6-1999

The Use of Curriculum-Based Measurement to Evaluate the Effects of a Remedial Education Program

Michael D. Hixson
Western Michigan University

Follow this and additional works at: http://scholarworks.wmich.edu/dissertations

Part of the Educational Assessment, Evaluation, and Research Commons, and the School Psychology Commons

Recommended Citation
Hixson, Michael D., "The Use of Curriculum-Based Measurement to Evaluate the Effects of a Remedial Education Program" (1999). Dissertations. 1509.
http://scholarworks.wmich.edu/dissertations/1509

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
THE USE OF CURRICULUM-BASED MEASUREMENT TO EVALUATE
THE EFFECTS OF A REMEDIAL EDUCATION PROGRAM

by

Michael D. Hixson

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Psychology

Western Michigan University
Kalamazoo, Michigan
June 1999
THE USE OF CURRICULUM-BASED MEASUREMENT TO EVALUATE THE EFFECTS OF A REMEDIAL EDUCATION PROGRAM

Michael D. Hixson, Ph.D.
Western Michigan University, 1999

This study examined the effectiveness of two Direct Instruction programs: Corrective Reading: Decoding and Connecting Math Concepts. The effectiveness of the interventions was evaluated using a single-subject experimental design with curriculum-based measurement (CBM) probes as the main means of assessment. Other methods of assessment included standardized achievement tests, school reports, and pre- and post-curriculum-based assessment probes. The CBM probes for all students were sensitive to the effects of the interventions, although there was a great deal of variability in the CBM reading probes and little correspondence across methods of assessment. The results indicate that the two Direct Instruction programs were effective, and that CBM probes can be used successfully to evaluate the effects of these interventions.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
ACKNOWLEDGEMENTS

My intellectual development can be crudely summarized as follows: B.F. Skinner's writings motivated me to learn about the science of behavior. Jack Michael taught me the principles of that science, and, Arthur Staats taught me the importance of understanding the cumulative effects of those principles.

Throughout my dissertation and graduate career, Jack Michael has been an invaluable source of guidance and knowledge. A great number of his past students have made important contributions to the science of behavior, I hope I can live up to their precedent. To me, Jack Michael is a behavior analyst of few peers, an intellectual, a health-conscious person, and a friend. I thank Jack for being all of those things because each one has inspired me.

I would also like to thank the other members of my committee: Kristal Ehrhardt, Galen Alessi, and Margaret McGlinchey. Kristal suggested my entry into school psychology, and she has always been a source of knowledgeable and friendly advice. Galen and Margaret gave me sound advice, and both taught first rate courses.

Lastly, I would like to thank my wife and family for their patience and support. I have been championed and sustained by the best people on earth.

Michael D. Hixson
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>History of Direct Instruction</td>
<td>2</td>
</tr>
<tr>
<td>Curriculum Based Measurement</td>
<td>13</td>
</tr>
<tr>
<td>Problems With Traditional Methods of Assessment</td>
<td>13</td>
</tr>
<tr>
<td>Features and Advantages of CBM</td>
<td>15</td>
</tr>
<tr>
<td>Problems With CBM</td>
<td>16</td>
</tr>
<tr>
<td>The Use of CBM in Research</td>
<td>18</td>
</tr>
<tr>
<td>METHOD</td>
<td>23</td>
</tr>
<tr>
<td>Subjects</td>
<td>23</td>
</tr>
<tr>
<td>Setting</td>
<td>24</td>
</tr>
<tr>
<td>Measures</td>
<td>24</td>
</tr>
<tr>
<td>Interobserver Agreement</td>
<td>29</td>
</tr>
<tr>
<td>Procedures</td>
<td>29</td>
</tr>
<tr>
<td>RESULTS</td>
<td>32</td>
</tr>
<tr>
<td>CBM Results</td>
<td>32</td>
</tr>
<tr>
<td>CBA Results</td>
<td>36</td>
</tr>
<tr>
<td>Achievement Test Results</td>
<td>38</td>
</tr>
<tr>
<td>Social Validity</td>
<td>38</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>43</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table of Contents—continued

Implications for Future Research and Weaknesses in the Present Study . . . . 45

APPENDIX

A. Protocol Clearance From the Human Subjects Institutional Review Board . . 48

BIBLIOGRAPHY . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
LIST OF TABLES

1. Math CBA Results for Mary ................................................................. 36
2. CBA Results for Reading Students ......................................................... 37
3. Grade Equivalence for Mary From the Woodcock Johnson
   Tests of Achievement ................................................................. 38
4. Grade Equivalents for Cecil From the Woodcock
   Reading Mastery Tests-Revised ..................................................... 39
5. Grade Equivalents for Tony From the Woodcock
   Reading Mastery Tests-Revised ..................................................... 40
6. Grade Equivalents for Oliver From the Woodcock
   Reading Mastery Tests-Revised ..................................................... 40
LIST OF FIGURES

1. CBM Probes ................................................................. 34
INTRODUCTION

According to both legal mandates and legislation (e.g., IDEA, 1997), children in the United States have a right to a free and appropriate public education. Significant public funds are spent on education because politicians and the citizenry believe that the education of children is of great importance for the continued advancement and development of the culture. However, the low performance of U.S. students in math and science, in comparison to students from other developed countries (e.g., Third International Math and Science Study, 1996), has created concern that this money is not being well spent. The Third International Math and Science Study compared the performance of eighth grade students from 41 industrialized countries in math and science. Statistical analysis found the following results: 20 nations were higher in math than the U.S., 13 were the same, and 7 were below. Japanese students scored significantly higher in math and science than U.S. students, even though U.S. students received more hours of math and science instruction than did Japanese students. The amount of time devoted to homework was approximately equal between the two countries, as was the amount of time spent watching TV. U.S. teachers also worked longer hours than Japanese teachers. One noteworthy difference found between Japanese and U.S. curriculum, is the number of topics presented. In Japan, fewer topics are presented, but each topic is mastered. Previous studies have found similar results. For example, the Report of the National Science Board (1992 as cited in Sagan, 1995) found that of 17-year-olds from thirteen different countries around the world, U.S. students came in last in algebra. However, U.S. students were the highest in one category—confidence. U.S. students were more confident in their knowledge of
math than students from countries who performed significantly better. In the area of reading, across the fourth, eighth, and twelfth grade levels, only 62, 74, and 77 percent of students read at or above a basic reading level (Donahue, Voelkl, Campbell, & Mazzeo, 1998). Sixteen percent of English-speaking young adults cannot read well enough to use a street map or to interpret information on a paycheck. (Worldbook Encyclopedia, 1997). Furthermore, the majority of unwed mothers, welfare dependents, the unemployed, and those arrested have poor basic skills (Berlin & Sum, 1988). Unfortunately, education is often governed by fancy over fact (Engelmann, 1992; Hirsch, 1996). Many popular educational reform approaches have little research support for their effectiveness (National Education Association, 1999). An educational approach that does have significant empirical support is Direct Instruction.

History of Direct Instruction

The curriculum used at the remedial education program that was the focus of the current study, has, as its general name, Direct Instruction (DI). There are DI programs across a variety of academic areas, such as spelling, reading, mathematics, language, comprehension, writing, and many others. DI grew out of the work of Carl Bereiter and Siegfried Engelmann at the University of Illinois in the early 1960s. They worked with disadvantaged preschoolers teaching them reading, language, and math. The children made average IQ gains of 24 points (Engelmann, 1992). In 1968, Engelmann began work with children in the primary grades as part of the world’s largest educational experiment. Sponsored by the Office of Education, the 500 million dollar study was called “Project Follow-Through” (Adams & Engelmann, 1996). Engelmann’s DI group, was but one of 20 sponsors, 9 of which qualified for inclusion in the evaluation. The sponsor’s orientations varied from the loosely structured Open
Education model to the structured Behavior Analysis and Direct Instruction models. The population was composed of over 10,000 students from schools across the United States.

A description of the findings from Project Follow-Through can be found in Research on Direct Instruction: 25 Years Beyond DISTAR (Adams & Engelmann, 1996). Achievement outcomes were compared across sponsors and to non-Follow-Through control schools. The outcome measures were grouped into three general categories: basic skills, cognitive, and affective. The DI model had the highest scores across all three categories and significantly above the controls. In contrast, most of the other models performed worse than the control schools. In the academic areas of reading, mathematics, spelling, and language, the DI model had consistently greater effect sizes than all other models.

