X^2(1) = 0.28$, $p = .60$, minority status, $X^2(1) = 2.93$, $p = .09$, federal lunch status, $X^2(1) = 0.19$, $p = .66$ and special education placements, $X^2(1) = 0.62$, $p = .43$, were similar across randomly-assigned treatment and control groups.

2.2 Procedure

Using the random-number generator in Microsoft Excel, students ($N=763$) were assigned to 18 treatment and 18 control classrooms/teachers across second ($n=209$), third ($n=183$), fourth ($n=208$), and fifth ($n=163$) grades. All students were in schools using *Growing with Mathematics* or Houghton Mifflin Math as a core instructional program.

Growing with Mathematics is a comprehensive Pre-K-5 curriculum developed through extensive field testing and in-depth research on how students learn mathematics. *Growing with Mathematics* is grounded in teaching practices that reflect the philosophy that mathematical understanding deepens when children (a) engage in a wide variety of carefully selected learning experiences, (b) build conceptual understanding of math ideas through learning that builds on their previous mathematics experiences, and (c) become better problem solvers and develop confidence because they are encouraged to share and discuss their ideas and to explain their thinking. The practices for *Growing with Mathematics* are also grounded in the belief that when children learn with understanding, they will learn more difficult concepts in the future more easily than if they just memorized a procedure or rule. The research-based materials of Houghton-Mifflin *Math* provide a comprehensive approach to support achievement, a strong focus on skill building, problem solving, and concepts mastery for every level of learning, structured daily content and test-taking practice, a variety of instructional features to meet individual teaching styles, and continuous assessments for use in monitoring progress.

2.2.1 Intervention

AM is a curriculum-based measurement system that provides appropriate practice and immediate feedback for students and teachers based on frequent progress monitoring assessments constructed from item libraries including objectives ranging from basic math skills through calculus. After participating in teacher-directed lessons focused on their mathematics content objectives, students work independently or in cooperative learning pairs on individualized assignments that are algorithm-generated and controlled so that each student has different problems when working on the same content. When students complete their work, they scan it into a computer for scoring, performance feedback, record keeping, and reporting. Teachers use feedback provided by the program to plan individual and small group re-teaching lessons and to determine when to assign and do mastery testing as a basis for moving students forward. Program developers provided professional development for all participating teachers and periodically provided support focused on answering questions related to instructional implementations. Participating teachers made all decisions related to how to integrate AM into their classrooms. The intervention practices that we evaluated were those recommended by the program developers and were similar in form and process to those studied in previous research.

2.2.2 Instrumentation

All students participated in pretesting in December and post testing at the end of the school year (May/June). All progress monitoring assessments were completed in school-based computer labs under the supervision of teacher or assigned assistants and all students had similar levels of experience with the measures (i.e., STAR Math and CTB/McGraw-Hill *Terra Nova*). The STAR Math [16,17] was administered in treatment and control classrooms in fall, winter, and spring while the *Terra Nova* was used in the fall and spring. The Normal Curve Equivalent (NCE) scores were used as the outcomes for all analyses.

STAR Math measures skills in number concepts, computation, and math application using 24 multiple-choice items that students are expected to complete in a maximum of 3 minutes each. It provides information related to individual performance in critical math skills for use in planning and placement of students at appropriate levels of AM library content. It also provides information on progress monitoring of improvements or lack of them over time. The technology used with the test continuously adjusts to the performance levels of the students (e.g., when questions are answered correctly, more difficult items are presented and when questions are answered incorrectly, less difficult items are presented). Grade equivalents, percentile ranks, and normal curve equivalent scores are available for summative evaluations and research.

