Using Curriculum-Based Assessment and Curriculum-Based Measurement to Guide Elementary Mathematics Instruction: Effect on Individual and Group Accountability Scores

Amanda M. VanDerHeyden
Vail Unified School District

Matthew K. Burns
University of Minnesota

No Child Left Behind mandates accountability data for school districts. This mandate has led to increased attention to instruction and academic remediation among educational researchers. The current study used schoolwide curriculum-based assessment (CBA) and curriculum-based measurement (CBM) data to plan and deliver mathematics instruction to examine if this would lead to improved student skill within one school year and improved group test scores between school years. The Screening to Enhance Equitable Educational Placement (STEEP) schoolwide problem-solving model was used in one elementary building, with CBM data used to track progress and CBA data used daily to track mastery at each skill level. Results suggested that children made significant progress within one school year, as measured by CBM, and the school significantly increased Stanford-9 mathematics scores after implementing the program. Potential implications for assessment and instructional practice are included.

Research regarding academic interventions is gaining increased attention presumably due to external influences such as the No Child Left Behind Act (NCLB; Daly & McCurdy, 2002). Reading instruction research has a long history, so much so that when the National Reading Panel (NRP) conducted its meta-analysis, it was faced with the daunting task of reviewing over 100,000 articles (NRP, 2000). Research on mathematics instruction has been comparably less prominent over the past 10 years (Badian, 1999; Daly & McCurdy, 2002), which is especially of concern given that only 31% of the nation's fourth-grade students scored at or above the proficiency standard on the 2003 National Assessment of Educational Progress in mathematics, an increase over the 2000 results (Manso & Galley, 2003). Further, mathematical proficiency has become closely linked to successful employment in many different occupations (Saffer, 1999).

Whereas there are considerably fewer studies related to math standards, assessment, and instruction than to reading, some interesting parallel findings exist. First, there is a large and enduring discrepancy between the performances of Caucasian students compared to Hispanic or African-American students (U.S. Department of Education, 2003) and children from low-poverty versus high-poverty backgrounds (Griffin & Case, 1997). These deficits are apparent by kindergarten (Fuchs, Fuchs, & Karns, 2001; Griffin & Case, 1997), persist and intensify without effective and early remediation (Griffin & Case, 1997), and are even more pronounced for children identified with learning disabilities (Fuchs & Fuchs, 2001). Teacher preparation, teacher retention, and specific instructional deficits in the primary grades have been linked to these math deficits (U.S. Department of Education, 2003).

The National Council of Teachers of

Mathematics (NCTM, 2000) described six principles for mathematics instruction. Specifically, they advised that schools should set high mathematics expectations, should provide strong support for all students, should ensure that teachers understand what students know and need to learn, and should ensure that assessments support learning and furnish useful information to both teachers and learners. Essentially, instructionally valid assessment data are a crucial aspect of a mathematics curriculum.

Several curriculum-based approaches to assessment have been developed in efforts to improve the instructional relevance of assessment data. For example, curriculum-based assessment (CBA; Gickling & Havertape, 1981) was developed to suggest instructional adaptations by examining the mismatch between curriculum and student skill (Kovaleski, Tucker, & Duffy, 1995), whereas curriculum-based measurement (CBM; Deno, 1985) was designed to assess the effectiveness of instruction and track student growth toward important learning outcomes over time. Both use the local curriculum as a measure of student progress through existing academic content and, when used together, can offer a comprehensive and instructionally relevant assessment model.

CBA and CBM were developed from a similar paradigm, but fundamental differences exist between them (Burns, MacQuarrie, & Campbell, 1999). CBA was designed to systematically assess the "instructional needs of a student based upon the on-going performance within the existing course content in order to deliver instruction as effectively as possible" (Gickling, Shane, & Croskery, 1989, pp. 344-345). CBA has been described for use within a mastery model of learning whereby a specific hierarchical sequence of skills and subskills are defined and the performance of each specific subskill is measured to ensure that the skill has been mastered before providing instruction on the next skill in the hierarchy. Instruction may be described as occurring within an instructional hierarchy where acquisition-level instruction is provided for a new skill, fluency-building instructional activity is provided for a slightly easier or lower-level skill that the child can perform accurately, and application-level instructional activity is provided for an even easier skill that the child can perform both accurately and fluently. Hence, CBA is used primarily to design and deliver instruction (Burns, Dean, & Klar, 2004), allowing for the identification of specific deficits in need of remediation (Daly, Witt, Martens, & Dool, 1997).

CBM's primary purpose is to assist in educational decision making (Deno, 1989) by directly measuring the desired functional outcome of instruction. Because CBM measures a general desired outcome (e.g., reading fluency), measurement can occur across instructional sessions and skill levels to relevantly and accurately reflect whether or not the child is achieving adequate growth toward the desired outcome (Fuchs & Deno, 1994). Hence, CBM is superior to CBA in tracking growth over time and estimating generalization whereas CBA may be more useful for suggesting specific interventions to guide instruction (Burns et al., 1999).