In addition to Project Follow-Through, there have been many studies on a much smaller scale that have demonstrated the effectiveness of DI. Gersten (1985) reviewed the research on the effectiveness of DI with special education students and found studies supporting its effectiveness with high risk students, the moderately retarded, and learning disabled students. A meta-analysis of DI research conducted by Adams provides an overall representation of the effectiveness of DI programs (Adams & Engelmann, 1996). Over 350 publications were reviewed for the meta-analysis. For inclusion in the meta-analysis the studies had to meet certain criteria to ensure the validity of the experimental design and the integrity of the intervention. Thirty-seven research articles met the criteria for inclusion. The data were analyzed in a number of ways. The simplest method compared the percentage of studies favoring DI to the percentage that favored the control or other intervention conditions. Eighty-seven percent of the studies favored DI, 0.6% tied, and 12% favored non-DI. A problem with
such a procedure is that the effect sizes are not taken into consideration. Hence, another method of analysis consisted of comparing statistically significant outcomes. This method found that 64% of the studies favored DI, 35% found no difference, and 1% favored non-DI. A more direct method of considering effect sizes is to calculate the average effect size per study. Effect sizes were calculated using the following formula:

\[
\frac{m_{exp} - m_{con}}{sdp} = \text{Effect Size}
\]

In this formula, the difference in means is divided by the pooled standard deviation. Therefore, the effect size is referenced to the normal distribution in terms of standard deviations. Using this method, the effect size of DI was 0.87. Adams reviewed the educational literature and found that no other educational program has come close to this effect size (Adams, in press as cited in Adams & Engelmann, 1996). The effect size of whole language was .08.

The components of DI programs have also been empirically validated (Adams & Engelmann, 1996). That is, the instructional features (see Engelmann & Carnine, 1991) and the presentation techniques have been demonstrated to be important and effective in teaching. The programs are also field tested before publication to ensure that the programs work with actual students. Many programs are not field tested (Engelmann, 1992).

Research has also been conducted on the specific DI curricula used in the present study; that is, Corrective Reading: Decoding (Engelmann, et al., 1988) and Connecting Math Concepts (Engelmann & Carnine, 1992). Gregory, Hackney, and Gregory (1982) found that five months in the Corrective Reading program produced an average of 1.88 year gains in reading ability compared to a 0.2 year gain for the
comparison group. Polloway, Epstein, Polloway, Patton, & Ball (1986) found that the Corrective Reading program was effective with both learning disabled (LD) and educably mentally retarded (EMR) students. The students in this study were administered the Corrective Reading program for an entire school year. LD and EMR students showed significantly greater improvement in reading recognition and comprehension in comparison to gains made in prior years. Campbell (1988) compared 7th and 8th grade low achieving students placed in Corrective Reading to similar students placed in regular English classes. The results of pretest and posttest scores on the Woodcock Reading Mastery Tests showed that students in the Corrective Reading group gained 2.2 years in 9 months of instruction while the control students gained 0.4 years. Similarly, another study found that low achieving students placed in Corrective Reading for one school year gained an average of 2.38 grade levels in word attack skills (Kasendorf & McQuaid, 1986 as cited in Goulter, 1997). Thomson (1992) compared the effects of three interventions for students with specific learning disabilities (SLD). The interventions were: (1) Corrective Reading, (2) Whole Language, and (3) a traditional/basal approach. After a school year of instruction, the students placed in Corrective Reading averaged 5 point gains in standard scores on the word recognition section of the Woodcock-Johnson Achievement Test. The students in the other groups made no gains on standard scores. Students’ reading fluency was also measured by counting the number of words read correctly per minute from a timed passage taken from the Dolch Story Reading Test. Students in Corrective Reading gained an average of 21 words read per minute across the academic year while students in the traditional basal gained 13 words read per minute and whole language students gained 7 words per minute.

Research has also been conducted on the effectiveness of the Connecting Math
Concepts (CMC) program (i.e., the math program used in the present study). Tarver and Jung (1995) compared CMC to a discovery learning math program entitled Math Their Way, with students in the first and second grades. The Comprehensive Test of Basic Skills—Mathematics (CTBS-M) was used as a pretest-posttest measure of gains. Students in CMC made significantly greater gains on the CTBS-M than Math Their Way students. At the end of second grade, CMC students obtained a total grade equivalent of 4.5 on the CTBS-M, while Math Their Way students obtained a total grade equivalent of 3.3. At the end of second grade, there were relatively fewer CMC students below average and relatively more CMC students above average; demonstrating that CMC benefitted both high- and low-performing students. On a measure of attitudes toward mathematics, students at the end of second grade in CMC indicated a greater positive attitude toward mathematics than students in Math Their Way.

Vreeland et al. (1994) found that disadvantaged children using CMC in third grade achieved average to above average rates of progress in both math calculation and application on the Kaufman Tests of Educational Achievement. They also found that on the Iowa Tests of Basic Skills CMC student scores stayed at about the 50th percentile in mathematics while student scores in other classrooms tended to decline. As a result of these findings, CMC was implemented in other classrooms and similar results were reported.

Snider and Crawford (1996) compared math instruction using a traditional math basal textbook published by Scott Foresman (SF) to CMC. The subjects were fourth grade students randomly assigned to either a SF or CMC classroom. Four dependent variables were used to measure student progress in a pretest-posttest experimental design. One assessment instrument was the National Achievement Test (NAT) (1989).
The NAT is a standardized test battery that contains three math sections—computation, concepts, and problem solving. The concepts and problem solving section are combined for scoring. The second assessment instrument was a math facts test that assessed recall of 72 multiplication facts under timed conditions. The third instrument was the cumulative tests taken from the SF program, and the fourth instrument was an experimenter-made test based on the topics covered in CMC. The third and fourth instruments were both types of curriculum-based assessments because they were based on the students’ curriculum. The overall score on the NAT showed no differences between groups, nor were differences found on the concepts and problem-solving subtest scores. The NAT computation subtest score did show significantly better performance for the students in CMC. CMC students performed significantly better on the math facts test than did SF students. CMC students also performed significantly better on both curriculum-based measures. It is particularly noteworthy that CMC students performed better than SF students on the cumulative tests from the SF program. Why there were few differences on the NAT is not clear. One problem that has been reported with standardized mathematics achievement tests is that they frequently fail to sample the concepts taught in mathematics curricula (Howell, Fox, & Morehead, 1993). Hence, mathematics achievement tests may be particularly insensitive to measuring the effects of interventions.

The effectiveness of DI programs is not only attributed to their being field-tested prior to publication but also because the curricular designs are based both on an empirically-validated theory of instruction (Adams & Engelmann, 1996; Engelmann & Carnine, 1991) and on research specific to each academic domain. For example, in the case of reading, the Corrective Reading program teaches skills that research has identified as important in learning to read, and the program targets the specific problems
of poor readers.

The Center for the Future of Teaching and Learning has identified many variables that are important in learning to read. The Center is a nonprofit organization that disseminates information on effective educational practices. In particular, the Center has focused on the research sponsored by the National Institute of Child Health and Human Development (NICHD). NICHD sponsored research is characterized by a high degree of scientific rigor (see the Center for the Future of Teaching and Learning, 1996). NICHD sponsors research questions that are studied using many children across different research sites. Only replicable findings become part of the knowledge base on how children learn to read. The Center for the Future of Teaching and Learning (1996) described the main implications of this research for early reading instruction as follows:

"1. Begin teaching phonemic awareness directly at an early age (kindergarten)."

(p. 39) Phonemic awareness is the ability to recognize individual sounds in words. The NICHD research shows that this skill does not develop naturally, rather, it is learned. The Corrective Reading program (as well as all other DI reading programs) teach phonemic awareness using a variety of techniques. For example, early in the program, children are taught to identify the first sound, middle sound, and last sound of words the teacher presents orally.

"2. Teach each sound-spelling correspondence explicitly."

(p. 40) Explicit instruction means that phonemes are taught in isolation. For example, the learner is presented with the letter "I," and taught to say /II/. This method is more effective than having children learn sound-spelling correspondences in the context of whole words and sentences. The Corrective Reading series explicitly teaches sound-spelling correspondences. For example, students are taught that when the letters "ol" are in a
word, they frequently make the sound /ol/ as in the word “old.”

“3. Teach frequent, highly regular sound-spelling relationships systematically.”

(p. 40) The curriculum should focus on the sound-spelling relationships that occur most frequently, and these relationships should be coordinated with material the children are asked to read. For example, in the Corrective Reading series, after children are taught to read “ol,” they receive immediate practice in reading words with this sound, such as “old,” “gold,” “hold,” and so on.