Norms for STAR Math are based on 25,800 students, stratified on the basis of geographic region, school location (urban, rural, and suburban), sex and ethnicity, who attended 256 schools in 42 states representative of the U.S. population. Correlations between scores on the STAR Math and other tests, including the California Achievement Test (CAT), Comprehensive Test of Basic Skills (CTBS), Iowa Tests of Basic Skills (ITBS), Metropolitan Achievement Tests (MAT), the Stanford Achievement Test, and several statewide measures of achievement, range from .60-.80 and reported concurrent and predictive validity scores are moderately high (i.e., median=.67-.80) across multiple studies and groups of students cf. http://www.rti4success.org/tools_charts/popups_screening/star-math_validity.php.

In addition to formative assessments, students in the participating district received regular summative standardized assessments. The *Terra Nova* [29] is a widely-used measure aligned with the National Assessment of Educational Progress (NAEP) achievement levels with norms based on more than 250,000 students in the United States, including those with disabilities. As Woodward and Brown [30] noted, the math portion of the *Terra Nova* contains select and constructed response formats, each reflecting different Item Response Theory (IRT) models. The select response items use the following three-parameter logistic model: item difficulty, item discrimination, and probability of a correct response by a low-scoring student. The constructed response items use two parameters: item discrimination and a difficulty parameter for each score point assigned to the item" (p. 156). The test requires computing using whole numbers, fractions, and decimals and solving word problems as well as completing items related to geometry and general measurement (e.g., perimeter, area) concepts and applications.

We used the *Terra Nova* as a standardized, global indicator of group comparability and progress on grade-level mathematics skills and competencies. It also provided an independent measure of growth for comparisons independent of the intervention-based STAR Math tests. We reasoned that using multiple measures, including one with nationally-representative technical characteristics albeit potentially misaligned with local curriculum

focus, provided a stronger approach than relying solely on curriculum- and program-based assessments as evidence of change.

2.3 Intervention Fidelity

Intervention fidelity is defined as the extent to which an intervention is given as conceived and planned [31]. Similar to previous research [23,24,25,26,27], intervention fidelity was not directly observed and the extent to which the intervention was delivered as intended was inferred from system-generated records of the number of times each week students completed a program-assigned practice sheet. In the context of progress monitoring and computer-based interventions, we reasoned that engagement rather than appointment was a preferred measure of intervention fidelity; that is, documenting assignment completion rather than merely documenting time spent at or not at a computer was desirable. We used the following decision rules from the program developers to identify high- and low-implementation fidelity groups:

- Any class with 75% or more of the students with an average of 75% correct or greater on all assignments received 1 point.
- - For Grade 2, any class with 75% or more of the class having .50 or more average objectives mastered per week received 1 point.
 -
 - For Grades 3 or higher, any class with 75% or more of the class having 1 or more average objectives mastered per week received 1 point.
- Classes scoring 2 points were rated HIGH.
-
- Classes scoring 0 points were rated LOW.
-
- Classes scoring 1 point were classified using additional decision rules.
- - If a Class scored 1 on the Objectives Rating and 0 on the Percent Correct Rating, the class average % correct on all assignments was used to determine the rating. Classes with an average percent correct of 75% or greater were scored HIGH.
 - If a Class scored 1 on the Percent Correct Rating and 0 on the Objectives Rating, the average objectives per week metric was examined. Classes were designated HIGH if the average objectives completed per week was above 1 in the case of grades 3+ or above .5 in the case of grade 2.

Application of these decision rules resulted in nine of the treatment classrooms being classified as low-fidelity and nine as high-fidelity implementation. Outcomes were compared across these groups as well as across treatment and control classrooms.

2.4 Design and Data Analysis

Multilevel analysis [32,33,34] was used to test the effect of the treatment on the outcome measures while nesting students within their classroom. Level 1 student level control variables included special education placement status and free or reduced-price lunch

status. Student minority status and gender were also entered into the child level models but were not retained as they did not account for any variance in the outcomes once special education status and free or reduced-price lunch status were already included. Level 2 classroom level control variables included proportion of students with special education placements, proportion of students with free or reduced-price lunch status, and class size. The percentages of the classroom composition made up of males and minority students were also tested but not retained as they did not contribute to the explanatory power of the models. The criterion of significance for all statistical tests was set at the .05 level of confidence and guidance from Cohen [35] and Thompson [36] was followed when evaluating and interpreting the practical significance of observed effects. We conducted both intent to treat analyses and analyses testing the effects of different levels of intervention fidelity in intervention classrooms.