The functional difference between CBA and CBM has often been a point of confusion (Burns et al., 1999). For a thorough discussion of the potential differences between CBA and CBM, the reader is referred to Shinn and Bamonto (1998). For the purpose of this study, the difference between CBA and CBM lies in the scope of the task items included on probe materials rather than the format of administration. CBA and CBM measures used in this study were both administered according to the procedures described by Shinn (1989), but the assessment probes were different. The CBA measures included items that sampled a single skill and skill level (e.g., subtraction with answers 0-5), whereas the CBM measures comprehensively sampled multiple skills and skill levels that the child would be expected to perform by the end of

archy where
provided for a
instructional
atly easier or
can perform
l instructional
n easier skill
th accurately
d primarily to
Burns, Dean,
identification
remediation
997).
assist in edu-
no, 1989) by
nctional out-
M measures a
eading fluen-
cross instruc-
to relevantly
r or not the
th toward the
Deno, 1994).
A in tracking
g generaliza-
useful for sug-
guide instruc-

erween CBA
t of confusion
gh discussion
een CBA and
o Shinn and
pose of this
BA and CBM
s included on
he format of
IM measures
administered
described by
t probes were
ncluded items
ill level (e.g.,
whereas the
ely sampled
hat the child
by the end of

the intervention program (e.g., addition, subtraction, multiplication, and division of basic facts). Thus, CBA was used to identify the particular level at which the child should receive instruction throughout the intervention, whereas CBM was used to judge the overall effectiveness of the program toward the desired functional outcome (e.g., fluency in basic math facts).

Proponents for CBA argue that academic difficulties are the result of a mismatch between student skills and the curriculum or instruction (Enggren & Kovaleski, 1996; Gravois & Gickling, 2002) whereby the curriculum can be too difficult resulting in student frustration, or too easy, resulting in student boredom. Thus, CBA attempts to find "the window of learning" (Tucker, 1985, p. 201) between boredom and frustration by matching skill level and challenge by assessing the difficulty of the instructional material for each child. The amount of known information within the task is determined by having the child complete the task and computing the percentage of known material. Gickling and Thompson (1985) suggested that if children are provided an appropriate level of challenge, called the instructional level, optimal learning will occur. Instructional material that is too difficult is termed a frustration level, and material that is not challenging enough is called a mastery level (Gravois & Gickling, 2002).

The instructional level for reading has been well defined and researched, but fewer data exist to identify an instructional level for mathematics (Burns, 2004a). Deno and Mirkin (1977) suggested that the instructional level for mathematics be computed with fluency measures such as those used in CBM. Table 1 lists the median digits correct proposed by Deno and Mirkin to equal the frustration, instructional, and mastery levels, respectively. Use of fluency measures to determine optimal challenge for mathematics makes intuitive sense, given that fluent computation is a goal for mathematics instruction (NCTM, 2000) and leads to more competent functional performance, but to date few data have been provided to identify what fluency rate represents an instructional level.

Burns et al. (1999) suggested that both CBA and CBM meet the three essential features of instructionally valid assessment outlined by Fuchs and Deno (1994). Principles for mathematics instruction outlined by the NCTM (2000) suggest that teachers should understand what students know and need to learn and that assessments should support learning and furnish useful information to teachers and learners. CBA and CBM 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. Previous research has found that using CBM data to demonstrate academic growth to students and teachers has led to improved academic outcomes (Fuchs, Fuchs, Hamlett, & Stecker, 1991; Fuchs et al., 1997). Thus, the use of CBA and CBM together seems to provide an assessment model that could address the principles outlined by the NCTM.

Research has found that providing appropriately challenging instruction, at the child's

Table 1 Instructional Categories for Fluency Measures of Mathematics

	Median Digits Correct/Minute		
	Frustrational	Instructional	Mastery
Grades 1 to 3	0 to 9	10 to 19	20 or more
Grades 4 to 12	0 to 19	20 to 39	40 or more

instructional level, has led to improved student outcomes (Burns, 2002; Gickling & Rosenfield, 1995; Shapiro & Ager, 1992). Previous research focused on using CBA data to plan and deliver instruction for individual students, rather than on a larger scale such as classwide and schoolwide. Given that NCLB calls for student performance data to be reported and evaluated for various groups, research is needed that explores schoolwide applications of using CBA data to improve instruction.

The current study implemented a schoolwide assessment model to examine the effectiveness of using CBA and CBM data to plan and deliver instruction. The two specific research questions that guided this study were (a) Will using CBA and CBM data to determine the instructional needs of children throughout an elementary building lead to significantly improved student skill within one school year? (b) Will guiding instruction within an elementary building with CBA and CBM data lead to improved group test scores used for accountability purposes between school years? The data were analyzed two ways for each question. First, means were computed and analyzed using an analysis of variance; second, scores for each question were converted to an ordinal scale and analyzed using a nonparametric test. Nevertheless, the study was largely descriptive in nature, and inferential statistics were used to clarify the specific questions rather than to make generalizations.

Method

Overview of STEEP Problem-Solving Model of Assessment

The school where this study was conducted was participating in a schoolwide problem-solving model of assessment called "Screening to Enhance Equitable Educational Placement" (STEEP; Witt, Daly, & Noell, 2000). STEEP is similar to other problem-solving models of assessment that rely on CBA and CBM data to identify problems and then plan, implement, and evaluate solutions to those problems (Fuchs & Fuchs, 1998; Good & Kaminski, 1996; Shinn, 1995). STEEP is a multiple-gate process that begins with schoolwide screening in reading, math, and writing using CBA probes. A series of functional assessment and intervention activities are then conducted for children who perform in the at-risk range during the screening. CBM probes are administered periodically to track growth. Ultimately, intervention responsiveness is used to determine whether or not children should be evaluated for special education. Use of the STEEP model has been found to accurately identify children who merit consideration for special education eligibility (VanDerHeyden, Witt, & Naquin, 2003).