“4. Show children exactly how to sound out words.” (p. 40) Students should not only be taught the sounds in isolation, but also how to blend those sounds together to make a recognizable word. For example, in the very first lesson of Corrective Reading Decoding A, children are first taught the spelling-sound correspondences for the letters “s”, “a”, “t”, “e”, and “m.” Later in that same lesson, they are taught to sound out the syllables “eem,” “me,” “ma,” “am,” “at,” and “eet.” Children get immediate practice in blending the sounds together. They are explicitly taught how to sound out the words without stopping between the sounds. The word “me” would be sounded out as /mmmeee/ rather than as /mmm_____eee/. Then they are taught to say the word the fast way (i.e., /me/).

“5. Use connected, decodable text for children to practice the sound-spelling relationships they learn.” (p. 40) Many programs teach some sound-spelling correspondences and then present sentences and stories that contain words that are undecodable, or conflict with what the children have learned. Less decodable text teaches children to use strategies other than sound-spelling correspondences to read words. Such children often resort to using the context of the sentence or pictures to identify words. Stanovich and Stanovich (1995) summarize the research on the use of context in reading acquisition:
Thus, the key point that top-down theorists got wrong, and the key error that has been perpetuated by the inheritors of the top-down view—the whole language movement—is the assumption that contextual dependency is always associated with good reading. In fact, the word recognition skills of the good reader are so rapid, automatic, and efficient that the skilled reader need not rely on contextual information. In fact, it is poor readers who guess from context—out of necessity because their decoding skills are so weak. (p. 92)

And, according to the Center for the Future of Teaching and Learning (1996):

In the NICHD intervention studies...teaching children to use context and prediction as strategies for word recognition resulted in greater numbers of reading disabilities than instruction that taught children to use their sound-spelling knowledge as the primary strategy for word recognition. (p. 39)

How the Corrective Reading series addresses these issues will be discussed after covering the final implication of the NICHD research.

“6. Use interesting stories to develop language comprehension.” (p. 41) The use of interesting stories is not ruled out by the NICHD research, but the stories must use decodable text and students must receive explicit instruction in developing comprehension. During the early stages of reading acquisition, the stories will be very simple due to the small number of words that are available for stories. More complex language comprehension skills can be developed by having the teacher read stories to the students. While the focus in the Corrective Reading: Decoding series is on decoding rather than comprehension, comprehension activities are included. Initially, students answer questions about a story that the teacher reads. Later, they answer questions about stories that they have read.

The Corrective Reading series specifically addresses the needs of the problem reader, and relatedly, the fifth implication from the NICHD research. The problem reader has learned to use context and word-guessing strategies to read. This is addressed by teaching students decoding strategies that are initially applied to words in isolation (i.e., word lists). Word lists force students to rely on decoding strategies. Students then read these words in sentences and passages. According to Engelmann et
al. (1988):

The sentences and passages are designed so they are relatively easy if the student approaches words as entities that are to be analyzed according to the arrangement of letters, but difficult if the student guesses on the basis of the context or syntax of the sentence. (p. G4)

This does not mean that students do not use comprehension strategies to read, but the comprehension strategies are applied after the word is decoded. In other words, the student first decodes the word, then the student determines whether the decoded word makes sense (i.e., is it a real word that would be used in this context). This is the opposite strategy that most problem readers use and the opposite of what is taught in many commercial reading programs where students first try to identify a meaningful word before attempting to decode the word.

The teaching methods and concepts in CMC are also based on an empirically-validated theory of instruction and on the mathematics instruction research literature. The Third International Math and Science Study (1996) found that Japanese text books introduce fewer topics than U.S. texts. This is because most U.S. math curricula are "spiral" programs; that is, students work exclusively on a particular topic for a few lessons. Then a new topic, often unrelated to the preceding one, is presented. The previous topic may not resurface until much later in the school year or the following year. This teaching method is justified by curriculum developers who believe that children learn particular math topics when they are "developmentally ready" (Engelmann, 1992). In contrast, math topics are organized into tracks in CMC. A track is an ongoing development of a particular topic. Within each lesson, work from three to five tracks is presented. For example in CMC Level D, students begin work on fraction representation problems in lesson 1 and continue work on these problems through lesson 102—this is one track. Skills are built systematically, cumulatively, and with enough practice to help ensure mastery.
The program is called *Connecting Math Concepts* because related concepts are taught in related ways (i.e., the students are shown the “connections” among concepts), and the range of examples shows how concepts are the same. For example, in teaching fractions, a program that initially presented the following examples: 1/2, 1/3, 2/3, 2/4 is likely to teach students the distorted view that fractions are numbers less than one. In CMC, fractions less than, equal to, and greater than one are initially presented. Students are specifically taught that the top number (numerator) tells the number of parts in each group and the bottom number (denominator) tells the number of parts you use. Students learn this rule when the full range of examples are presented. Students also learn the relationships between fractions, ratios, division, and decimals. Students learn number families which demonstrate the commutative property and the relation between addition and subtraction, and the relation between multiplication and division. For example the number family 3—4—>7 is used to teach all of the following relations: 3+4=7; 4+3=7; 7-3=4; and 7-4=3. Students are taught to find any one of the three numbers that compose a number family. The number family format is then used to solve word problems. For example, to represent all subscriptions in a household, one could write: magazines----newspapers-->subscriptions. The student would then plug the specific values in the word problem into the number family. There are many other examples in CMC that illustrate this teaching method (see Carnine & Engelmann, 1990). Research has demonstrated that using such principles is more efficient and produces better concept learning (Engelmann, Carnine, & Steely, 1991).

Overall, DI programs have a strong empirical base on multiple levels. First, the programs are based on a theory of instruction whose principles have been empirically validated (Adams & Engelmann, 1996; Becker, 1992; Engelmann & Carnine, 1991). Second, the programs are field-tested to ensure student learning. Third, as previously
discussed, a great deal of research has been conducted demonstrating the superiority of DI programs overall and of the specific programs used in the present study. Last, DI programs target skills and incorporate teaching strategies that have been identified as important in the research literature.

**Curriculum-Based Measurement**

Not only is the current experiment an evaluation of DI as it is used in a remedial education program, but it is also an evaluation of curriculum-based measurement (CBM). CBM is considered “best practice” among many school psychologists (e.g., Shinn, 1989, 1995). There are a number of reasons for this advocacy of CBM, which will be discussed shortly. The development of CBM can be attributed to the work of Deno, Fuchs, and Mirkin with the Minneapolis public schools (e.g., Deno & Mirkin, 1977; Fuchs & Deno, 1981). Their major concern was the development of assessment procedures to measure individual student progress: “Where educational placements were once made to serve the majority interests, placements now must be made to serve the needs of the individual student” (Deno & Mirkin, 1977, p. 5). This focus on the individual student and individual progress monitoring made clear that the traditional forms of assessment in schools were inadequate.

**Problems With Traditional Methods of Assessment**

There are a variety of measures that could be used to assess academic performance. One is teacher report. An obvious problem with this measure is the potential subjectivity of the report. Teachers may have different student expectations or biases in their reports. However, research suggests that teachers can accurately distinguish nondisabled from disabled students (Gresham, MacMillan, & Bocian,
1997), but, traditionally, the academic problems reported by teachers are further assessed by a school psychologist using standardized testing methods with published tests. The interpretation of test scores often varies from state to state and district to district. In many cases, more than one test is administered and the relationship or discrepancy between the tests is considered important. In the simplest case, a low score on a standardized published test is used as one basis for classifying the child as disabled. There are a number of potential or definite problems with the use of such tests, as has been indicated by Marston (1989). First, published tests sometimes fail to sample the skills taught in the curriculum. Jenkins and Pany (1978 as cited in Marston, 1989) found that grade scores using published tests varied widely as a function of the reading curriculum. Second, the results of published tests are often of little use for instructional planning. In the case of mild disabilities, the same treatment goals and strategies are used, regardless of the specific diagnosis (e.g., learning disability, mild mental retardation, or behavior disorder) (Reschly & Ysseldyke, 1995). Third, many published tests fail to assess the rate of emission of academic behaviors (i.e., so-called "behavior fluency"). As will be seen, traditional CBM directly measures the rate of behavior across skill types. Fourth, at best, published tests allow for a pretest-posttest assessment of skills. A more frequent measure of skills would allow a more dynamic and rapid interchange between the data and educational decision-making. Last, published tests are often insensitive to skill acquisition. For example, Marston, Fuchs, and Deno (1986) found no change on a pretest-posttest, but significant change using CBM. Interestingly, the CBM was more positively correlated with teacher report than the results of the standardized published test.
Features and Advantages of CBM

The focus of CBM on the assessment of individual behavior resulted in a number of unique features as identified by Marston (1989):

[CBM had to be] (1) tied to a students' curricula, (2) of short duration to facilitate frequent administration by teachers/educators, (3) capable of having many multiple forms, (4) inexpensive to produce in terms of time in production and in expense, and (5) sensitive to the improvement of students' achievement over time. Another necessary characteristic was the identification of academic behaviors in the basic skill content areas that educators could measure reliably and validly. (p. 30)

Shinn and Bamonto (1998) also said that "CBM can be used more comprehensively in a Problem-Solving model to make a variety of decisions." (p. 5) They also pointed out, as did Marston, that the focus of CBM is the assessment of basic skills:

The measures are designed only to assess student performance in the basic skills areas of reading, spelling, mathematics computation, and written expression. The measures were not designed to assess student performance in content areas such as science or even to assess a student's reading skills in the written materials in these subjects. (p. 7)

The focus on basic skills has implications for the use of CBM that will be discussed later.

