School Psychology Review,
2007, Volume 36, No. 3, pp. 453–467

Effect of Technology-Enhanced Continuous Progress Monitoring on Math Achievement

Jim Ysseldyke
University of Minnesota

Daniel M. Bolt
University of Wisconsin—Madison

Abstract. We examined the extent to which use of a technology-enhanced continuous progress monitoring system would enhance the results of math instruction, examined variability in teacher implementation of the program, and compared math results in classrooms in which teachers did and did not use the system. Classrooms were randomly assigned to within-school experimental and control groups. Participating students were pre- and post-tested using two standardized, nationally normed tests of math achievement. When teachers implemented the continuous progress monitoring system as intended, 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. Failure to take into account intervention integrity would have made it look like continuous progress monitoring did not enhance math results.

School personnel are focused increasingly on improvement of the overall achievement of their students, and there is a need and a requirement (under the provisions of the No Child Left Behind Act) to find ways to implement evidence-based instructional practices. Lack of systematic, usable data on individual student performance and progress at the classroom level is a major bottleneck to improving teaching and learning. At the same time publishers are producing comprehensive technology-enhanced progress monitoring systems that provide teachers with the data they need to differentiate instruction, group students on the basis of comparable goals, and manage or adapt instruction based on student performance (Edformation, 2004; Good & Kaminski, 2002; McGraw-Hill Digital Learning, 2002; Renaissance Learning, 1998a). It is argued that such systems enable professionals to engage in data-driven instructional decision making and are effective in addressing the diversity of academic skills evidenced by the students now attending U.S. schools. School psychologists bring to schools considerable training and expertise in data-driven decision making and accountability (Ysseldyke et al., 2006), and one way in which they can apply

---

The authors acknowledge the assistance of Michael Patterson of the Research Department at Renaissance Learning with data collection and manuscript editing.

Correspondence regarding this article should be addressed to Jim Ysseldyke, School Psychology Program, University of Minnesota, 350 Elliott Hall, 75 East River Road, Minneapolis, MN 55455; E-mail: jim@umn.edu

those skills is by helping school personnel implement continuous progress monitoring systems.

When school psychologists talk about progress monitoring, they often do so in the context of discussions of curriculum-based assessment (CBA; e.g., Tucker, 1985), curriculum-based measurement (CBM; e.g., Deno, 1985; Hintze, Christ, & Keller, 2002; Shinn, 1989), curriculum-based evaluation (e.g., Howell & Morehead, 1987), problem solving (Tilly, 2002), or response to intervention (Batsche et al., 2005). Clearly, there is a focus on progress monitoring in the major applications of response to intervention provisions, although the progress monitoring is used to identify the extent to which instruction or intervention is working and to make changes when it is not (Salvia, Ysseldyke, & Bolt, 2007). Such is the case in the Heartland (Iowa) Area Education Agency model (Grimes, Kurns, & Tilley, 2006; Upah & Tilley, 2002), the Pennsylvania Instructional Support Team model (Kovaleski, 2003), the Ohio Instructionally-Based Assessment model, and the Minneapolis Problem-Solving model (Marston, Muyskens, Lau, & Canter, 2003). Within progress monitoring models, progress monitoring tools are envisioned as a methodology to be used to monitor student response to instruction or response to instructional interventions (Burns & Ysseldyke, 2005).

The alleged benefit of progress monitoring systems, whether they be periodic (given frequently, typically weekly or biweekly) like the Dynamic Indicators of Basic Early Literacy Skills (Good & Kaminsky, 2002), AIMSweb (Edformation, 2004), and Yearly Progress Pro (McGraw-Hill Digital Learning, 2002), or continuous (given frequently, often daily or hourly) like Accelerated Math™ (AM; Renaissance Learning, 1998a), is that the systems can be used to keep teachers aware of the performance and progress of every student in their class and enable them to make changes in instruction for students experiencing difficulty. In the sections that follow, we describe CBA, CBM, and progress monitoring or instructional management systems.

CBA, CBM, and Progress Monitoring Systems

CBA Systems

CBA (sometimes referred to as "mastery measurement") was defined by Tucker (1985) as a system of monitoring student progress through existing course content. Burns, MacQuarrie, and Campbell (2003) identify differences between CBA and CBM, indicating that the definition of CBA generally has been misunderstood in the research literature "to be a generic term for any technique measuring student skill in the given curriculum" (p. 4). They indicate that the architects of the original CBA model began using the term "Instructional Assessment" (Kovaleski, Tucker, & Duffy, 1995) to refer to an accuracy-based approach (e.g., focus on measuring percentage of known to unknown material in computing instructional level) first advocated by Gickling and Thompson (1985). CBA was designed to assess "instructional needs of a student based upon ongoing performance within the existing course content in order to deliver instruction as effectively as possible" (Gickling, Shane, & Croskery, 1989, pp. 344–345).

CBM Systems

CBM (sometimes referred to as "general outcomes measure") is a well-established, empirically based technology that can be used to monitor student performance across time, and that has been shown to be reliable and valid for enhancing the level of information educators need to modify individual instruction for students (Black & Wiliam, 1998; Deno, 1985; Fuchs & Fuchs, 1988). A CBM system is a "standardized methodology that specifies procedures for selecting test stimuli from students' curriculum, administering and scoring tests, summarizing the assessment information, and using the information to formulate instructional decisions in the basic skills areas" (Fuchs & Fuchs, 1988, p. 3). More specifically, the performance of students across standard tasks can then be used by teachers to monitor progress and adapt instructional programs as needed for each student individually

(Deno, 1985, 1986). The overall goal of this type of instructional system is to frequently assess ongoing work, monitor individual progress, provide informative feedback to students, adapt instruction as needed, and ultimately improve student overall performance.