In the present study, prior to schoolwide screening, the school-based consultant met with general and special education leadership at the school to provide an overview of the process and identify on-site "coaches" to assist with the schoolwide screening. Classwide screenings were scheduled in one-hour rolling blocks by grade level, and a sufficient number of coaches were trained (or borrowed from other sites) to allow for one coach per classroom at each grade level during a given one-hour period (typically four to five coaches per school). During the classwide screening, teachers administered two math probes, one selected to represent current skill placement (CBA) and one to track growth (CBM) throughout the year. The teachers used scripted instructions and digital timers to administer all probes using directions adapted from Shinn (1989). Coaches provided assistance as needed to ensure integrity. Mathematics probes were administered to the entire class simultaneously. The classwide screening for math required about 5 minutes to complete. The school-based consultant attended the next grade-level planning meeting to train teachers to score the math probes. Teachers were then provided a graph showing their stu-

roblems and
ate solutions
Fuchs, 1998;
inn, 1995).
s that begins
ading, math,
A series of
ention activ-
ren who per-
he screening.
riodically to
nterventeion
line whether
ed for special
del has been
hildren who
ducation eli-
& Naquin,

schoolwide
sultant met
n leadership
rview of the
hes" to assist
. Classwide
-hour rolling
ient number
rowed from
ch per class-
a given one-
coaches per
screening,
probes, one
ll placement
wth (CBM)
used script-
to adminis-
dapted from
assistance as
Mathematics
entire class
creening for
o complete.
ttended the
ing to train
es. Teachers
ng their stu-

dents' performance relative to the instructional standards used with STEEP. It should be noted that STEEP uses similar procedures for reading and writing as well, which were also implemented, but only mathematics is discussed here.

These data were used to determine whether or not a classwide problem existed. If a classwide problem was found, classwide intervention occurred prior to any individual assessment of children in that particular class. A classwide problem was identified when the median score for the class fell below the instructional minimum standard (Deno & Mirkin, 1977; see Table 1). If no classwide problem was detected, children performing below the 16th percentile and below the instructional minimum participated in an individual assessment of the influence of incentives upon performance (skill/performance deficit assessment). Children who continued to perform below the instructional standard were given incentives and participated in individual, protocol-based intervention performed in the classroom and monitored for integrity and effectiveness each day for at least 10 consecutive sessions.

Description of School

Data were collected from one elementary school in a rural community approximately 15 miles outside of Tucson, Arizona. (Although collecting data in only one school suggests potential difficulties with external validity of their interpretation, the intensity of the intervention was such that it precluded a wider implementation at this point.) Gender was equally distributed for grades three through five during the 2002-2003 school year. In terms of ethnicity, 79% of students were Caucasian (not Hispanic), 16% were Hispanic or Latino, 4% were African American, and less than 1% were Asian and Native American. Thirty-eight students (11%) within the third through fifth grades received special education services, nine (3%) were eligible for Title I services, and six (1.7%) were English-language learners. There were four classrooms per grade level with approximately 25 students per class. Teaching experience for third- through fifth-grade teachers ranged from less than 1 year to 25 years.

Identification of Mathematics Problem

The first screening occurred in the fall. In the participating school, third-grade teachers selected mixed addition and subtraction problems with answers 0 to 18. Fourth-grade teachers selected multiplication facts 0 to 9. Fifth-grade teachers selected mixed multiplication and division facts 0 to 9. Classwide math problems were detected in all classrooms grades four through five. Several classwide problems were also detected at the third grade level. Specifically, 22% of third graders performed below the instructional standard for basic addition and subtraction facts at a time of year when instruction was proceeding to multiplication, 69% of fourth graders performed below the instructional criterion for basic single-digit multiplication facts, and 58% of fifth graders performed below the instructional criterion for basic single-digit multiplication and division facts. Hence, a schoolwide math problem was suspected, and an intervention plan was subsequently proposed to improve math skills for all children in grades three through five.

Mathematics Curriculum

The school used the Saxon Math Curriculum (Saxon Publishers, 2004) to guide mathematics instruction for third-grade students and Harcourt Math (Harcourt Publishing, 2003) to guide instruction for fourth- and fifth-grade students. The sequence of mathematics skills instruction followed the sequence developed from Arizona standards (see Table 2). These curricula had been used at this school for the preceding two years.

Table 2 Skill Sequence Developed from State and District Standards 2002-2003 Following Mastery of Basic Facts

3rd Grade Sequence
1. mixed probe of addition and subtraction with and without regrouping
2. single digit multiplied by double digit with and without regrouping
3. double digit multiplied by double digit with and without regrouping
4. estimation

4th Grade Sequence
1. single digit multiplied by double digit with and without regrouping
2. double digit multiplied by double digit with and without regrouping
3. single-digit divisor into double digit with and without remainders
4. adding and subtracting fractions with like denominators
5. estimation

5th Grade Sequence
1. mixed multiplication with and without regrouping
2. single-digit divisor into double digit with remainders
3. review of all digit combinations with remainders
4. mixed conversion of mixed fractions to improper and improper to mixed
5. reducing fractions to simplest form
6. finding least common denominator
7. adding and subtracting fractions with like denominator
8. estimation

By the time the screening was conducted (November), the instructional calendar and curricula had progressed to higher-level skills that required mastery of the basic facts assessed during the screening. Hence, the problem was first defined as a fluency problem for basic math skills. Teachers were instructed to continue with their instructional plan, covering all the required objectives in the state standards, and to consider the intervention plan to be a supplemental activity. The general education intervention was designed to provide concentrated practice opportunities for the basic computational skills that the students had not mastered as expected and to progress through a sequence of computational skills that were required for each grade level by the state standards under the assumption that students would not master those skills without extra practice.