3. RESULTS AND DISCUSSION

Means, standard deviations, and number of students for STAR Math and *Terra Nova* comparisons are in Table 1 and Table 2. In general, STAR Math scores were higher in the spring than when school began in the fall; and, they generally improved as the school year progressed, except in second grade when the performance of both groups dropped on the winter assessment. We reasoned that this variation was normal (i.e., resulting from naturally-occurring differences in general instruction across groups) and the improvements across all grades for the treatment group ($M_{\text{fall}}=44.02$, $M_{\text{winter}}=46.16$, $M_{\text{spring}}=51.16$) were larger than those for the control group ($M_{\text{fall}}=42.26$, $M_{\text{winter}}=41.08$, $M_{\text{spring}}=44.71$) supporting the expected intervention effect hypothesis. Similar variation was evident for *Terra Nova* scores with drops in performance evident across treatment and control groups in second and fifth grade. We attributed this variation to a lack of experience with standardized testing for younger students and to a possible lack of motivation for the older students; and again, the improvements across all grades for the treatment group ($M_{\text{fall}}=47.95$, $M_{\text{spring}}=49.34$, difference = +1.40) were greater than those for the control group ($M_{\text{fall}}=48.46$, $M_{\text{spring}}=46.90$, difference = -1.56) supporting the expected intervention effect hypothesis.

For the STAR Math, we used three-level hierarchical linear modeling to test hypotheses. Measurements were nested within student and students were nested within classrooms. Performance was measured on fall, winter, and spring assessments. The level-one model estimated initial status at the beginning of the school year and monthly growth rate across the year. The level-two model contained the following student level covariates to model their association with both initial status and growth rate: free or reduced-price lunch status, minority status, and special education status. The level-three model contained a variable that indicated whether a classroom was assigned to the treatment condition. The treatment and control conditions were equivalent in initial status. There was a statistically significant effect for monthly growth rate. The control group grew at a rate of .76 NCE points per month, or 6.87 NCE points across the academic year, and this rate was statistically significant. The treatment group grew at a statistically significant rate of .63 NCE points *more than the control group per month*; and this translates into a total growth rate of 16.67 NCE points across the academic year. When expressed as effect sizes, or standard deviation units, the control group change was .32 which would be considered a small effect size. The treatment group change reflected as an effect size was .79 which would be considered a large effect and represents a .47 standard deviation unit advantage in growth for the treatment group. With respect to the child level covariates, there were three effects that were statistically significant. Minority-group students began the year 4.64 points below majority-group students, students with an IEP began the year 14.73 points below students who did not have

an IEP, and students who qualified for free or reduced-price lunch status demonstrated a growth rate that was .47 points per month lower than their peers who did not qualify.

Table 1. Means, Standard Deviations, and Number of Students for STAR MATH Comparison