CBM measures are tied to student curricula by using the material from the curricula as assessment material. For instance, if the evaluator is interested in the reading performance of a student, reading passages from the student's curriculum would be used in the assessment. Using curricular materials for the assessment also means that the assessment tools are inexpensive, easy to produce, and permit the development of multiple forms. The CBM's that will be discussed in this paper have standardized methods for administration and scoring, and have been demonstrated to be both reliable and valid measures of academic behavior (Marston, 1989).

The following are CBM's that have been developed for specific content areas:
In the area of reading, students read aloud from a basal reader for 1 minute. The
number of words read correctly per minute constitutes the basic metric. In spelling, students write words that are dictated at specified intervals for 2 minutes. The number of correct letter sequences and words spelled correctly is counted. In written expression, students write a story for 3 minutes after being given a story starter. The number of words written, spelled correctly, and/or correct word sequences is counted. In mathematics, students write answers to problems on worksheets for 2 to 5 minutes. The number of digits written correctly is counted.

Ideally, the implementation of a sensitive repeated measure should result in improved student achievement, at least when the assessment device is used within the context of a problem-solving model. In the research conducted by Fuchs, Fuchs and colleagues, they found improved student achievement after instructing teachers in the use of CBM (Fuchs, Fuchs, & Hamlett, 1989; Fuchs, Fuchs, Hamlett, & Stecker, 1991). However, this improved student performance is likely due to more than the collecting and graphing of the CBM data; teachers must also review the specific types of student errors and implement appropriate instructional modifications based on those errors (Fuchs, Fuchs, Hamlett, & Allinder, 1991).

Problems With CBM

If CBM is intended as a general measure of the aforementioned academic areas, then the measures are only samples of those domains, and are, therefore, inferential—just like standardized published tests. Also, CBM probes sometimes sample skills that correlate with, but are not the same, as the skills of interest. For example, reading rate is not the same as reading comprehension, and reading comprehension is a key concern. Sometimes, CBM reading probes are supplemented with reading questions in order to assess comprehension, but standardized procedures
have not been developed for creating and scoring the questions (Fuchs, 1989).

Research has shown a correlation between reading rate and comprehension (Deno, Mirkin, & Chiang, 1982), but, clearly, reading rate and reading comprehension are not the same. A similar argument can be made for the written expression CBM: the number of words written correctly per minute is not the same as writing a good story. In spelling, the number of correct letter sequences and the number of words written correctly are direct measures of the behavior of interest. In math, the number of digits written correctly is also a direct measure, but not under all conditions. Story problems, geometry, and time are skills taught in many basal curricula, but their relation to math CBM probes is more tenuous than the addition, subtraction, multiplication, and division skills. Also, digits correct does not directly assess the strategies and algorithms involved in much of mathematics.

In behavior analysis, direct measures of the behavior of interest are most often taken. CBM is supposed to be the same:

It [CBM] borrows the operant research methodology of repeated behavior sampling and time-series analysis. Employing direct and frequent evaluation, a teacher collects repeated, short samples of a student’s behavior within the curriculum, over a time period, and under different teaching strategies. (Fuchs, & Deno, 1981, p. 6)

But, as the previous argument pointed out, CBM probes are not always direct and complete measures of the behavior of interest. This situation may be inevitable in education, where students are rapidly acquiring a variety of new complex behaviors. Or, CBM probes can be developed that target each specific behavior, for example, if the sounds of the letters “a,” “r,” and “b” are being taught, CBM probes can be developed specific to these skills (Fuchs, 1989). But standardized procedures have not been developed at these more specific levels. Howell, Fox, and Morehead (1993) describe curriculum-based assessment (CBA) methods that are direct and have a greater level of
specificity. But, while these methods make logical sense and have some empirical validity, they have not been standardized and validated to the same degree as the traditional CBM's. CBM refers to the specific assessments discussed previously—CBA is a broader term that refers to assessment procedures whose focus is on the identification of specific skills and deficits for the purpose of determining intervention components (Shinn, 1995). Unlike CBM, CBA probes are not typically administered repeatedly for the purpose of progress monitoring.

Shinn and Bamonto (1998) do much to clear up the confusion. They agreed that CBM measures will not replace achievement tests because CBM is limited to the assessment of basic skills. However, they consider basic skills to be an integral part of the elementary curriculum. Therefore, the assessment of basic skills using a sensitive measurement system is necessary. They also argue that CBM’s utility is in the context of a problem-solving model, in which the academic problems of individual students are being addressed. The authors do concede that CBM is inferential:

The selection of potential dynamic indicators was driven by an interest in measuring behavior as holistically as possible; that is, attempts were made to select terminal behaviors that included as many of the important subcomponents, so that by directly assessing the global skill of interest, inferences could be drawn about a number of specific behaviors. (p. 20)

These issues will come up again in the Discussion section of the current paper as they relate to the DI curriculum and the results of the current study.

The Use of CBM in Research

Because of the utility of CBM in assessing the performance of individual students over time, it has functioned as a useful dependent variable in research. For example, Marston (1987-1988) evaluated the effectiveness of special education for 11 mildly handicapped students using CBM reading scores (i.e., the number of words
read correctly per minute). The students’ performance in general education functioned
as the baseline condition. CBM probes were administered once per week for
approximately 7 weeks in general education, and then for another 14 weeks
(approximately) in special education. The slopes across the two conditions were
compared, and it was found that there was a greater positive slope in special education
for 10 out of 11 students—suggesting that the special education services were effective.
Special education would be expected to have a gradual, rather than an immediate, effect
on the CBM probes. This makes collecting a greater number of data points under each
condition, and comparing the difference in slopes between conditions, important
components of the experimental design and analysis.

For interventions with more immediate effects, CBM probes can be used with
relatively fewer observations per condition. This is the case in drug evaluation research.
For example, one study demonstrated the utility of using CBM probes to evaluate the
effects of methylphenidate on two students diagnosed with attention deficit
hyperactivity disorder (Stoner, Carey, Ikeda, & Shinn, 1994). CBM math and reading
probes were used across baseline, placebo, and three dosage levels of methylphenidate
to evaluate the effects of the drug. For both students, an optimal dose was found. But
in this study, and in many other studies that have used CBM probes in a single-subject
design, there is considerable variability in the probes within conditions. For example,
in the baseline condition for one of the subjects, the CBM math probes varied from
approximately 9 to 35 digits correct.

The variability in CBM probes may be inherent to them because, unlike the
traditional discrete behaviors that are typical of behavior analytic research, CBM probes
(a) compile a number of discrete behaviors into one observation and (b) the behaviors
measured and the controlling variables are slightly different on each administration. For
instance, in CBM reading probes, the passages measure a variety of phonic and sight-word reading skills. While a rat’s lever press can be broken into simpler responses, the number of responses measured in a reading probe appears much greater. Also, different passages are used for each administration—which is good, otherwise, any change in the reading score could be attributed to practice. But, this means that different skills (i.e., different word reading and phonetic skills) could be measured on each probe, hence it is not surprising that the scores are variable. For CBM reading probes, some researchers screen the probes to ensure that the readability of the passages (as calculated using readability formulas, such as the Flesch-Kincaid) is similar across passages. But the criterion is never so stringent that the passages must all have the same readability score, and even passages with the same readability can vary in their difficulty for a particular child. Therefore, students will be reading passages that have different levels of reading difficulty. In spite of this, the studies reviewed so far, and many others, have documented that CBM probes are sensitive to treatment effects under many conditions, although Stoner et al. (1994) warn that “it would be more appropriate to search for and control the sources of such variability in current or recent stimulus conditions” (p. 111). Indeed, the variability in the CBM probes in Stoner et al.’s research made interpreting drug effects more difficult.