CBM and CBA may be used on a frequent basis to determine specifically what children know and do not know, to design instruction that addresses skills in need of additional remediation, and to show progress in the local curriculum (VanDerHeyden & Burns, 2006).

Progress Monitoring and Instructional Management Systems

Recently, AM (Renaissance Learning, 1998a) has shown promise as a technology-enhanced progress monitoring and instructional management system. This system provides a possible solution for managing the complex set of tasks faced by educators today that are nearly impossible to do without the assistance of technology. The intervention is based on six "Renaissance principles": (a) increased time to practice essential skills; (b) match between student skill level and level of instruction; (c) direct and immediate feedback to teachers and learners; (d) personalized goal setting; (e) use of technology to process, store, and report information; and (f) universal success. We found the program attractive because it incorporated evidence-based principles of effective instruction as outlined by Ysseldyke and Christenson (Christenson & Ysseldyke, 1989; Ysseldyke & Christenson, 1987a, 1987b, 2002). The specific instructional components incorporated include those listed in Table 1.

AM is relatively straightforward to implement. Students are pretested using a 15-min computer-adaptive test (STAR Math™, Renaissance Learning, 1998b) and, based on their performance, assigned to appropriate instructional levels. The computer generates on-level practice exercises from an extensive database of items. Students respond to these exercises and score them by scanning them. Then the computer provides immediate feedback to the student and the teacher. Teachers are provided with summary information detailing the quantity and quality of math performance for each student in their class. They use this information to drive instruction, individualize instruction, adapt instruction, and group students for instruction. At face value, AM *should* work to enhance student math outcomes; it gives educational professionals the information they need and it incorporates nearly all of the evidence-based components of effective instruction.

By using the AM software, teachers can manage multiple instructional tasks (i.e., matching practice items to students' skill level, providing a continuous stream of practice items, monitoring individual and class progress, and providing immediate feedback on performance via numerous reporting features). Although much of this information is presently available at a classroom level, it is often unrealistic and unmanageable for teachers to organize without the assistance of computer technology. It is hypothesized that by using computer technology educators will be able to organize classroom level information—heretofore unmanageable—into useful individual student level programming and therefore enhance the performance of participating students. It is also thought that it will be necessary to use technology to manage information if teachers and school psychologists are to monitor the response of individual students to regular classroom instruction and to the differentiated instruction that occurs when specific interventions are put in place for individual students who evidence nonresponsiveness to instruction (Salvia et al., 2007).

Effectiveness Studies

Most research on CBM has been at the level of individual students rather than at classroom, school-wide, or district-wide levels. Fuchs, Fuchs, Hamlett, and Stecker (1997) showed that use of CBM data to demonstrate academic growth to both students and teachers led to improved academic outcomes. VanDerHeyden and Burns (2006) used school-wide CBA and CBM data to plan and deliver mathematics instruction and to examine the extent

Table 1
Components of Effective Instruction

Component	Description
Instructional match	The student's needs are assessed accurately, and instruction is matched appropriately to the results of the instructional diagnosis.
Instructional expectations	There are realistic, yet high, expectations for both the amount and accuracy of work to be completed by the student, and these are communicated clearly to the student.
Classroom environment	There is a positive, supportive classroom atmosphere, and time is used productively.
Instructional presentation	Instruction is presented in a clear and effective manner, the directions contain sufficient information for the student to understand the kinds of behaviors or skills to be demonstrated, and the student's understanding is checked.
Cognitive emphasis	Thinking skills and learning strategies for completing assignments are communicated explicitly to the student.
Motivational strategies	Effective strategies for heightening student interest and effort are used with the student.
Relevant practice	The student is given adequate opportunity to practice with appropriate materials and a high success rate.
Informed feedback	The student receives relatively immediate and specific information on his or her performance or behavior; when the student makes mistakes, correction is provided.
Academic engaged time	The student is actively engaged in responding to academic content; the teacher monitors the extent to which the student is actively engaged and redirects the student when the student is unengaged.
Adaptive instruction	The curriculum is modified within reason to accommodate the student's unique and specific instructional needs.
Progress evaluation	There is direct, frequent measurement of the student's progress toward completion of instructional objectives; data on the student's performance and progress are used to plan future instruction.
Student understanding	The student demonstrates an accurate understanding of what is to be done and how it is to be done in the classroom.

to which this could lead to improved math skills and scores on math tests. They used the Screening to Enhance Equitable Placement model (VanDerHeyden, Witt, & Naquin, 2003; Witt, Daly, & Noell, 1999) and noted significant improvements in math performance. Researchers have found consistently that the provision of challenging material at the right instructional level leads to improved educational outcomes in reading (Burns, 2002; Gickling & Rosenfield, 1995) and math (Burns, 2002; Gickling et al., 1989).