Grade	Group	Descriptor	Assessment		
			Fall	Winter	Spring
Second	Treatment	<i>M</i>	48.47	46.71	51.36
		<i>SD</i>	17.64	19.36	19.91
		<i>n</i>	78	67	64
	Control	<i>M</i>	44.04	41.63	47.24
		<i>SD</i>	19.01	19.78	22.72
		<i>n</i>	109	96	103
Third	Treatment	<i>M</i>	41.39	45.88	52.49
		<i>SD</i>	20.14	19.89	19.03
		<i>n</i>	114	108	112
	Control	<i>M</i>	43.21	45.07	47.79
		<i>SD</i>	20.93	18.81	22.38
		<i>n</i>	55	49	52
Fourth	Treatment	<i>M</i>	42.57	46.16	51.01
		<i>SD</i>	19.70	20.34	22.26
		<i>n</i>	83	86	81
	Control	<i>M</i>	40.77	40.32	44.00
		<i>SD</i>	19.35	18.61	23.18
		<i>n</i>	102	92	102
Fifth	Treatment	<i>M</i>	44.90	46.07	49.27
		<i>SD</i>	22.66	22.36	23.16
		<i>n</i>	81	68	80
	Control	<i>M</i>	41.02	38.21	39.53
		<i>SD</i>	18.96	20.49	22.23
		<i>n</i>	71	69	70
All	Treatment	<i>M</i>	44.02	46.16	51.16
		<i>SD</i>	20.23	20.35	20.96
		<i>n</i>	356	329	337
	Control	<i>M</i>	42.26	41.08	44.71
		<i>SD</i>	19.51	19.45	22.75
		<i>n</i>	340	308	329

When intervention fidelity was added to the STAR Math models, there was a statistically significant effect on monthly growth rate for the high-implementation group. High-implementation classrooms grew, on average, at a statistically significant rate that was 1.10 points per month greater than control classrooms. The low-implementation classrooms grew,

on average, at a rate that was .19 points per month greater than the control group classrooms and this difference was not statistically significant.

Table 2. Means, Standard Deviations and Number of Students for Terra Nova Math Comparison

Grade	Group	Descriptor	Assessment		
			Fall	Spring	Difference
Second	Treatment	<i>Mean</i>	53.34	45.58	-7.76
		<i>SD</i>	20.52	14.50	15.84
		<i>n</i>	59	59	59
	Control	<i>Mean</i>	53.26	48.48	-4.78
		<i>SD</i>	18.45	17.58	14.05
		<i>n</i>	73	73	73
Third	Treatment	<i>Mean</i>	41.76	53.32	11.56
		<i>SD</i>	18.58	20.10	17.46
		<i>n</i>	79	79	79
	Control	<i>Mean</i>	40.63	52.14	11.51
		<i>SD</i>	16.71	19.64	14.69
		<i>n</i>	35	35	35
Fourth	Treatment	<i>Mean</i>	46.16	48.18	2.02
		<i>SD</i>	19.91	24.10	12.95
		<i>n</i>	50	50	50
	Control	<i>Mean</i>	48.38	46.07	-2.32
		<i>SD</i>	18.82	21.81	14.73
		<i>n</i>	89	89	89
Fifth	Treatment	<i>Mean</i>	50.86	47.49	-3.37
		<i>SD</i>	19.00	21.07	11.26
		<i>n</i>	70	70	70
	Control	<i>Mean</i>	47.08	42.49	-4.59
		<i>SD</i>	18.17	19.33	12.27
		<i>n</i>	51	51	51
All	Treatment	<i>Mean</i>	47.95	49.34	1.40
		<i>SD</i>	19.73	19.83	16.44
		<i>n</i>	256	256	256
	Control	<i>Mean</i>	48.46	46.90	-1.56
		<i>SD</i>	18.62	19.93	14.98
		<i>n</i>	248	248	248

For the *Terra Nova*, children were assessed at two points during the year. For these analyses, two-level hierarchical linear models were used in which change between the two assessments was modeled as the outcome. The level-one model estimated student gain, and the associations between free or reduced-price lunch status and special education

status and achievement improvement represented as gain. The level-two model estimated the effect for classrooms that were randomly assigned to the treatment condition and included the following classroom level covariates: proportion of the classroom that qualified for free or reduced-price lunch status, the proportion of the classroom that had an IEP, and class size. The control group showed a small decline from fall to spring of 1.86 NCE points across the academic year and this rate was not statistically significant. The treatment group children, on average, made an additional 3.29 NCE point gain across the academic year; however, this gain was also not statistically significant. No other predictors resulted in statistically significant effects.