The goal of the intervention was to remediate deficits in basic computational skills in grades three through five using a fluency approach. A fluency-building intervention was selected for two reasons. First, universal screening data identified a basic facts fluency problem in all three grades. Second, the fluency with which a student can perform a skill directly impacts the degree to which the student can apply that skill to the solution of more complex problems under conditions that differ from the conditions used during training (Binder, 1996). Hence, fluent performance of basic computational skills was a prerequisite for achieving the more comprehensive goals reflected in the state standards and taught within the school's math curriculum.

General Education Intervention Plan

During a one-hour faculty meeting, data from the screening were presented to the faculty, and an intervention plan was described and discussed. The basic plan was outlined and teachers were given an opportunity to ask

questions and make suggestions. Specifically, teachers gave input on how materials would be provided to them, how materials would be organized, layout of the datasheet, period of time each day when intervention would occur, start date, and procedure for turning in data each week.

For the basic facts sequence, children were divided into tutoring pairs and a format of intervention was used similar to that reported by Greenwood and colleagues (i.e., classwide peer tutoring; Greenwood, Carta, & Hall, 1988) and Fuchs and colleagues (i.e., peer-assisted learning strategies; Fuchs, Fuchs, Mathes, & Simmons, 1997; Fuchs, Fuchs, Phillips, Hamlett, & Karns, 1995). The key elements of the intervention included (a) paired peer instructional-level skill practice with multiple massed opportunities to respond at a brisk pace and with immediate corrective feedback, (b) independent timed practice for a score, (c) slightly delayed corrective feedback, and (d) progression to the next skill level contingent on meeting a mastery criterion on the lower-level skill. Teachers were trained to identify children with high error rates and to provide a scripted 5-minute lesson outside of the 30-minute classwide intervention period. A "coach card" or protocol for the intervention was developed and provided to teachers specifying each step of the intervention in observable terms. (The coach cards used for basic facts and the higher-level skills are provided in Appendices A and B.) The teacher's role was to monitor during the intervention and to ensure high engagement and accurate implementation of procedures by the children.

The principal and teachers elected to perform the intervention on each individual child's instructional skill level and to move children through a pre-specified sequence of skills and difficulty levels based upon individual child performance. The sequence of skills began with basic facts and progressed through a series of computational skills required in the state standards for each grade level. (The sequence of skills used is provided in Table 2.) Hence, prior to beginning intervention, the skill and level at which each child performed above the instructional standard was identified. Multiple forms of probes sampling each skill in the sequence were developed and organized for teachers. Teachers were also provided with the protocol for the intervention, the sequence of skills, a data sheet for recording student performance, a digital timer, approximately 12 sets of flashcards for each of the basic facts groups, and peer practice sheets for the higher-level computational skills.

Training for the intervention was provided using the following sequence of activities. First, the steps of the intervention were described and provided to the teachers in writing. The experimenter then role-played each of the steps with the teachers and answered questions. On the first day of implementation, a trained coach was assigned to each classroom to ensure that each step of the intervention was correctly conducted and to assist the teacher in training the students to perform the intervention and score probes. Following training, teachers performed the intervention four days a week and administered a mixed probe of basic facts each Friday in lieu of the intervention session. Each day of intervention, students earned a score of digits correct in 2 minutes on the skill level on which they were working. The teacher recorded these data on a datasheet, which was given to a consultant every Friday for graphing. Teachers were trained to move children to the next highest skill in the sequence on the following day given a mastery-level score during the daily timed assessment.

Monitoring and Ensuring Integrity of the Intervention

At least one class was randomly selected each week for integrity monitoring. During this period, an independent, trained observer made an unannounced visit to the classroom

to note the occurrence or nonoccurrence of each scripted step of the intervention protocol. The teacher received immediate, corrective feedback to implement any omitted steps, which was then recorded on the integrity observation form. This information was also provided to the principal on a weekly basis so that she could provide additional feedback if needed. In addition, the observer noted any behaviors that might be interfering with proper implementation or the effectiveness of each intervention component (e.g., classroom management issues) and provided the teacher with ideas for how to resolve such issues. The observer also provided this information to the principal so that the principal could provide additional feedback and support if deemed appropriate.

Each week, teachers received graphs of their students' scores across days of the intervention. Individual verbal feedback was provided based upon the data. Also, approximately every three weeks, the consultant and principal attended a grade-level planning meeting to clarify procedures, answer questions, and remind teachers to turn in data. This meeting served as an opportunity to facilitate teachers sharing ideas about how to best manage the intervention. At the end of each quarter, the consultant attended a faculty meeting and presented data on the progress of teachers and grade levels, clarified procedures, and answered teacher questions and addressed teacher concerns.

The consultant completed a checklist of the critical components of the intervention each week for each teacher, and this checklist was used by the principal to give the teacher specific verbal feedback along with the teacher's graphs for the week. The checklist included the following items: Were the data turned in? Was the sequence correct? Did the intervention occur every day? Was the mixed probe administered on Friday? On average, 75% of participating teachers correctly completed the above steps on a weekly basis. The most frequently missed step was failing to turn in student data to the office. To contend with this error, the data-entry person simply contacted teachers whose data had not been submitted and prompted them to send the information to the office. In all cases, this corrective action was effective.