A literature search found only one article that specifically evaluated the effects of DI using a CBM measure (Marston, Deno, Kim, Diment, & Rogers, 1995). The authors did not use a single-subject experimental design. Instead, they considered the first CBM reading probe prior to the interventions as the pretest score, and the last CBM score as the posttest score. They then compared gains in the number of words read correctly per minute across six interventions, one of which was DI. This study found only modest positive effects of DI; the students gained an average of 9.1 words
per minute over the 10 week study. The CBM reading probes were taken from another curriculum—not the DI curriculum. Also, it is not clear how well the teachers implemented the DI program. Other studies have used CBM probes within a DI curriculum, but in these studies the main independent variable was something other than the DI curriculum. For example, one study evaluated the effects of a re-reading contingency and reward system on CBM reading probes taken from the DI curriculum (Holz, Peck, McLaughlin, & Stookey, 1996).

The purpose of the present study is to evaluate DI in the context of a remedial education program using CBM probes. Students needing either reading or math remediation participated in the study. The DI math curriculum is called *Connecting Math Concepts* (CMC) (Engelmann & Carnine, 1992), and the DI reading program is called *Corrective Reading: Decoding* (Engelmann, Carnine, & Johnson, 1988). The reading CBM probes were taken from the Macmillan basal curriculum (Arnold & Smith, 1987). The CBM reading probes were not made from the DI reading curriculum for a number of reasons:

1. All of the reading students placed in Decoding level A, and this level does not have reading passages.
2. The DI reading passages from other levels are highly related (in terms of the specific words read) to the lessons.
3. Of greater interest was how the skills learned from the DI curriculum would generalize to a standard reading basal; especially since it is the main goal of the remedial education program (described later) to assist students in performing better in their respective schools. Scores from CBM reading probes using different curricular materials are highly correlated (Fuchs & Deno, 1992). Therefore, using reading probes made from another curriculum is justified. Unlike the CBM reading probes, the CBM
math probes were based on the objectives found in the CMC curriculum.

The present study is also an evaluation of CBM. Perhaps DI reading curricula and CBM reading probes taken from a traditional basal reader are too different from one another for much generalization to take place. In the case of math, CMC teaches much more than the basic skills that the CBM math probes measured. Skills such as time, geometry, and story problems are also taught. Therefore, the sensitivity of the math probes to the intervention was of interest.
METHOD

Subjects

The participants were four first-time students at the remedial education program. One student was referred for low grades in math, and the other three for reading problems. Mary was ten years old and in fifth grade. She scored at level D (approximately a fourth grade level) on the CMC placement test. Cecil was a 9 year old male in third grade who was receiving extra services in school for speech and reading, and he received ritalin three times per day. His parents said that he sometimes “shut down,” that is, he would stop working. He scored at level A (the lowest level of the Corrective Reading program) on the Corrective Reading placement test. Tony was an 8 year old male in a second grade special education class who received special services for reading. According to his parents, Tony becomes frustrated very quickly. He was receiving dexadrine for ADHD and tegretol for seizures. Tony scored at level A on the Corrective Reading placement test. Oliver was an 8 year old male in second grade. His parents said he was approximately 1 year behind in reading. Oliver scored at level A on the Corrective Reading placement test.

For the purpose of a pretest-posttest comparison, the students were administered achievement tests related to their area of academic difficulty. Mary took the three math sections of the Woodcock-Johnson Tests of Achievement. Her grade-equivalent scores were: Calculation, 5.4; Applied Problems, 4.9; and Quantitative Concepts, 5.5. The test was administered in January; therefore, because she was in fifth grade, her scores were close to grade level on all sections except Applied
Problems. The reading students were administered the *Woodcock Reading Mastery Tests-Revised* (WRMT-R).

Cecil obtained the following grade-equivalents: Word Identification, 1.4; Word Attack, 1.4; Word Comprehension, 1.3; Passage Comprehension, 1.4. All grade-equivalents were significantly below the student's 3rd grade level. Tony obtained the following grade equivalents: Word Identification, 1.3; Word Attack, K.9; Word Comprehension, 1.4; Passage Comprehension, 1.2. Because Tony was over half-way through second grade, all grade-equivalents were at least 1 year below grade level. Oliver obtained the following grade equivalents: Word Identification, 2.3; Word Attack, 2.0; Word Comprehension, 2.4; Passage Comprehension, 1.9. These scores were several months below grade level.

Setting

Project Help is an after school remedial education service sponsored by the School Psychology Program and the Department of Psychology at Western Michigan University. Tutoring sessions ran for two hours Monday through Thursday in accordance with the WMU semester schedule. Students received a 15-minute break approximately half-way through the session. The students were taught by trained tutors, most of whom were majors or minors in psychology. During the first two weeks of the semester, the tutors were taught the DI teaching procedures, and they were supervised by school psychology graduate students throughout the semester.

Measures

CBM reading and math probes were administered weekly prior to and during the Project Help semester. This is an AB single-subject design. Administering the
probes repeatedly both prior to and during Project Help is necessary in order to detect the variability and trends within the CBM probes. A single data point prior to Project Help would be insufficient. In the case of reading, the reading probes were taken from the *Macmillan Connections Reading Program* (Arnold & Smith, 1987). To determine the grade level of reading probes each student should be administered, curriculum-based assessment methods were followed as described in Shinn (1989). This method consists of administering a set of three reading probes from each level of the basal (i.e., pre-primer, primer, 1st grade, 2nd grade, and so on), until an instructional level of material is found. Students read each passage for one minute, and the number of words read correctly and incorrectly per minute is recorded. The median score on the three probes is compared to certain criteria to determine each student’s reading ability on passages at that level. The criteria used in the present study are outlined in Shapiro (1996). Based on the number of words read correctly and incorrectly per minute, reading ability is categorized as frustrational, instructional, or mastery. Because all three of the reading students were in third grade or below, and because the students were described as very low readers according to their parents, they were first administered the pre-primary level probes. According to the Shapiro criteria, the instructional level is 40-60 words read correctly per minute. All of the students read below 40 words per minute on the pre-primary level probes, therefore they were not administered more difficult passages. This level of probe is called “frustrational” for all students because it was read below 40 words correct per minute. Because of these results, all of the probes administered throughout the study were at the pre-primary level—the easiest level that is available. Prior to administering the reading probe, students were given the following standard instructions (Shinn, 1989):

> When I say, “start,” begin reading aloud at the top of this page. Read across the page [demonstrate by pointing]. Try to read each word. If you come to a word
you don’t know, I’ll tell it to you. Be sure to do your best reading. Are there any questions? Start.

The examiner followed along on a separate copy and marked the words read incorrectly. Additions, inaccurate decoding, omissions, and pauses of 3-seconds were counted as errors. If students paused for three seconds, they were told the word. After 1 minute was up (as indicated by a beeping 1-minute timer), they were asked to stop reading. The number of words read correctly and incorrectly per minute was the main dependent variable.

Shapiro’s placement criteria are sometimes used by school psychology practitioners and researchers (e.g., Noell et al., 1998; Shriver, 1999). Shapiro (1996) cites Fuchs and Deno (1982) for the criteria, but this author was unable to find these criteria in Fuchs and Deno (1982). In an earlier paper, Fuchs and Deno (1981) list several criteria very similar (but not identical) to that reported by Shapiro (1996). Fuchs and Deno (1981) tested seven different placement criteria, many of which were derived from the work of others (e.g., Haughton, 1972). Shapiro says of the Fuchs and Deno (1982) placement criteria: “The recommended reading levels of Fuchs and Deno for mastery, instructional, and frustrational levels are based on a ‘best guess’ approach for instruction.” (p. 116) Despite the seemingly wide-spread acceptance of the criteria cited by Shapiro, they should probably be much more cautiously used and interpreted—especially since what “best guess” means is not explained. Shapiro says that a better method is to develop placement criteria from local norms. Despite the lack of justification for these criteria, they were used in the present study. This was because practicing school psychologists and researchers have found these placement criteria useful. Also, the development of local norms is not feasible with the students at Project Help. The students come from different school districts, and, more importantly, Project Help does not get average and above average students, and scores from these students
are needed to determine an average reading ability.

A different set of reading CBA probes, from the same pre-primary book as the first set of probes, was administered near the end of the Project Help semester. The median score on the three probes was used to measure gains in reading rate from the first CBA reading probes. Mary was administered math CBA probes at the beginning and near the end of the Project Help semester. The CBA probes assessed digits correct and incorrect per minute on single digit multiplication, simple division (no remainders), and 2 digit minus 1 digit subtraction problems. Like the CBA reading probes, three probes of each type were administered and the median score was recorded.