There was variability in the size of the changes across classrooms on the *Terra Nova* measure. Therefore we included an analysis that contrasted control classrooms with high- and low-implementation treatment classrooms. In this analysis, a similar two-level model, the control group made an average decline of 2.18 NCE points, the low-implementation treatment classrooms made an average decline of 2.15 NCE points, and the high-implementation treatment classrooms made an average gain of 8.38 NCE points which was statistically significant. These effects translate into the following effect sizes Control -1.03, Low-Implementation Treatment: -.10, High-Implementation Treatment .40. Therefore, in classrooms where the treatment was more fully implemented, there was a moderately-sized advantage for the treatment condition. In this model, class size also had a statistically significant effect in that for every additional child above the average class size, the average classroom level gain was .79 points higher.

4. DISCUSSION

Using regular assessments of academic progress to inform instruction is a hallmark of both promising and evidence-based education practices and researchers have also found that these formative scores are highly predictive of performance on summative state-wide high-stakes tests. Scores on frequently administered measures are also valuable in efforts to improve the quality of teaching through adjustment of curriculum and instructional practices and in documenting implementation fidelity and student growth over time.

In a recent study, Crawford et al. [6] found that total time and intensity or concentration of time spent by middle school students using a computer-based math program was “significantly and positively related to increased performance” (p.10). In our research, there was considerable variability among teachers in the level of progress monitoring implementation fidelity that was achieved; however, there was a statistically significant advantage for the treatment condition as evidenced by faster rates of growth on the STAR Math measures. This finding was consistent across treatment classrooms and grade levels. *Terra Nova* results were positive for the treatment condition, but only in high implementation classrooms where there was an overall moderately-sized advantage in growth rate.

4.1 Implications for Research and Practice

Several of our findings have implications for future research and the improvement of practice. Similar to other studies, there was difficulty achieving full implementation in some classrooms, there was considerable variability among teachers in the quality of the implementation that was achieved, and the elementary *Terra Nova* results were positive for the treatment condition, but only in high implementation classrooms where there was an overall moderately-sized advantage in growth rate. Questions related to optimal allocations

of teacher and student time as well as the benefits accrued within and between intervention sessions (e.g., attention to key aspects of program at different “participation” points) remain to be addressed. Similarly, documenting the relative contribution of the fidelity constructs of structure and process studied by Crawford et al. [6] to short- and long-term achievement gains will greatly inform practice. Future observational research focusing on relationships between key features of intervention fidelity (e.g., amount and concentration of time spent actively engaged in core features of the program, levels of teacher and student engagement with all aspects of the program) across multiple outcomes is clearly warranted to better understand and extend the value of technology-based interventions in general and progress monitoring grounded in AM with elementary school students at risk for academic failure in particular. In a meta-analysis, Yeo [11] found that CBM was a useful tool for predicting performance on statewide reading assessments; however, the landscape of benefits for mathematics achievement is underdeveloped. The lack of consistent findings across measures with different levels of curriculum alignment in our study suggests an area of needed research in elementary and other school settings.

4.2 Limitations

While demographic characteristics were similarly distributed across treatment and control groups as a result of randomization procedures followed in our study, the uncontrolled effects of classroom-level composition had potential unintended influences on our findings. Regardless, we believe our sample was comparable to and representative of other difficult-to-teach groups in elementary schools across the country and our effects remain general in similar settings with comparable students at risk for school failure. Similarly, while all teachers received initial and ongoing professional development support related to use of the technology-based intervention included in our study, natural, but uncontrolled, variation in their general and specific experience and effectiveness were potential limits in our work. In the end, the restrictions resulting from these limitations were judged minimal in light of the likelihood that research on comparable interventions would be similarly influenced (i.e., teacher experience and/or willingness to engage a new intervention are difficult to control in any school-based study).