Measures

Multiple equivalent forms were created for all of the CBM and CBA measures used during the study. CBA was used daily to track individual growth and mastery at each skill level (and to determine when to increase task difficulty). Two outcome measures were used to assess mathematics skills. The weekly CBM probes provided data regarding mathematics fluency, and group test data were obtained from the Stanford Achievement Test (9th ed.; SAT-9). The reliability and validity of data from CBM and CBA are well documented (Burns, 2004b; Marston, 1989), and the SAT-9 is one of the oldest and most widely used group measures of student achievement with sufficient reliability to assess groups of scores (Haladyna, 1998). All CBA and CBM probes were administered according to the procedures described by Shinn (1989). The SAT-9 was administered at the beginning of May according to standardized procedures.

Results

Student Growth Within One School Year

Individual student CBM data were analyzed by randomly selecting one probe from January, February, March, and April of 2002-2003 for each student. Data from those probes are represented in Table 3. Random selection of one probe was used, as opposed to computing a monthly mean or median, to utilize an element of randomness to allow for the variance to be analyzed through a repeated-measure ANOVA. Results of this analysis found significant effects for each grade and for the total sample. Further, Cohen's d was computed as estimates of effect sizes, which ranged from .49 to .97 standard deviation units.

contend with
n simply con-
not been sub-
end the infor-
s, this correc-

ere created for
ures used dur-
daily to track
at each skill
increase task
res were used
weekly CBM
mathematics
were obtained
Test (9th ed.;
alidity of data
l documented
and the SAT-
st widely used
evement with
oups of scores
CBM probes
to the proce-
). The SAT-9
uning of May
dures.

School Year

ata were ana-
se probe from
April of 2002-
n those probes
dom selection
ed to comput-
, to utilize an
v for the vari-
repeated-mea-
analysis found
le and for the
l was comput-
which ranged
viation units.

Table 3 Mean Digits Correct/2 Minutes Scores for Randomly Selected Monthly Mathematics Probes

	January		February		March		April		
Grade	M	SD	M	SD	M	SD	M	SD	F
Third	27.9	9.9	37.9	13.4	35.3	13.0	38.8	12.3	13.45*
Fourth	39.7	17.6	50.1	22.0	53.4	25.3	58.9	27.0	35.01*
Fifth	47.6	22.9	50.8	22.7	52.0	24.8	59.3	25.1	25.52*
Total	41.1	20.5	48.0	21.5	49.2	24.0	55.1	25.1	64.29*

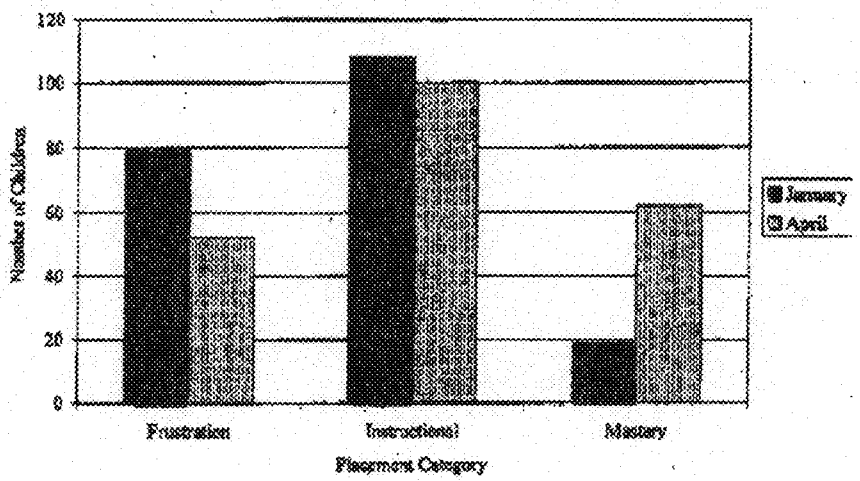
* $p < .001$

Cohen's *d* (effect size between January and April scores): third .37; fourth .86; fifth .49; and total .62.

Thus, according to Cohen's (1988) interpretative approach, moderate effects were noted for students in the fifth grade and for the total sample, but strong effects were found in grades three and four.

Individual student data were also interpreted by using the mathematics fluency score to place each child's score into the categories of frustration, instructional level, and mastery, as suggested by Deno and Mirkin (1977). These nonparametric data were interpreted with a Wilcoxon Signed Ranks test. Next, the percentage of students in each category was computed for January and April. Figure 1 graphically displays the data, suggesting a shift from frustration and instructional level in January (when the intervention had been underway for approximately five weeks) to a larger percentage of mastery in April. In January, there were 79 children (38%) whose scores fell within the frustration range while 19 (9%) scored within the mastery category. The remaining 108 (52%) of the January scores were categorized as instructional. However, by April the number of children scoring in the mastery category was 62 (29%) with only 52 (24%) student scores falling in the frustration category. Further, the number of students who scored in the instructional category during the April probe had decreased to 100 (47%). Thus, category placements were reliably

Figure 1 Frequency graph of student CBM data for 2002-2003 by fluency categories.



higher ($Z = 6.89$, $p < .001$) in April than in January.