There have been a number of studies on the efficacy of AM as a continuous progress monitoring and instructional management tool. Spicuzza and Ysseldyke (1999) used AM to manage the math instruction of students enrolled in a summer school program mandated because they had failed state tests. Results indicated that students using AM showed an average gain of 5.75 normal curve equivalent units on the Northwest Achievement Levels Test, a district math achievement test, over the summer. This was six times the rate of growth they had shown over the prior school year. Spicuzza et al. (2001) found that use of AM resulted in gains for high-, middle-, and low-functioning students, and that it also increased the in-classroom incidence of behaviors known to enhance academic outcomes (e.g., academic engaged time, immediate feedback, checking for student understanding, and instructional match). Ysseldyke, Spicuzza, Kosciolek, and Boys (2003) found that students enrolled in classrooms where the technology-enhanced progress monitoring and instructional management system was used as an enhancement, consistently outperformed students using only the Everyday Math curriculum (University of Chicago School Mathematics Program, 2004). They reported that gains were consistent for high-, middle-, and low-performing students and gains were greatest when teachers implemented the program to a greater degree and with the highest fidelity of treatment. Spicuzza et al. (2003) found that use of a curriculum-based instructional management system enhanced math achievement of students in urban schools.

Ysseldyke and Tardrew (in press) reported the results of a quasi-experiment of differentiating math instruction for a group of more than 2,000 students. They reported that students in AM classrooms experienced significantly greater gains in math compared to students in control classrooms. The difference in the gains in one semester ranged from 7 percentile points in sixth grade to 14 percentile points in third and fourth grades. Gains were experienced across the achievement spectrum and consistent for low-, middle-, and high-achieving students. The gain for students whose teachers implemented the program with a high level of implementation integrity averaged 18 percentile points, about nine times the gain of control students.

Purpose of Study

The purpose of this study was to investigate the effect of teacher use of AM to monitor the progress of students in elementary and middle school classrooms and to manage their instruction based on the results of the progress monitoring. The extent to which adding the progress monitoring system made it easier for teachers to differentiate instruction and enhanced student math performance was of interest. We conducted a randomized controlled experiment to permit the strongest possible inferences about effects of the implementation of the continuous monitoring system; however, because of the inevitability of variability in implementation integrity, implementation was monitored at the student level. This strategy leads to the following research questions:

1. Is there significant improvement on standardized math tests for students in classrooms implementing the progress monitoring and instructional management system compared to control classrooms?
2. To what extent are there differences in implementation of the progress monitoring and instructional management system in elementary versus middle school classes?
3. What is the range of variability in implementation? To what extent do students whose teachers are high, moderate and low implementers of the continuous progress monitoring system differ in gains on standardized math tests?

The latter two questions were intended to investigate anticipated experimental effects in a way that would allow a more direct connection of such effects to actual use of the progress monitoring and instructional management system. Ultimately, our analytic approach mimics the simultaneous use of "intent-to-treat" (student assigned to experimental classroom) and "completer" (student actually used AM as intended) analyses frequently used in social science intervention studies in which participants frequently vary in the degree to which they implement the intended intervention.

Method

With the cooperation of Renaissance Learning staff, the principals at schools who had shown interest in the AM program (by requesting a product quote in the last 18 months), but who had not subsequently purchased the program, were contacted during the spring of 2003; those who were interested in participating in the study were sent an application. To apply, elementary schools had to have at least three teachers per grade level willing to use the assessments, and middle school teachers had to have at least two sections (or classes) of students per content area.

In exchange for the significant dedication required of potential participants, Renaissance Learning offered schools the AM and STAR Math assessments, the software libraries covering the school's range of grades (higher or lower if needed for advanced or lower functioning students and special state- or textbook-tagged libraries if available), and support materials including scanners, scan cards, guides, and learning cards. Schools were also provided with unlimited access to Renaissance Learning's technical support toll-free number, and up to five on-site visits from a Renaissance Learning AM consultant.

Applicants were interviewed to determine their willingness and ability to successfully participate in the study. In selecting the sample, an attempt was made to represent at least three different states, historically disadvantaged groups, and schools receiving Title I funding. The initial sample selected by the researchers for participation represented 8 states and school districts and included 9 schools with approximately 3,000 students in 122 classrooms.

Despite the care taken in choosing the sample, it became evident early in the study that one large, urban middle school was not able to dedicate the time and resources needed to use AM. Given the fact that the school as a whole did less than a month's worth of expected work in AM over the course of the entire school year, the decision was made to exclude the school from the treatment versus control analysis (it remains in the implementation level analysis), which lowered the sample by 22 classrooms and approximately 670 students.

The continuous progress monitoring system (AM) was implemented in 8 schools in 7 school districts in 7 states. Forty-one classrooms were considered treatment classrooms, and 39 classrooms were included in the control group. An additional 20 classes were classified as "B" classrooms. These were elementary classes directed to implement AM only during the second half of the school year, within the context of a multiple-baseline design. Students from the B classes are only included in the implementation group analysis, and not the treatment versus control analysis. In Table 2 we list the race–ethnicity and gender of the 1,880 students included in the initial treatment versus control analysis. Demographic characteristics are shown for each school and for those in the treatment and control groups. The table also shows the state in which each school is located and the math curricula used.

The reader should note that the quantity of students listed in the demographic table might not match the numbers included in our statistical analyses. This is because students sometimes were missing for pre- or post-tests, and when this was the case we excluded them from the specific analyses. For example, if a student took both the pre- and post- STAR Math Test, but missed the post-test for the Terra Nova, we retained the student in the sample, but excluded that student from the pre–post change score analysis for Terra Nova.