5. CONCLUSION

Curriculum-based measurement helps teachers to make instructional change decisions by using assessments aligned with academic content; the process is effective and efficient and generally more sensitive to changes in performance because it provides opportunities for direct and continuous monitoring of achievement over short as well as long frames of time [12]. The processes, structures, and opportunities for change that are the foundation for reform practices, such as RTI, were “viewed by many as a new way to think about both early intervention and disability identification” and that promised much needed change in efforts to promote success of America’s most vulnerable students [37] (p.129). Unfortunately, education is filled with eras of innovative thinking and unfulfilled promises that are often the result of a failure to implement practices with fidelity sufficient to bring about change and sustain meaningful growth [38,39,40].

Clearly, “helping students log on to a computer and then walking away does not constitute appropriate implementation” and “to foster positive student outcomes, teachers must implement computer-based programs with the same level of rigor that they implement teacher-led interventions” or with the same attention to fidelity that is an expectation of

clinical trials and other implementation research [6] (p.234). Like others, we found positive outcomes associated with implementation fidelity in our study; and, the conclusion here is that simply measuring pretest to posttest improvements is potentially an insufficient base when judging the value of school-based interventions implemented in the context of authentic education environments. Finally, our work supports an underlying advantage inherent in the use of curriculum-based assessments for progress monitoring as evidenced by the more consistent effects evident for STAR Math which was by design more closely aligned to the district curriculum than the *Terra Nova*.

ACKNOWLEDGEMENTS

We would like to thank all the administrators, teachers and students who participated in this study.

COMPETING INTERESTS

We, the authors have no competing interests that could potentially bias our work or the reporting of it here.

REFERENCES

1. Safer N, Fleischman S. How student progress monitoring improves instruction. *Educational Leadership*. 2005;62(5):81-83.
2. Century J, Rudnick M, Freeman C. A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*. 2010;31:199-218.
3. Mowbray CT, Holter MC, Teague GB, Bybee D. Fidelity criteria Development, measurement, and validation. *American Journal of Evaluation*. 2003;24:315-340.
4. O'Donnell CL. Defining, conceptualizing and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*. 2008;78:33-84.
5. Zvoch K. Treatment fidelity in multisite evaluation. *American Journal of Evaluation*. 2009;30:44-61.
6. Crawford L, Carpenter DM II, Wilson MT, Schmeister M, McDonald M. Testing the relationship between fidelity of implementation and student outcomes in math. *Assessment for Effective Intervention*. 2012;37:224-235.
7. National Center on Student Progress Monitoring. How progress monitoring works; 2012. Available: <http://studentprogress.org/progresmon.asp#2>
8. Burns MK. Matching math interventions to students' skill deficits a preliminary investigation of a conceptual and procedural heuristic. *Assessment for Effective Intervention*. 2011;36:210-218.
9. Deno SL. Developments in curriculum-based measurement. *Remedial and Special Education*. 2003;37:184-192.
10. Fletcher JM, Vaughn S. Response to intervention Preventing and remediating academic difficulties. *Child Development Perspectives*. 2009;3:30-37.
11. Yeo S. Predicting performance on state achievement tests using curriculum-based measurement in reading: A multilevel meta-analysis. *Remedial and Special Education*. 2010;31:412-422.
12. Deno SL. Curriculum-based measurement the emerging alternative. *Exceptional Children*. 1985;52:219-232.