Group Score Data Between School Years

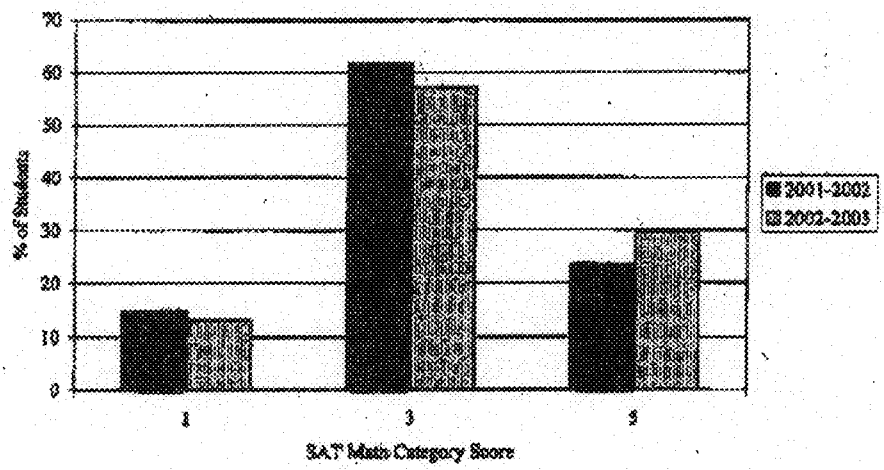
Mathematics standard scores from SAT-9 data for children in the participating school were compared before (2001-2002 school year) and after (2002-2003 school year) implementation. Variance for these data was examined with a t test, rather than a repeated-measure ANOVA, because these data did not contain any element of randomness. Table 4 displays the data by grade and for the compiled samples. As illustrated, scores post-implementation were significantly higher than the previous year for all grades and for the total sample. Effect sizes were also computed using Cohen's d; they ranged from .29 to .45, which suggest small to moderate effects. SAT-9 data from 2000-2001 resulted in mean standard scores of 596.28 for third grade, 628.10 for fourth grade, and 646.86 for fifth grade. Although no analyses on these scores were conducted, they appeared slightly higher than average scores for 2001-2002 and were lower than the 2002-2003 data.

Categorical scores were also provided for the SAT-9 mathematics data using three categories, below average, average, and above average. Figure 2 shows the change in category scores from Year 1 (2001-2002) to Year 2 (2002-2003) and suggests a shift toward the above-average category. Specifically, data from the 2001-2002 school year showed that 71 students (24%) scored in the above-average category, 186 students (62%) scored in the average category, and 45 students (15%) scored in the below-average category. However, the number of students scoring in the SAT-9 above-average category during 2002-2003 increased to 105 (30%), with 202 students (57%) scoring in the average range and 46 (13%) scoring in the below-average range. Thus, the number of children scoring in the above-average range increased by 6% between years, and scores within the average and below-average range decreased by 4% and 2%, respectively. These ordinal data were further analyzed with the Mann-Whitney test, which found a significant effect between Years 1 and 2 ($Z = 3.37$, $p < .001$).

Discussion

The school in which this study was conducted had a pervasive problem with basic mathematics fact fluency. Most of the children in grades 3 through 5 lacked the basic computational skills to perform at grade level

Figure 2 Frequency graph of SAT-9 mathematics data between years by categories.



ange in catego-
002) to Year 2
ift toward the
cifically, data
ar showed that
he above-aver-
2%) scored in
nudents (15%)
age category.
ents scoring in
tegory during
0%), with 202
average range
below-average
ildren scoring
creased by 6%
in the average
sed by 4% and
data were fur-
-Whitney test,
ffect between
01).

study was con-
ern with basic
st of the chil-
cked the basic
at grade level

2001-2002
2002-2003

and to benefit from instruction related to applied or problem-solving tasks. This finding mirrored a national trend (Loveless, 2003) and suggests a challenge for schools attempting to comply with the requirements of NCLB. In this study we investigated the effectiveness with which direct measures of student math performance could be obtained and used to build an intervention program to improve performance for all children at multiple grade levels. Previous studies have found that using CBA data to modify instruction led to increased outcomes for individual students (Burns, 2002; Gickling & Rosenfield, 1995; Shapiro & Ager, 1992). The data reported in this article suggest that using CBA and CBM to guide remediation efforts substantially increased growth within one year, and at the same time led to significantly improved schoolwide scores on a standardized norm-referenced group test used for accountability reporting following intervention implementation.

The SAT-9 was not an ideal measure to use given the psychometric concern about test sensitivity associated with any norm-referenced measure (Salvia & Ysseldyke, 2004). This should be taken into account when interpreting the small to moderate effect sizes. Nonetheless, the state where this study was conducted used the SAT-9 along with a criterion-referenced state-developed test for accountability purposes. No technical data were available on the state-developed test at the time of this study. Further, the state test had been modified each year that it had been used and, therefore, could best be considered under development. Hence, the SAT-9 was selected as the strongest tool appropriate for use in this study, despite its limitations. Further, proficiency categories provided by the state when reporting to individual districts and teachers and schools commonly used these data to judge the effectiveness of their instructional efforts (e.g., percentage of children meeting or exceeding standards). These categories were set apriori by the state. The data were reported here both as scaled scores and by category scores.

By conducting universal screening using CBA and CBM, deficits were also identified at the first- and second-grade levels in math. Teachers reported frustration and loss of instructional time at the beginning of each year as they sampled back to teach or re-teach skills that should have been mastered at the lower grade levels. Although the effects that were obtained with the current general education intervention were notable, they were still moderate in effect size or practical significance. Whereas teachers and district administrators reported satisfaction with the improvements noted, the study authors desired to obtain greater growth and achieve-

Table 4 SAT-9 Standard Scores and t-test Results for Pre- and Post-Implementation Years by Grade

	2001-2002			2002-2003			
Grade	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*
Third	85	562.06	143.80	129	602.54	35.20	3.07**
Fourth	116	611.09	120.61	117	638.22	33.39	2.35*
Fifth	113	636.73	109.86	107	659.17	35.77	2.01*
Total	314	607.04	126.83	353	631.53	41.93	3.42**

* $p < .05$. ** $p < .01$.
Cohen's *d* (effect size between years): third .45; fourth .35; fifth .31; and total .29.

ment in future years while reducing the "cost" of the intervention simultaneously. Hence, in the future, the intervention would be modified to enhance efficiency or ease of implementation and to use a fluency-building component for basic computational skills required in the state performance standards but also to include scripted practice with problem-solving using the same classwide peer tutoring format. The revised intervention plan would occur within 30 minutes to supplement math instruction in general education grades 1 through 5. In fact, a follow-up study is currently underway to examine the effectiveness of a more efficient model where teachers are trained to increase skill difficulty based upon a class median score in the mastery range on a specific skill.