Teachers in the 80 classroom settings were randomly assigned to experimental and control conditions. Different assignment methods were used for middle and elementary school participants. Given that elementary teachers typically teach the same class of students, we randomly assigned the three teachers from each grade at each school to either the experimental (use the progress monitoring and instructional management system October to May) or control (no use of the progress monitoring and instructional management system) conditions. The third group of elementary school teachers (which we called the B group) was assigned to begin using AM midway through the school year, giving us a multiple-baseline design as well as greater anticipated variability across students in the amount of AM implementation over the school year. In the middle schools, in which one teacher often teaches multiple sections of the same subject, *classrooms of students* were randomly assigned to one of the two conditions: used or

Table 2
Number of Students in the Treatment (AM) and Control Groups at Each School

	Reagan		Jefferson		Washington		Clinton		Nixon		Carter		Ford		Lincoln	
	C	AM	C	AM	C	AM	C	AM	C	AM	C	AM	C	AM	C	AM
Race																
Asian				1		1	7	3	8	12					2	3
African American	12	38	47	43	37	37	102	116	22	29					98	82
Hispanic	1		1	3	37	31	7	11	33	24	83	80	85	84		
Native American								1	2	5					1	1
Caucasian	42	78	47	45			70	94	95	124	1				13	15
Unknown	58						2	1	1				4			
Gender																
Male	61	55	43	44	41	30	87	105	78	89	46	36	53	39	60	45
Female	44	61	52	48	33	39	100	121	82	103	38	44	33	45	54	56
Unknown	8						1		1	2			3			
State[a]	AL		SC		FL		MS		MI		TX		TX		NC	
Curriculum[b]	Harcourt, Silver Berdett Math		Houghton Mifflin Math Central		Houghton Mifflin Math Central		Glencoe		Multiple[c]		Sharon Wells Math		Sharon Wells Math; Harcourt Math		Glencoe; McGraw-Hill; state curriculum	

Note. C = control group; AM = Accelerated Math group.

[a] All schools received state standards-tagged AM libraries.

[b] Schools using Glencoe or Houghton Mifflin received AM libraries aligned to that textbook.

[c] Scott Foresman Middle School Math; Consumer Math; Everyday Math; Transition Math (Prentice Hall); Chicago Math.

did not use the progress monitoring and instructional management system. Participating teachers in middle schools taught at least one experimental group and one control group.

Those in the experimental condition were trained to use and expected to use the continuous progress monitoring and instructional management system as an enhancement to their regular math curriculum, whereas those in the control condition simply used their regular math curriculum. AM can be used with any curricula because teachers have the ability to only assign the AM objectives that align with what is being taught.

All students in the 80 classrooms were pretested with STAR Math, a computer-adaptive mathematics achievement test, and with the mathematics subtest of the Terra Nova (CTB/McGraw Hill, 2001) in October 2003. Teachers received training from Renaissance Learning on how to test students using the STAR Math assessment, as well as on how the results of the assessment could be used to inform placement in AM. Teachers also received detailed instructions on testing students with the Terra Nova. Each teacher received an appropriate inventory of Terra Nova testing materials early enough to have time to review directions and ask questions or request additional materials if needed. Teachers were given a weeklong window in which to test students.

Two notable events occurred during the pretesting phase. One elementary school misunderstood the testing directions and only administered the STAR Math assessment to students in their experimental classrooms. No students in the B or control classrooms were pretested with STAR Math, although all students did pretest with the Terra Nova. At a second elementary school, a few teachers decided not to test some second-grade students with the STAR Math assessment. This school has a high English as a second language population, and the teachers determined that many of the second-grade students' reading levels were below that needed to read the test. These students were given the Terra Nova pretest.

Teachers in the experimental group were asked to use data from the STAR Math pretest to make informed assignments of individual students or groups of students to specific instructional libraries (levels) and to use the instructional management system to track student performance, assign work, and inform both the student and teacher of the accuracy of the work.

The goal of AM is to serve as a tool for educators to help students master math objectives through continuous practice. Students do not sit in front of a computer and work math problems. When a teacher assigns an objective to a student, the software prints a paper copy of a practice assignment for the student to complete. The student finishes the practice assignment, which is then scanned into and scored by the software. If the student's performance is satisfactory (5 out of 6 problems correct), the program indicates that he or she is "ready to test," and the teacher is able to print the test for the student. If the student then scores well enough on the test (4 out of 5 problems correct), the objective is considered "mastered," and the program prints a practice assignment for the next objective on the student's list. If the student is struggling with an objective and needs more work, the program notifies the teacher, who can then intervene. Students can master objectives through diagnostic testing as well. This assignment is similar to a regular test, but can be printed by the teacher without the student completing any practice work, an option many teachers find useful for "testing out" advanced students. The program also includes a review feature that provides students with additional practice on previously mastered objectives, by mixing review problems in with practice assignments. The software automatically keeps track of these data, so teachers are able to use them to monitor progress.

During the course of the school year, each teacher using AM received three to five visits from a Renaissance Learning Math Consultant who guided the teachers on how to improve their use of the progress monitoring system. Each teacher also had unlimited access to technical support. At the end of the year (May), students took both STAR Math and Terra Nova as post-tests.

We were able to use the data from AM in our analyses. We kept track of the number of practice items, review items, and test items that students attempted. We also counted the percent correct they earned on the practice items, review items, test items, and the number of objectives that students mastered.