In the case of the one math student, the CBM math probes were developed based on the objectives of the CMC level D program. Some subjectivity is involved in this process. The main guideline is that the probes should assess basic skills that can be measured by digits correct per minute. Math probes (i.e., worksheets) were made that consisted of a total of 25 problems of the following types:

1. Four addition problems with numbers from 10 to 999.
2. Two subtraction problems that did not require borrowing with numbers from 1 to 999.
3. Two subtraction problems that required borrowing with numbers from 1 to 999.
4. Four multiplication problems with numbers from 1 to 99.
5. Three evenly divisible division problems with a dividend from 10 to 99 and a maximum divisor of 10.
6. Three not-evenly divisible division problems with a dividend from 10 to 99 and a maximum divisor of 10.
7. Two fraction addition problems with common denominators and numerals 10
or less.

8. Two fraction subtraction problems with common denominators and numerals 10 or less.

9. Three fraction multiplication problems with numerals 10 or less and not necessarily common denominators.

Unlike the reading probes, the student was given as much time as necessary to complete the math worksheets. This is a less common practice. Typically, students are given 2 to 5 minutes to complete as many problems as possible. But measuring the time to complete an entire worksheet is a variation that is recommended in the literature (Shapiro, 1996). Because the math probe worksheets assessed multiple skills, this allowed for a sort of multiple baseline across behaviors, which may not have been possible if the student had not been given the opportunity to attempt all of the problems.

Before administering the worksheet, the participant was given the following standard directions:

There are several types of problems on the sheet. Some are addition, some are subtraction, some are multiplication, and some are division. Look at each problem carefully before you answer it. When I say “start,” turn over the page and begin answering the problems. Start the first problem on the left on the top row [point]. Work across and then go to the next row. If you can’t answer a problem make an “X” on it and go to the next one. Work quickly and carefully. Are there any questions? Start.

These directions are nearly identical to those recommended by Shinn (1989). The number of digits correct and incorrect was the main dependent variable.

All students were administered achievement tests during the first and last two weeks of Project Help tutoring. The reading students were administered the Woodcock Reading Mastery Tests-Revised (WRMT-R), and the math student was administered the mathematics sections of the Woodcock-Johnson Tests of Achievement (i.e., tests 24, 25, and 33). For the reading students, a different form of the test was administered.
on the second occasion. All tests were administered by school psychology graduate
students trained in psychoeducational assessment.

Interobserver Agreement

A tape recorder was used to record the reading probes. Agreement was assessed
by having a second observer score a sample of each student’s recorded reading probes.
Agreement on the math probes was assessed by having a second observer score a
sample of the completed math worksheets. Agreement was assessed on 33% percent of
the CBM reading probes (range of 30-36% across students) and on 40% of the CBM
math probes. The percentage of agreement was calculated by dividing the number of
agreements by the number of agreements plus disagreements and multiplying by 100%.

Procedures

Probes were administered once per week for all students. Ideally, all of the
probes would have been administered in the same setting because research has
demonstrated that the setting does have a small but significant effect on CBM scores
(Derr-Minneci, & Shapiro, 1992). But this was logistically impossible for two
students. All but two probes were administered at Project Help for Cecil, and all but
one probe was administered at Project Help for Tony. All of Oliver’s probes were
administered at Project Help, and all of Mary’s probes were administered at her house.
The probes were administered at the beginning of the tutoring sessions for the three
students who received the probes at Project Help. The other student received the probes
at her home immediately after Project Help. Each student received the probe on the
same day of the week, unless the student was absent in which case the probe was
administered the following day. The directions to the students and administration
procedures for the CBM probes were the same as those for the CBA probes described earlier.

DI lessons are scripted. This means that everything that the tutor does and says is explicitly stated in the teacher presentation book. The reading students were placed in Corrective Reading: Decoding A. The lessons are divided into parts. Typically, the lessons begin by teaching and reviewing letter and letter-combination sounds. For example, when the letter "a" is touched, the student is taught to make the /aaa/ sound as in the word "apple." The next section of the lesson typically teaches phonemic awareness skills, the ability to recognize that words can be broken into different sounds. One exercise that is used to teach this is to say three words, for example "mit," "mat," and "met," and ask the students to say the word that had a particular middle sound. For example, "Which word has the middle sound /aaa/," for which the correct answer is "mat." The next portion of the lesson typically teaches the students to sound out words and to read the words the "fast way" (i.e., to read each word at a normal reading rate). Next, students do workbook activities that involve writing sounds, reading words, writing words, reading sentences, and changing words to make new words. There are also activities to strengthen visual search and recognition skills. The lessons gradually introduce new sounds and words. Students must demonstrate correct responding on each skill that is taught in the lesson. This is done to help ensure mastery on all skills. A key component of this is the correction procedure. When students make an incorrect response, they are immediately told the correct response, they make the correct response, and then they start that section over. Students are also periodically administered mastery tests. Lessons may be repeated based on the mastery test results.

For the student placed in the Connecting Math Concepts level D program, the lessons began with teacher-directed activities. A number of skills were taught and
reviewed. Even in the teacher-directed portion of the lesson, the student received considerable practice in working problems. After the teacher-directed activities, the student did independent work on the skills learned during the lesson. As in all DI programs, the lessons follow a hierarchical skill structure, and concepts are systematically reviewed to help ensure mastery on all skills.
RESULTS

Mary completed 63 lessons in CMC in approximately 53 hours of actual instructional time (i.e., subtracting breaks and other Project Help activities). Cecil completed 58 lessons of *Corrective Reading: Decoding A* in approximately 54 hours of actual instructional time. Tony completed 63 lessons of *Corrective Reading: Decoding A* in approximately 56 hours of actual instructional time. Oliver completed the greatest number of lessons. He went through all the lessons of *Corrective Reading: Decoding A* (65 lessons) and 8 lessons of *Corrective Reading: Decoding B* for a total of 73 lessons. He received approximately 62 hours of actual instruction.

CBM Results

The main dependent variable for all subjects was the weekly CBM probe scores (Figure 1). Interobserver agreement on the CBM reading probes was 99.6% for correct responses and 97% for incorrect responses. Interobserver agreement on the CBM math probes was 99% for correct responses and 99.6% for incorrect responses.

All of Mary's math probes were administered at her house. Mary's digits correct per minute (DCPM) increased from baseline through Project Help, but the increase during Project Help could have been due to the increasing trend that began in baseline. However, the increase in baseline may not have been due to an increase in skill acquisition. Informal observations of Mary suggested that during baseline she was learning how to do the CBM worksheets more rapidly by quickly putting an "x" on problems she was unable to solve. In other words, initially she pondered over each problem, but she soon learned to recognize problems she could not answer—mainly
those involving division and the multiplication of fractions. This hypothesis is supported by considering that during baseline and at the beginning of Project Help both correct and incorrect digits per minute were increasing. Starting at week 10, Mary began correctly solving fraction multiplication problems, and starting at week 12 she began correctly solving multiplication problems that required carrying and those involving double digits. Mary had been taught how to solve fraction multiplication problems starting at week 10 and how to solve multiplication problems with carrying starting at week 9. Tutor notes indicated that Mary had difficulty learning how to solve multiplication problems with carrying and those involving double digit multiplication. This may explain why correct solutions to these problems did not occur until week 12. Up to week 9, Mary either left these types of multiplication problems blank or solved them incorrectly. For example, her answer to the column multiplication problem "54 x 51" was "254." This type of error was made consistently through week 8. Starting at week 9, she left all of these problems blank, until she began correctly solving them at week 12. Correctly solving fraction multiplication and multiplication problems with carrying are the main factors that resulted in the decrease in digits incorrect per minute (DIPM). Mary never got to the lessons in CMC that taught how to solve long division problems. She, therefore, remained incorrect on these problems throughout the study.

Cecil’s reading probes for weeks 2 and 3 were administered at his house. All other probes were administered at Project Help. Cecil’s words correct per minute (WCPM) during baseline remained relatively stable (Figure 1). During Project Help, he demonstrated a gradual increase in WCPM as indicated by the positive slope of 0.6. The lowest data point throughout the study for Cecil occurred during week 10. All three of the reading students scored lower than usual during this week. Week 10 immediately
Figure 1. CBM Probes.
Each data point represents one reading or math CBM probe. Solid lines are lines of best fit. Linear regression equations represent digits or words correct.
followed spring break, so all three of the students had not received instruction at Project Help for one week, although they were still attending their respective elementary schools. Cecil had also missed the first two days of week 10 due to illness, and his CBM probe was administered the first day he returned.