Several practical implications are worth noting. First, universal screening was used to detect and to define the problem. Without the benefit of universal screening data, the problem was not likely to be detected or properly defined, and children at this school would be disproportionately likely to be identified for special services if evaluated and compared to national comparison groups (VanDerHeyden, 2001; VanDerHeyden et al., 2003). The screening was time-efficient, requiring only approximately 5 minutes of time in the classroom and approximately 20 minutes of scoring time. The screening was inexpensive to conduct and was correctly administered by classroom teachers in the general classroom setting. Hence, using CBA and CBM data for universal screening would be easy for schools and would yield data upon which accurate decisions could be based (Marston, Mirkin, & Deno, 1984; Shinn, Tindal, & Spira, 1987). Because the process is inexpensive and time-efficient, it could also be repeated several times during the year to identify not only those children who are performing below the expected level of performance but also those who are growing at a slower rate (Fuchs & Fuchs, 1998). Frequently repeating the measurement allows for ongoing progress monitoring to evaluate the effectiveness of the attempted solution (e.g., classwide, gradewide, or schoolwide intervention).

Limitations

As stated earlier, research has consistently supported teaching students at their individual instructional level (Burns, 2002; Gickling & Rosenfield, 1995; Shapiro & Ager, 1992). The current study was an attempt to expand that instructional methodology to an entire school building. Although the data supported this practice, this study was mostly descriptive in nature. Thus, changes in practice based on these data should be made only cautiously.

Further, limitations need to be noted. Given the scope of the intervention, it was necessary to implement the intervention in only one building, which limited the external validity of the data. Additional research is needed to examine the effects across different buildings in different settings in order to assess generalizability. Second, the study examined data within one year using the same group of students, and between two years using two different cohorts of students. The growth shown during one school year could have been due to maturation, rather than the intervention, so a between-year analysis was conducted as well. Even though the two analyses found consistent results, data from the within-year comparison should be interpreted with maturation in mind. Future research with a control group is needed to attempt to control for this possible confound.

Conclusions

The intervention in this study focused on fluency building. Whereas fluent performance of basic computational skills is necessary to establish accurate higher-level or more complex problem solving, it is probably not sufficient. Hence, following establishment of fluent performance in basic computational skills, students should receive instruction in using these skills

ing to evaluate
mpted solution
or schoolwide

as consistently
t their individ-
2002; Gickling
& Ager, 1992).
mpt to expand
gy to an entire
data supported
stly descriptive
actice based on
y cautiously.
to be noted.
vention, it was
ntervention in
ed the external
nal research is
across different
s in order to
nd, the study
using the same
een two years
students. The
ool year could
rather than the
ar analysis was
ough the two
lts, data from
ould be inter-
mind. Future
s is needed to
ible confound.

dy focused on
at performance
is necessary to
more complex
not sufficient.
of fluent per-
l skills, students
ing these skills

to solve more complex and applied problems. In this school, guided practice and instruction for solving complex problems was provided through the regular math curriculum. In future years, however, a similar format may be attempted to provide routine and structured practice applying fluent knowledge to enhance outcomes on problem-solving tasks.

NCLB has expanded the focus of intervention research beyond the individual student, perhaps leading some to argue for a shift toward prevention rather than a reactive approach to intervention with individual students (Shapiro, 2000; Fuchs & Fuchs, 2001). In fact, CBA and CBM have been discussed as examples of instructional technology widely supported for use in universal screening and in monitoring individual student growth, but lacking data addressing systemic implementation (Shapiro, 2000). This study described the use of CBA and CBM data to determine which intervention should be delivered to all children participating in general education instruction at the building level and constituted a systemic early intervention approach as called for in the literature.

It is hoped that this article will be the first of many that demonstrate effective schoolwide interventions while providing a model for practitioners interested in empirically investigating the effectiveness of their own practices. These data are also ideal for tracking, reporting, and ensuring adequate yearly progress as required through NCLB. Further, because the data are universal and capture both the level and trend of performance, they can also be used effectively to identify children who are struggling in the general education curriculum and might need to be considered for special services (Fuchs & Fuchs, 1998). In the current context of accountability, tools that allow districts to accomplish more than one assessment objective (e.g., monitoring general education effectiveness and universal screening) are critical to limit time spent in assessment and maximize time devoted to instruction.

References

Badian, N. (1999). Persistent arithmetic, reading or arithmetic, or reading disability. *Annals of Dyslexia, 49*, 45-70.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *Behavior Analyst, 19*, 163-197.

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. (2004a). Empirical analysis of drill ratio research: Refining the instructional level for drill tasks. *Remedial and Special Education, 25*, 167-175.

Burns, M. K. (2004b). Using curriculum-based assessment in the consultative process: A useful innovation or an educational fad. *Journal of Educational and Psychological Consultation, 15*, 63-78.