Results

Our primary concern was to examine whether the implementation of a progress monitoring system would result in greater improvement on standardized math test scores for students in experimental classrooms when compared to those in similar classrooms not given the intervention. The B classrooms that only implemented AM during the latter half of the school year were not included in this initial analysis.

In the course of this analysis, it became evident that a better understanding of the intervention could be gained through further analyses exploring the extent to which those in the experimental condition actually implemented AM. It was necessary to formulate an index of intervention integrity in order to examine the extent to which student performance, progress, and outcomes differed as a function of whether teachers used AM. This index was formulated by using the data collected from natural implementation of the program to form three groups of students based on the extent to which they actually participated in the instructional program. The groups were formed by counting the numbers of objectives the students mastered. Working on and mastering objectives is a key component of AM. Teachers use the reports provided by the program to track student performance, and teachers are taught to expect the students to accomplish four objectives per week, or ideally a year's worth of content in a school year. In addition, teachers are trained to work with students to set personalized goals (numbers of objectives to master) that are challenging and attainable. Given the centrality of mastering objectives to proper program implementation, this metric provides a good indicator of the extent to which the program is being implemented.

The overall range of objectives mastered was 0–197. The distribution of this variable across students demonstrated a strong positive skew. We chose to define implementation groups such that students with less than nine objectives mastered were considered "nonimplementers"; students with 10–36 objectives mastered were "low implementers"; and students with more than 36 objectives mastered were "high implementers." These cut points are somewhat arbitrary but were determined on the basis of an inspection of the frequency distributions and consultation with Renaissance Learning technical staff about the levels of progress necessary to accomplish a year's worth of objectives in a year as well as rates of progress considered high, moderate, and low. Use of actual student performance data provided us with several advantages. Student performance data were used to determine the extent of implementation of the program rather than student enrollment in classes taught by teachers assigned to implement the program. If we had considered all students as participants or nonparticipants based on their assignment to classrooms of teachers who were or were not in the experimental group, the fact that teachers did not implement the program with some students in the experimental group would have been masked. In addition, number of objectives mastered appears to function as an effective measure of implementation. In AM, students start at an objective level appropriately matched to their level of ability, making mastery of objectives more a matter of effort than achievement. Indeed, we observed several classrooms in which all students in the class mastered a large number of objectives, despite possessing a wide range of ability.

All students were pretested in October of the experimental year and post-tested in May. Two dependent measures were used: STAR Math (Renaissance Learning, 1998b), a computer-adaptive math test, and the math subtests of the Tera Nova (CTB/McGraw-Hill, 2001), with scores on both tests considered in terms of normal curve equivalents.

In comparing the experimental (progress monitoring and instructional management system using AM) versus control (no AM) conditions, a linear regression analysis was conducted in which post-test scores were regressed onto (a) pretest scores, (b) dummy variables related to main effects for experimental condition and school, and (c) product dummy variables that accounted for interactions between school and experimental condition. A couple of important features of the regression model should be noted. First, pretest scores were used as covariates, which ultimately allowed for the effects related to experimental condition and school to be studied in terms of *residualized change* in the post-test score. One concern with using ordinary gain scores (i.e., postscore–prescore) as the outcome in this analysis is the problem of regression to the mean. The use of residualized change scores in the context of the current analysis helps address such effects, thus permitting a more meaningful aggregation across students that differed in terms of pretest scores. In effect, residualized change scores can be conceptualized as follows:

$$Y - b_{YX}(X - M_X) - M_Y$$

where

Y = observed post-test score
b_{YX} = slope of the regression equation $Y = a + bX$
X = observed pretest score
M_X = group mean of pretest scores
M_Y = group mean of post-test scores

Use of residualized change scores allows for more effective control of pretest differences in evaluating intervention effects.

Second, by adding school effect dummy codes and school by experimental condition product variables to the regression model, our analysis also accounted for school effects related to both the control and experimental conditions. Controlling for school effects was found useful, as there are school-specific factors that lead to differential gains in achievement over the school year and thus represent an important source of variance unrelated to experimental condition. Given that the original design randomized classrooms within schools, school can be viewed as a blocking variable, thus making it natural to introduce school as a control in the analyses. Likewise, interactions between school and experimental condition control for the potential of a differential experimental effect across schools, as might be attributed to differential implementation of AM across schools. As mentioned earlier, a large proportion of the second-grade students at one elementary school were not given the STAR Math assessment during the pretesting phase because their teachers determined that the students' reading levels were too low to take the test. It should be noted that this school serves primarily Hispanic students. For this reason, a decision was made to remove the entire second grade at this school from our initial treatment versus control analysis.

The results of the regression analyses with respect to both the STAR Math and TerraNova standardized tests are reported in Table 3. The main effects of the experimental condition are evaluated by the condition (AM) coefficient, main effects related to school by the school coefficient, and interaction effects by the AM × School coefficient. A significant main effect for AM was observed for STAR Math, with an expected score change difference of approximately 5 score points (Cohen's effect size = 0.37). By contrast, a significant main effect was not found for TerraNova. The stronger effects observed for STAR Math are not unexpected because of the closer alignment of AM with this assessment method; however, the nonsignificance of effects for TerraNova is unexpected. Perhaps most important, substantial school effects were also found in this analysis, both in the form of main effects and interaction effects. In each analysis, Nixon School served as the baseline condition, both for main effects and interactions with experimental condition. For both the STAR Math and TerraNova analyses, significant variability across schools is demonstrated both in terms of main effects as well as interactions, as several coefficients of each type are significant in both analyses. One explanation for these effects may be variability in imple-