Tony’s reading probe for week 4 was administered at his house. All other reading probes were administered at Project Help. Tony’s baseline WCPM were very stable (Figure 1). Inexplicably, his WCPM remained at the same level through baseline and much of the Project Help tutoring. Tony had a diagnosis of ADHD, and his tutor reported that it was difficult to keep him on task. He had difficulty sitting through entire lessons, so the tutor allowed him to sometimes stand during instruction. During week 12, he asked to stand during the CBM probe (he did not stand on the prior probes), and this was allowed. During the administration of the probe, it appeared that he had difficulty reading the passage because of the distance between his eyes and the page lying on the table. For the rest of the CBM probes he was not allowed to stand. All but 3 probes during Project Help tutoring are above Tony’s highest baseline point. Also, the last 3 weeks of Project Help show a clear increase in WCPM. The slopes during baseline and Project Help indicate a greater positive slope during Project Help than baseline.

All of Oliver’s reading probes were administered at Project Help. Oliver’s WCPM increased in level from baseline to Project Help (Figure 1). There was a rapid increase the first 3 weeks of Project Help which tended to level off the final weeks, but the trend line has a positive slope. There is no data point for the fourteenth week because Oliver was absent this entire week.

There is a correlation between the WCPM on the CBM probes and the number of lessons each student completed. Oliver had the highest WCPM throughout the study,
and he completed the greatest number of lessons. Cecil had the lowest WCPM (at least by the end of Project Help), and he completed the fewest lessons.

The number of WIPM remained relatively stable for all reading students. This is particularly noteworthy since during the latter part of Project Help students were reading many more words; so there were more opportunities to make errors.

CBA Results

Mary was administered CBA subtraction, multiplication, and division probes at the beginning and near the end of Project Help (Table 1). Mary made gains on all areas that were assessed. Gains were most significant on single digit multiplication and simple division. In these areas, both DCPM and DIPM improved.

<table>
<thead>
<tr>
<th>CBA Type</th>
<th>Beginning of Project Help</th>
<th>End of Project Help</th>
<th>Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single digit multiplication</td>
<td>28 DCPM, 12 DIPM</td>
<td>34 DCPM, 2 DIPM</td>
<td>+6 DCPM, -10 DIPM</td>
</tr>
<tr>
<td>Simple division</td>
<td>7 DCPM, 5 DIPM</td>
<td>19 DCPM, 1 DIPM</td>
<td>+12 DCPM, -4 DIPM</td>
</tr>
<tr>
<td>2 digit minus 1 digit subtraction</td>
<td>15 DCPM, 0 DIPM</td>
<td>18 DCPM, 0 DIPM</td>
<td>+3 DCPM, 0 DIPM</td>
</tr>
</tbody>
</table>

DCPM=Digits correct per minute
DIPM=Digits incorrect per minute

Table 2 shows the results of the CBA probe administrations for the reading
students. All of the CBA probes, like the CBM probes, were from the pre-primer level of the *Macmillan Connections Reading Program* (Arnold & Smith, 1987). The post-Project Help CBA score for Tony was similar to his CBM scores from the final weeks. The post-Project Help CBA score for Oliver was slightly lower than the CBM scores from the final weeks. Cecil’s post-Project Help CBA score was much higher than any of his CBM scores. Hence, the gains may overestimate Cecil’s ability and underestimate Oliver’s. While all of the probes indicate improvement, the results were quite variable. The results are not reconciled by considering the number of lessons each student completed. Oliver completed the greatest number of lessons (73) and made the lowest gains, while Cecil completed the fewest (58) lessons and made the highest gains.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Pre-Project Help CBA</th>
<th>Post-Project Help CBA</th>
<th>Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cecil</td>
<td>8</td>
<td>44</td>
<td>36</td>
</tr>
<tr>
<td>Tony</td>
<td>16</td>
<td>37</td>
<td>21</td>
</tr>
<tr>
<td>Oliver</td>
<td>27</td>
<td>36</td>
<td>9</td>
</tr>
</tbody>
</table>

Achievement Test Results

The achievement test results for Mary are presented in Table 3. She was
administered the math sections from the *Woodcock-Johnson Tests of Achievement* prior to and near the end of Project Help. While she gained half a grade on the Calculation and Quantitative Concepts subtests, her score declined on the Applied Problems subtest.

Table 3

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Pre-Project Help</th>
<th>Post-Project Help</th>
<th>Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculation</td>
<td>5.4</td>
<td>5.9</td>
<td>+0.5</td>
</tr>
<tr>
<td>Applied Problems</td>
<td>4.9</td>
<td>4.4</td>
<td>-0.5</td>
</tr>
<tr>
<td>Quantitative Concepts</td>
<td>5.5</td>
<td>6</td>
<td>+0.5</td>
</tr>
</tbody>
</table>

All of the reading students were administered the *Woodcock Reading Mastery Tests-Revised* prior to and near the end of Project Help. A different form of the test was administered on the second occasion. Table 4 contains the test results from Cecil. He made modest but fairly consistent gains across the subtests. The test results for Tony are displayed in Table 5. Again, gains were made on all subtests, especially on the Word Attack subtest.

The test results for Oliver are displayed in Table 6. Oliver had completed the greatest number of lessons and his WCPM on the CBM probes was the highest, but, unexpectedly, he showed essentially no gain on the Word Identification and Word Attack subtests. He did show some increase in comprehension, but the focus of
Corrective Reading: Decoding is on decoding rather than comprehension. Therefore, one would expect the greatest gains to be seen on the Word Identification and Word Attack subtests.

Social Validity

Student grade and progress notes provided some information on whether the skills acquired at Project Help generalized to the students' performances at school. According to Mary's report card from the Kalamazoo Public schools, she received a "D" in Math for the second marking period of the school year. The second marking period ended a few weeks prior to beginning Project Help. During the third marking period, Mary received a "B" in math. The third marking period ended approximately

Table 4
Grade Equivalents for Cecil From the Woodcock Reading Mastery Tests-Revised

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Pre-Project Help</th>
<th>Post-Project Help</th>
<th>Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Identification</td>
<td>1.4</td>
<td>1.7</td>
<td>+0.3</td>
</tr>
<tr>
<td>Word Attack</td>
<td>1.4</td>
<td>1.8</td>
<td>+0.4</td>
</tr>
<tr>
<td>Word Comprehension</td>
<td>1.3</td>
<td>1.6</td>
<td>+0.3</td>
</tr>
<tr>
<td>Passage Comprehension</td>
<td>1.4</td>
<td>1.5</td>
<td>+0.1</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
### Table 5
Grade Equivalents for Tony From the Woodcock Reading Mastery Tests-Revised

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Pre-Project Help</th>
<th>Post-Project Help</th>
<th>Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Identification</td>
<td>1.3</td>
<td>1.7</td>
<td>+0.4</td>
</tr>
<tr>
<td>Word Attack</td>
<td>1.9</td>
<td>1.9</td>
<td>+1</td>
</tr>
<tr>
<td>Word Comprehension</td>
<td>1.4</td>
<td>1.7</td>
<td>+0.3</td>
</tr>
<tr>
<td>Passage Comprehension</td>
<td>1.2</td>
<td>1.7</td>
<td>+0.5</td>
</tr>
</tbody>
</table>

### Table 6
Grade Equivalents for Oliver From the Woodcock Reading Mastery Tests-Revised

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Pre-Project Help</th>
<th>Post-Project Help</th>
<th>Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Identification</td>
<td>2.3</td>
<td>2.4</td>
<td>+0.1</td>
</tr>
<tr>
<td>Word Attack</td>
<td>2.0</td>
<td>1.8</td>
<td>-0.2</td>
</tr>
<tr>
<td>Word Comprehension</td>
<td>2.4</td>
<td>2.8</td>
<td>+0.4</td>
</tr>
<tr>
<td>Passage Comprehension</td>
<td>1.9</td>
<td>2.3</td>
<td>+0.4</td>
</tr>
</tbody>
</table>
two weeks prior to the end of Project Help. None of Mary’s grades in other academic areas made this much improvement, and her grade in reading declined. Cecil received a “2-” in reading for both the second and third marking periods. A “2” indicates satisfactory effort and “-” indicates a weakness. His report card did show a 30% improvement on a reading test from the second to the third marking periods. Also, the parents reported that the teachers told them that Cecil’s reading had improved since starting Project Help.