Burns, M. K., Dean, V. J., & Klar, S. (2004). Using curriculum-based assessment in the responsiveness to intervention diagnostic model for learning disabilities. *Assessment for Effective Intervention, 29* (3), 47-56.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Daly, E. J., & McCurdy, M. (2002). Getting it right so they can get it right: An overview of the special series. *School Psychology Review, 31*, 453-458.

Daly, E. J., Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review, 26*, 554-574.

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

Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 1-17). New York: Guilford Press.

Deno, S., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children.

Enggren, R., & Kovaleski, J. F. (1996). *Instructional assessment*. Harrisburg, PA: Instructional Support System of Pennsylvania.

Fuchs, L. S., & Deno, S. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children, 61*, 15-24.

Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice, 13*, 204-219.

Fuchs, L. S., & Fuchs, D. (2001). Principles for the prevention and intervention of mathematics difficulties. *Learning Disabilities Research and Practice, 16*, 85-95.

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

Fuchs, L. S., Fuchs, D., & Karns, K. (2001). Enhancing kindergartners' mathematical development: Effects of peer-assisted learning strategies. *The Elementary School Journal, 101*, 495-510.

Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., Katzaroff, M., & Dutka, S. (1997). Effects of task-focused goals on low-achieving students with and without learning disabilities. *American Educational Research Journal, 34*, 513-543.

Fuchs, D., Fuchs, L., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal, 34*, 174-206.

Fuchs, L. S., Fuchs, D., Phillips, N. B., Hamlett, C. L., & Karns, K. (1995). Acquisition and transfer effects of classwide peer-assisted learning strategies in mathematics for students with varying learning histories. *School Psychology Review, 24*, 604-620.

Gickling, E. E., & Havertape, S. (1981). *Curriculum-based assessment*. Minneapolis, MN: School Psychology Inservice Training Network.

Gickling, E., & Rosenfield, S. (1995). Best practices in curriculum-based assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (3rd ed., pp. 587-595). Washington, DC: 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., & Thompson, V. (1985). A personal view of curriculum-based assessment. *Exceptional Children, 52*, 205-218.

Good, R. H., & Kaminski, R. A. (1996). Assessment for instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills. *School Psychology Quarterly, 11*, 326-336.

Gravois, T. A., & Gickling, E. E. (2002). Best practices in curriculum-based assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 885-898). Bethesda, MD: National Association of School Psychologists.

Greenwood, C. R., Carta, J. J., & Hall, R. V. (1988). The use of peer tutoring strategies in classroom management and educational instruction. *School Psychology Review, 17*, 258-273.

Griffin, S., & Case, R. (1997). Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education, 3*, 1-49.

Helmstadter, T. M. (1998). Review of the Stanford Achievement Test (9th ed.). In J. C. Impara & B. S. Plake (Eds.), *The thirteenth mental measurements yearbook* (pp. 928-930). Lincoln: The University of Nebraska-Lincoln.

Harcourt Publishers. (2003). *Harcourt math*. Orlando, FL: Author.

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

Loveless, T. (2003). *Trends in math achievement: The importance of basic skills*. Paper presented at the Secretary's Summit on Mathematics. February 6, 2003. Retrieved November 20, 2003 from http://www.ed.gov/rschstat/research/progs/mathscience/loveless.html.

Manzo, K. K., & Celley, M. (2003). Math climbs, reading flat on '03 NAEP. *Education Week, 23*(12), 1, 18.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford Press.

Marston, D., Mirkin, P., & Deno, S. (1984). Curriculum-based measurement: An alternative to traditional screening, referral, and identification. *The Journal of Special Education, 18*, 109-117.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.

Saffer, N. (1999). Core subjects and your career. *Occupational Outlook Quarterly, 43*(2), 26-40.

Saxon Publishers. (2004). *Saxon math*. Norman, OK: Saxon Publishers.

Salvia, J., & Ysseldyke, J. E. (2004). *Assessment in special and inclusive education* (9th ed.). Boston: Houghton Mifflin.

Shapiro, E. S. (2000). School psychology from an instructional perspective: Solving big, not little problems. *School Psychology Review, 29*, 560-572.

Shapiro, E. S., & Ager, C. (1992). Assessment of special education students in regular education programs:

Linking assessment to instruction. *Elementary School Journal*, 92, 283-296.

Shinn, M. R. (1989). Identifying and defining academic problems: CBM screening and eligibility procedures. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 90-129). New York: Guilford Press.

Shinn, M. (1995). Best practices in curriculum-based measurement and its use in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (3rd ed., pp. 547-567). Washington, DC: National Association of School Psychologists.

Shinn, M. R., & Bamonto, S. (1998). Advanced applications of curriculum-based measurement: "Big ideas" and avoiding confusion. In M. R. Shinn (Ed.) *Advanced applications of curriculum-based measurement* (pp. 1-31). New York: Guilford Press.

Shinn, M. R., Tindal, G. A., & Spira, D. A. (1987). Special education referrals as an index of teacher tolerance: Are teachers imperfect tests? *Exceptional Children*, 54, 32-40.

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

U.S. Department of Education. (2003). *Mathematics and science initiative concept paper*. Retrieved November 30, 2003 from http://www.ed.gov/rschstat/research/progs/mathscience/conceept_paper.pdf.

VanDerHeyden, A. M. (2001). *Comparison of several methods for identifying students in need of special services in the schools: Accuracy of the problem validation screening model*. Unpublished doctoral dissertation, Louisiana State University.

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. (2000). *Functional assessments*. Longmont, CO: Sopris West.

Author Note

The authors thank the principal of Acacia Elementary School, Laurie Emery, the teachers and students, and the district administrators who supported this work.