Table 3
Regression Analysis for Evaluation of Accelerated Math Effects, Experimental versus Control Students, Controlling for School Effects

	STAR Math			TerraNova		
Variable	*b*	*SEb*	*t*	*b*	*SEb*	*t*
Intercept	21.24	1.44	14.79*	16.09	1.38	11.66*
Condition (AM)	5.37	1.70	3.15*	0.83	1.44	0.58
Pretest score	0.73	0.02	34.55*	0.72	0.02	38.46*
Jefferson	1.83	2.03	0.90	−0.78	1.72	−0.46
Washington	−8.23	2.29	−3.59*	−7.76	1.98	−3.92*
Reagan	−9.56	2.03	−4.70*	−1.15	1.62	−0.71
Carter	−6.07	2.38	−2.55*	−3.72	2.16	−1.72
Ford	−2.01	2.21	−0.91	−0.41	1.78	−0.23
Lincoln	−8.07	1.98	−4.07*	1.18	1.62	0.73
Clinton	−7.04	1.59	−4.42*	−6.67	1.37	−4.88*
AM × Jefferson	−0.01	2.96	−0.00	0.44	2.53	0.17
AM × Washington	−3.04	3.25	−0.93	−10.83	2.80	−3.86*
AM × Reagan	−4.27	2.82	−1.52	5.74	2.33	2.46*
AM × Carter	−10.48	3.44	−3.05*	−0.27	3.04	−0.09
AM × Ford	—[a]	—[a]	—[a]	1.68	2.58	0.65
AM × Lincoln	−2.31	2.81	−0.82	−2.06	2.34	−0.88
AM × Clinton	−3.72	2.37	−1.57	−3.64	2.01	−1.81

[a] Ford was excluded from the interaction effects analysis because a large number of students did not pretest with STAR Math.

*$p < .05$.

mentation, the focus of the second research question.

Thus, in regard to Research Question 2, we next examined the extent to which teachers differed in their implementation of the continuous monitoring and instructional management system. As noted, the classrooms or teachers in this study were randomly assigned to experimental (AM) and control (no AM) conditions. Those in the experimental group were asked to implement AM with *all* of the students in their classes. We were surprised to learn that many teachers chose not to implement the program with some students. Teachers did not implement the program at all for 1046 of the 2645 (39.5%) students in the experimental group. A follow-up analysis was conducted in an effort to ascertain the extent to which this failure to implement was systematic. That is, we asked whether teachers chose not to implement the program with specific kinds of students (e.g., girls, members of specific racial or multicultural groups, low-functioning students, high-functioning students, and English language learner students). There was no systematic method for exclusion of students. We intentionally did not compel teachers to implement the program, as it was our desire to document what happens in a naturalistic implementation of an intervention. Therefore, we compared gain in achievement for students who were high, low, and nonimplementers.

Separate one-way analyses of variance were performed to test for differences in posttest scores on STAR Math and TerraNova across the three implementation groups. In each analysis, pretest scores and school were entered as covariates. In Table 4 we report the mean residual STAR Math gain and mean

Table 4
Mean Residualized Gains for Students with Varying Levels of Implementation, Controlling for School Effects

Group	Implementation Level No	Low	High
	STAR Math NCE		
Grades 2–5	1.346	0.633	7.077
Grades 6–8	−2.340	−1.658	4.213
	Terra Nova NCE		
Grades 2–5	−1.209	−1.511	4.739
Grades 6–8	−0.565	−0.360	2.092

Note. NCE = normal curve equivalent.

residual TerraNova gain for the entire year for students in the no-, low-, and high-implementation groups, This is shown separately for students in Grades 2–5 and 6–8. For both outcome measures, there were significant differences between groups for both Grades 2–5 (STAR Math, *F*[2, 777] = 9.289, $p < .001$; TerraNova, *F*[2, 954] = 13.240, $p < .001$) and Grades 6–8 (STAR Math, *F*[2, 994] = 18.354, $p < .001$; TerraNova, *F*[2, 1047] = 4.066, $p < .019$). Most notable are the consistently large and positive gains observed for the high-implementation groups. Such results demonstrate the need for high implementation of AM (i.e., >36 objectives mastered) to achieve noticeable effects. It should be noted that the seemingly negative gains here are in part an artifact of using the residualized gain scores in the analyses. A negative value does not indicate that students achieved lower scores on the post-test compared to the pretest. Rather, it indicates the magnitude of their gain relative to the average gain made by all students, as the average gain always is set to zero.

Results of one-way, within-subjects analyses of variance on STAR Math gain for the entire year and on TerraNova gain for the entire year are shown in Table 5, here pooled across grade levels, but again using pretest scores and school as covariates. There was a significant difference ($p < .001$) in the amount of gain evidenced by students in the no-, low-, and high-implementing groups on both the STAR Math Test and the TerraNova Achievement Test. Results indicate that intervention integrity had a significant effect on gain in math scores.

Discussion

Students whose teachers use continuous progress monitoring and instructional management systems significantly outperformed those whose teachers solely use the math curricula being used in their district. Continuous progress monitoring and data-driven decision-making enhances progress toward meeting standards and results in higher test scores. This finding is consistent with the argument of Black and Wiliam (1998) that formative evaluation and CBA are extremely powerful ways to move students toward standards mastery.