From the second to the third marking periods, Tony went up in 3 out of 10 categories in reading/writing and down in one. Tony’s report card indicated a significant decrease in math and most other academic/behavioral categories from the second to the third marking periods. His teachers reported significant behavior problems and off-task behaviors during the third marking period. The teachers hypothesized that Tony’s behavior problems contributed to his lack of success in the classroom.

Oliver’s parents made more information available on him than was available for the other students. Oliver improved in 4 out of 5 areas of reading from the second to the third marking periods. In all 4 cases, the improvements were of 1/2 grade. The second marking period grades indicated that he made satisfactory progress in only 2 of the 5 areas of reading. At the end of the third marking period, he made satisfactory progress in all 5 areas. Oliver also improved both of his spelling grades by 1/2 grade. One of the two spelling grades had been unsatisfactory for the second marking period, but both were satisfactory for the third marking period. Oliver’s report card indicates that he also made substantial improvements in math during the third marking period, but improvements in other academic areas were non-existent or minimal. At his school, Oliver was administered the Word Identification and Passage Comprehension subtests
of the Woodcock Reading Mastery Tests-Revised in late January and in April. In Word Identification, he obtained a 1.9 grade equivalent in January and a 2.2 grade equivalent in April. In Passage Comprehension, he obtained a 2.0 grade equivalent in January and a 2.4 grade equivalent in April. The 0.3 gain in Word Identification was greater than the gain assessed at Project Help, which was 0.1.
DISCUSSION

This study showed that Corrective Reading: Decoding and CMC can be successfully used within a remedial education program to improve academic performance. The success of the program is indicated by the increase in the CBM probes, the gains made in the CBA probes, the achievement test results, and the report cards and reports from parents. However, the results were variable across students and within students on the various dependent measures.

Mary made consistent gains throughout baseline and Project Help on the CBM math probes. The greatest indicator of improvement was the decrease in DIPM during Project Help. The CBM reading probes for the other students were relatively stable during baseline, and the scores of all three increased during Project Help. But there was a great deal of between-subject variability. Cecil’s scores increased gradually throughout Project Help, while Tony’s increased near the end of tutoring, and Oliver’s increased most significantly at the beginning of Project Help. Unlike the math CBM probes, there was also significant within-subject variability on the reading probes. Cecil and Tony had both their highest and lowest CBM scores during Project Help.

The variability in CBM reading scores could be due to a lack of overlap between the skills measured on the reading probes and the skills taught in the curriculum. In Tony’s case, little improvement is evident the first seven weeks of Project Help, although his performance during tutoring indicated that he was learning. The variability between and within subjects on the CBM reading probes suggests that they were not particularly sensitive to the effects of the intervention. Although, they may be sensitive enough under many conditions. The probes are also easy to make and quick to
administer, but the data suggest that they need to be administered over a number of weeks before any conclusions can be drawn. CBM reading probes from the Corrective Reading program may have been more sensitive, but there are no passages in Corrective Reading: Decoding A to make CBM probes from. There are passages in the rest of the Corrective Reading series.

The achievement test results are generally positive but variable across students. Mary made gains of 1/2 grade on the Calculation and Quantitative Concepts sections of the Woodcock-Johnson Tests of Achievement, but her score declined by 1/2 grade on the Applied Problems section. There is great variability in what is taught in different math curricula and in what is tested on different math achievement tests (Howell, et al., 1993). This has lead Howell et al. (1993) to conclude that "the grade level statements from all math tests are useless." (p. 324) The math test results should, therefore, be interpreted cautiously because the tests are likely to be fairly insensitive to student progress.

Cecil made modest gains across all sections of the Woodcock Reading Mastery Tests-Revised. Tony made higher gains, particularly on the Word Attack subtest where he gained an entire grade level. Oliver made gains on the comprehension subtests, but he made only a 0.1 gain in Word Identification and a -0.2 loss in Word Attack. This was unexpected because he had completed the greatest number of lessons. Most of his errors on the Word Attack subtest consisted of saying the long vowel sound rather than the short sound. For example, when presented with the nonsense word “mem,” he said /meem/ and when presented with “eb,” he said /eeb/. Tony, on the other hand, did not consistently make these types of errors. One possible explanation for Oliver’s performance on test day was that he had received only one day of instruction at Project Help in the last 13 days. His absences were due to vacation and illness. Also, Oliver’s
pre-test score on the WRMT-R was much higher than that of the other students, perhaps instruction on skills more advanced than those encountered in *Corrective Reading: Decoding A* is necessary to achieve a higher score.

There was also significant variability between the dependent variables. For example, Oliver completed the most lessons and made the greatest gains on the CBM probes, yet his gains were the smallest on the CBA probes and in the achievement test results. Tony made significant gains across all measures but showed little progress at school. There are other results that can be gleaned from the present study.

In the case of the three reading students, there was a correlation between WCPM on the initial CBA probes and the number of lessons each student completed. Higher reading fluency was associated with a greater number of lessons completed. Informal observations and tutor notes also indicated that Cecil and Tony exhibited behaviors that interfered with the progression of the lessons. Cecil refused to work on approximately four occasions. These refusals lasted over an hour on two occasions. Tony frequently tried to engage the tutor in a conversation about something other than the lesson. Tony also did not sit still, and he completed portions of lessons while standing. These off-task behaviors may be the main reason for the reduced number of lessons completed by Cecil and Tony.

Implications for Future Research and Weaknesses in the Present Study

Given the variability in results between students, the variability within the CBM reading probe scores within students, and the variability between the different measures within students, a more direct and sensitive measure of learning may be warranted. In the present study, the CBM probes were sensitive to the effects of the interventions, but for the purposes of intervention evaluation a more sensitive measure may be desirable.
If the student is reading passages, then the standard CBM procedures may be sensitive enough. CBM reading probes based on the student’s curriculum would probably be more sensitive than probes based on another curriculum. If the student is not yet reading passages, then one could create word/letter lists based on what was being taught in the curriculum. Repeatedly administering the lists could serve as a repeated measure. This would have been an interesting dependent variable in the present study.

In the case of math curricula, worksheets could be created that directly assessed skills other than the basic skills, such as problem-solving skills. These alternatives should be viewed as supplements to the CBM probes. Research has documented the reliability and validity of CBM methods, and CBM often does provide a more holistic measure of skill acquisition. For example, using word/sound lists as the repeated measure is not as closely tied to the terminal behavior of interest as is a CBM reading probe, which measures passage reading.

A common criticism of published achievement tests is their lack of content validity. But this criticism could also hold for CBM probes under some conditions. A reading lesson could teach letter-sound relationships that are not measured on a particular CBM reading probe. In the present study, observations of students during the lessons and performance on the DI mastery tests indicated that the students were learning new skills each lesson. But these changes in behavior were often not detected by the CBM probes. Also, because the focus of CBM is on the assessment of basic skills, it can underestimate the skills acquired in a particular program. In the present study, Mary learned a variety of problem solving skills in CMC, but these skills were not directly assessed by the math probes. Shinn and Bamonto (1998) argue that because basic skills are so important, the assessment of these skills is critical. This is a reasonable point, but it should be kept in mind that students can be learning or missing...
skills that are not directly assessed using the traditional CBM probes.

As previously mentioned, it would have been interesting to have a more sensitive measure of student learning that more directly assessed the skills taught in the lessons. Comparing this measure to the traditional CBM probes may have provided useful information as to the sensitivity of both measures. There are other methodological problems in the current study. The number of baseline data points was small—consisting of only 2 to 4 observations. Hence, the slopes computed using the baseline data points should be viewed with caution. Also, the multiple-baseline across subjects is very weak. At most, two observations separated subjects. This problem is difficult to overcome at Project Help because the tutoring schedule is governed by the University's semester schedule. The multiple-baseline in this study was only possible because two of the students started Project Help late. In spite of these difficulties, the data strongly suggest that the DI programs used at Project Help produce significant improvements in academic behavior, and that CBM methods can be successfully used to measure the effects of these programs.
Appendix A

Protocol Clearance From the Human Subjects
Institutional Review Board
Date: 23 November 1998

To: Jack Michael, Principal Investigator
   Michael Hixson, Student Investigator for dissertation

From: Sylvia Culp, Chair

Re: HSIRB Project Number 98-10-15

This letter will serve as confirmation that your research project entitled “The Use of CBM to Evaluate the Effects of a Remedial Reading Program and a Token Economy” has been approved under the full category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may only conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

Please remember to include the HSIRB project number on all future correspondence in order to speed processing.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: 23 November 1999
BIBLIOGRAPHY