13. Shinn MR. Curriculum-based measurement Assessing special children. New York, Guilford; 1989.
14. Stecker PM. Monitoring student progress in individualized educational programs using curriculum-based measurement. Washington, DC: National Center on Student Progress Monitoring undated.
15. Fuchs LS, Fuchs D. What is scientifically-based research on progress monitoring? Washington, DC: National Center on Student Progress Monitoring; 2007. Available: [http://www.studentprogress.org/library/What is Scientificall %20Based Research.pdf](http://www.studentprogress.org/library/What_is_Scientificall_%20Based_Research.pdf)
16. WI. Renaissance Learning Inc. Accelerated Math. Wisconsin Rapids; 1999.
17. WI. Renaissance Learning Inc. STAR Math user's manual. Wisconsin Rapids; 1999.
18. WI. Renaissance Learning Inc. Comparison of the STAR Reading Computer-Adaptive Test and the Scholastic Reading Inventory Test. Wisconsin Rapids; 2000.
19. WI. Renaissance Learning Inc. STAR Math: Technical manual. Wisconsin Rapids; 2008.
20. WI. Renaissance Learning Inc. The foundation of the STAR assessments. Wisconsin Rapids; 2010. Available from <http://doc.renlearn.com/KMNet/R003957507GG2170.pdf>
21. What Works Clearinghouse. WWC Intervention Report: Middle School Math-Accelerated Math. Washington, DC: U. S. Department of Education, Institute of Education Sciences, What Works Clearinghouse; 2008. Available from: http://ies.ed.gov/ncee/wwc/pdf/wwc_accelmath_093008.pdf.
22. Spicuzza R, Ysseldyke JE. Using Accelerated Math to Enhance Instruction in a Mandated Summer School Program. Minneapolis, MN: National Center on Educational Outcomes, University of Minnesota; 1999.
23. Spicuzza R, Ysseldyke J, Lemkuil A, Kosciolk S, Boys C, Teelucksingh E. Effects of curriculum-based monitoring on classroom instruction and math achievement. *Journal of School Psychology*. 2001;39:521-542.
24. Ysseldyke JE, Spicuzza R, Kosciolk S, Boys C. Changes in mathematics achievement and classroom structure in 4th and 5th grade classrooms resulting from implementation of a learning information system. *Journal of Educational Research*. 2003;96:163-174.
25. Nunnery JA, Ross SM. The effects of the School Renaissance program on student achievement in reading and mathematics. *Research in the Schools*. 2007;14(1):40-59.
26. Burns MK, Kanive R, De Grande M. Effects of a computer-delivered math fact intervention as a supplemental intervention for math in third and fourth grades. *Remedial and Special Education*. 2012;33:184-191.
27. Ysseldyke J, Bolt DM. Effect of technology-enhanced continuous progress monitoring on math achievement. *School Psychology Review*. 2007;36:453-467.
28. Ysseldyke J, Tardrew S. Use of a progress monitoring system to enable teachers to differentiate mathematics instruction. *Journal of Applied School Psychology*. 2007;24:1-28.
29. McGraw-Hill. CTB *Terra Nova*. Monterrey, CA: CTB McGraw-Hill; 2002.
30. Woodward J, Brown C. Meeting the curricular needs of academically low-achieving students in middle grade mathematics. *The Journal of Special Education*. 2006;40:151-159.
31. Dusenbury L, Brannigan R, Falco M, Hansen WB. A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*. 2003;18:237-256.
32. Bickel R. Multilevel analysis for applied research. New York: The Guilford Press; 2007.
33. O'Connell AA, McCoach DB. Editors Multilevel modeling of educational data. Charlotte, NC: Information Age Publishing; 2008.

34. Radenbush S, Bryk A. Hierarchical linear models. 2nd ed. Thousand Oaks, CA: Sage; 2002.
35. Cohen J. Statistical power analysis for the behavioral sciences 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
36. Thompson B. Fundamentals of behavioral statistics. New York: The Guilford Press; 2006.
37. Fuchs D, Deshler DD. What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research and Practice*. 2007;22:129-136.
38. Fixsen DL, Naoom SF, Blase KA, Friedman RM, Wallace F. Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231); 2005.
39. Fixsen D, Blase K. Creating implementation support in education. Unpublished U.S. Department of Education Implementation Brief; 2011.
40. U.S. Department of Education. Prevalence and implementation fidelity of research-based prevention programs in public schools: Final report, Washington, DC: Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service; 2011.

© 2014 Lambert et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here:
<http://www.sciencedomain.org/review-history.php?iid=379&id=21&aid=3315>