Intervention integrity is critical. We were surprised to learn that many teachers in the experimental group chose not to implement the progress monitoring system. This is a limitation of the study. Schmoker (1999) might have predicted as much. As he indicated, teachers are confronted with "initiatives du jour," and unless there is explicit monitoring of implementation and some kind of reward for those who implement the initiatives, teachers do not do so. Goodlad, Klein, and associates (1970) observed this same phenomenon more than 30 years ago. They reported that "innovation is not enough, behind the classroom door even teachers who think they are implementing an innovation are often only twisting it right back into what they have always done" (Goodlad et al., 1970, p. 72).

As investigators look for evidence that specific kinds of instructional interventions impact student performance, it is critical that they take into account intervention integrity. We were able to illustrate that there is significant difference in the extent to which teachers implement instructional interventions as evidenced in this study by the considerable variability in magnitude of student engagement and work. Failure to take this into account

Table 5
Residualized Gains, Cohen's *d*, and ANOVA for the Three Levels of Implementation Pooled Across Grade Levels

Group	*n*	*M*	*SD*	Cohen's *d*	*F*	*t*
		STAR Math NCE				
Implementation level						
No	815	−0.851	14.855	—[a]		
Low	512	−0.534	14.579	0.021		
High	459	6.449	14.497	0.491		
Comparison of all 3 groups					40.529*	
Comparison of no vs high						8.543*
		TerraNova NCE				
Implementation level						
No	1077	−0.657	13.582	—[a]		
Low	492	−1.559	12.254	−0.066		
High	446	3.741	11.655	0.324		
Comparison of all 3 groups					23.843*	
Comparison of no vs. high						5.986*

Note. ANOVA = analysis of variance.
[a] Cohen's *d* was not computed for the nonimplementing group, since it served as the comparison group in the analyses.
$*p < .001$.

would have resulted in misinterpretation of findings. Just as it is important in drug studies to demonstrate that those in the experimental group actually took their drug(s), so too it is critical to demonstrate that students we assume participated in an intervention actually did so. When we took this into account, it firmed up our earlier findings that AM is a highly successful continuous monitoring and instructional management program.

The failure of many of the teachers to implement progress monitoring limited our ability to use an analysis that conformed with our original randomized control group design. Instead, we needed to form implementation integrity groups based on the numbers of objectives that students mastered. This measure of implementation has the potential of being somewhat confounded with achievement. Our inspection of data from multiple implementations of AM indicated that teachers were able to achieve very high levels of objectives mastered in classrooms in which the students represented very diverse ability levels. In such instances each student started at a different level of difficulty, but amount of progress was consistent across ability levels.

School psychologists can take several messages from this study. First, teachers will not always comply with your well-intended recommendations. This was true as much as 40% of the time. Second, there is considerable variability in fidelity or integrity of implementation across and within teachers. Third, when teachers find interventions effective they tell their colleagues and this can result in their receiving additional difficult-to-teach students attending their classrooms for part of a day and actually participating in an intervention. Fourth, failure to consider intervention integrity will mask the actual effectiveness of specific instructional interventions. Fifth, when teachers implement continuous progress monitoring and use data derived from the monitoring to make instructional decisions for individual students, those students profit significantly.

References

Batsche, G., Elliott, J., Graden, J., Grimes, J., Kovaleski, J., Prasse, D., & Tilly, W. D. (2005). *Response to intervention: Policy considerations and implementation.* Alexandria, VA: National Association of State Directors of Special Education.

Black, P. J., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–144.

Burns, M. K. (2002). Utilizing a comprehensive system of assessment to intervention using curriculum-based assessments. *Intervention in School and Clinic, 38,* 8–13.

Burns, M. K., MacQuarrie, L. L., & Campbell, D. T. (1999). The difference between curriculum-based assessment and curriculum-based measurement: A focus on purpose and result. *NASP Communique, 27*(6), 18–19.

Burns, M. K., & Ysseldyke, J. E. (2005). Comparison of existing responsiveness-to-intervention models to identify and answer implementation questions. *California School Psychologist, 10,* 9–20.

Christenson, S., & Ysseldyke, J. (1989). Assessing student performance: An important change is needed. *Journal of School Psychology, 27,* 409–426.

CTB/McGraw-Hill (2001). *Terra Nova.* Monterrey, CA: CTB/McGraw-Hill.

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

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

Edformation. (2004). *AIMSweb response to intervention.* Edina: MN: Edformation (www.AIMSweb.com).

Fuchs, L. S., & Fuchs, D. (1988). Curriculum-based measurement: A methodology for evaluating and improving student programs. *Diagnostique, 14,* 3–13.

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

Gickling, E. E., & Rosenfield, S. (1995). Best practices in curriculum-based assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (pp. 587–596). Bethesda, MD: National Association of School Psychologists.

Gickling, E. E., Shane, R. L., & Croskery, K. M. (1989). Developing math skills in low-achieving high school students through curriculum-based assessment. *School Psychology Review, 18,* 344–356.

Gickling, E. E., & Thompson, V. (1985). A personal view of curriculum-based assessment. *Exceptional Children, 52,* 205–218.

Good, R., & Kaminski, R. (2002). *Dynamic Indicators of Basic Early Literacy Skills (DIBELS).* Longmont, CO: Sopris West. Also available at www.dibels.uoregon.edu

Goodlad, J., Klein, M. F., & Associates. (1970). *Behind the classroom door.* Worthington, OH: Charles A. Jones.

Grimes, J., Kurns, S., & Tilley, W. D. (2006). Sustainability: An enduring commitment to success. *School Psychology Review, 35,* 224–244.

Hintze, J. M., Christ, T. J., & Keller, L. A. (2002). The generalizability of CBM survey-level mathematics assessments: Just how many samples do we need? *School Psychology Review, 31,* 514–528.

Howell, K. W., & Morehead, M. K. (1987). *Curriculum-based evaluation for special and remedial education.* Columbus, OH: Merrill.

Kovaleski, J. F. (2003). *The three tier model for identifying learning disabilities: Critical program features and system issues.* National Research Center on Learning Disabilities Responsiveness to Intervention Symposium, Kansas City, KS.

Kovaleski, J. F., Tucker, J. A., & Duffy, D. J. (1995, June). School reform through instructional support: The Pennsylvania Initiative (Part I). *Communique, 23*(8), Insert.

Marston, D., Muyskens, P., Lau, M., & Canter, A. (2003). Problem-solving model for decision making with high-incidence disabilities: The Minneapolis experience. *Learning Disabilities Research and Practice, 18*(3), 187–200.

McGraw-Hill Digital Learning. (2002). *Yearly Progress Pro.* Columbus, OH: Author.

Renaissance Learning. (1998a). *Accelerated Math™* (computer program). Wisconsin Rapids, WI: Renaissance Learning (http://www.renaissancelearning.com).

Renaissance Learning. (1998b). *STAR Math.* Wisconsin Rapids, WI: Renaissance Learning.

Salvia, J., Ysseldyke, J., & Bolt, S. (2007). *Assessment: In special and inclusive education.* Boston, MA: Houghton-Mifflin.

Schmoker, M. (1999). *Results: The key to continuous school improvement.* Alexandria, VA: Association for Supervision and Curriculum Development.

Shinn, M. R. (Ed). (1989). *Curriculum-based measurement: Assessing special children.* New York: Guilford Press.

Spicuzza, R., & Ysseldyke, J. (1999). *Using Accelerated Math™ to enhance instruction in a mandated summer school program.* Minneapolis: University of Minnesota.

Spicuzza, R., Ysseldyke, J., Kosciolek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance math achievement in urban schools. *Journal for the Education of Students Placed at Risk, 8,* 247–265.

Spicuzza, R., Ysseldyke, J., Lemkuil, A., Kosciolek, S., Boys, C., & Teelucksingh, E. (2001). Effects of using a curriculum-based monitoring system on the classroom instructional environment and math achievement. *Journal of School Psychology, 39,* 521–542.

Tilly, W. D., III. (2002). Best practices in school psychology as a problem-solving enterprise. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 21–36). Bethesda, MD: National Association of School Psychologists.

Tucker, J. A. (1985). Curriculum-based assessment: An introduction. *Exceptional Children, 52,* 199–204.

University of Chicago School Mathematics Program. (2004). *Everyday mathematics.* DeSoto, TX: S Wright Group/McGraw-Hill.

Upah, K. R., & Tilly, W. D. (2002). Best practices in designing, implementing, and evaluating quality interventions. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 483–501). Bethesda, MD: National Association of School Psychologists.

VanDerHeyden, A., & Burns, M. K. (2006). *Examination of the utility of various curriculum-based measures of mathematics performance within a response to intervention model.* Manuscript submitted for publication.

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development and validation of a process for screening referrals to special education. *School Psychology Review, 32*, 204–227.

Witt, J. C., Daly, E., & Noell, G. (1999). *Functional assessments.* Longmont, CO: Sopris West.

Ysseldyke, J. E., Burns, M. K., Dawson, M., Kelley, B., Morrison, D., Ortiz, S., et al. (2006). *School psychology: A blueprint for training and practice III.* Bethesda, MD: National Association of School Psychologists.

Ysseldyke, J. E., & Christenson, S. L. (1987a). Evaluating students' instructional environments. *Remedial and Special Education, 8,* 17–24.

Ysseldyke, J. E., & Christenson, S. L. (1987b). *The instructional environment system 2.* Longmont, CO: Sopris West.

Ysseldyke, J. E., & Christenson, S. L. (2002). *Functional assessment of academic behavior: Creating successful learning environments.* Longmont, CO: Sopris West.

Ysseldyke, J. E., Spicuzza, R., Kosciolek, S., & Boys, C. (2003). Effects of a learning information system on mathematics achievement and classroom structure. *Journal of Educational Research, 96*, 163–174.

Ysseldyke, J. E., & Tardrew, S. (in press). Use of a progress monitoring system to enable teachers to differentiate mathematics instruction. *Journal of Applied School Psychology.*

Date Received: July 7, 2006
Date Accepted: March 12, 2007
Action Editor: John Hintze ■

Jim Ysseldyke, PhD, NCSP, is Emma Birkmaier Professor of School Psychology in the Department of Educational Psychology at the University of Minnesota. He specializes in research and development focused on enhancement of student competence and building the capacity of systems to meet students' needs.

Daniel Bolt, PhD, is Associate Professor of Quantitative Methods in the Department of Educational Psychology at the University of Wisconsin—Madison. He specializes in statistical methodology related to testing and measurement.

TITLE: Effect of Technology-Enhanced Continuous Progress Monitoring on Math Achievement
SOURCE: Sch Psychol Rev 36 no3 S 2007