JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1994, 27, 101–113 NUMBER 1 (SPRING 1994)

THE UTILITY OF CURRICULUM-BASED MEASUREMENT FOR EVALUATING THE EFFECTS OF METHYLPHENIDATE ON ACADEMIC PERFORMANCE

GARY STONER, SEAN P. CAREY, MARTIN J. IKEDA, AND MARK R. SHINN

UNIVERSITY OF OREGON

Two case studies were conducted to investigate the utility of curriculum-based measurement of math and reading for evaluating the effects of methylphenidate on the academic performance of 2 students diagnosed with attention deficit hyperactivity disorder. Following baseline measurement, double-blind placebo-controlled procedures were employed to evaluate each student's response to three levels (5 mg, 10 mg, and 15 mg) of the medication. Results of the first study suggest that the curriculum-based measures were sensitive indicators of the student's response to medication. This finding was replicated in the second study. In the second study, when the student's follow-up dose of medication was based on trial-phase data, follow-up performance was improved compared to baseline performance. These case studies suggest that further research is warranted on the utility of curriculum-based measurements for monitoring and evaluating stimulant medication interventions with children with this disorder.

DESCRIPTORS: medication evaluation, treatment utility, academic behavior

It is estimated that 3% to 5% of children in the United States meet the current diagnostic criteria for attention deficit hyperactivity disorder (ADHD; American Psychiatric Association, 1987). Children with this diagnosis typically experience school-related difficulties in the areas of academic performance and achievement, including completing assignments, following teacher directions, and mastering basic literacy skills (August & Garfinkel, 1990; Frick & Lahey, 1991). As a group, their scores on standardized achievement tests are up to one standard deviation below their peers (Barkley, 1990); more than 40% of these students receive special education services in school under the categories of specific learning disabilities or behavior disorders (Weiss & Hechtman, 1986).

Stimulant medication, although controversial, is the most common treatment for children with ADHD who experience school-related problems. As many as 750,000 children, more than 2% of the school population, are prescribed these drugs annually, with methylphenidate (Ritalin®) being the most commonly used (DuPaul, Barkley, & McMurray, 1991; Safer & Krager, 1988). Although studies indicate that methylphenidate treatment has a positive impact on academic productivity in 70% to 80% of cases (Rapport, 1987), the drug's effects are idiosyncratic and task specific (Rapport et al., 1987). Unfortunately, methylphenidate's effects cannot be predicted reliably from a child's size, weight, or age (Rapport, DuPaul, & Kelly, 1989). Difficulty in predicting the effects of methylphenidate is further complicated by the issue of determining an effective dose. For any child, the same dose of methylphenidate may produce positive, negative, or no changes in performance, depending on the evaluation task (Rapport, 1987). Therefore, one of the critical issues inherent in the use of stimulant medication for a given child is determining what dose, if any, is appropriate.

Most commonly, stimulant medication doses are determined by subjective parent report (Gadow, 1981). For example, a relatively small dose typically is prescribed initially, and 1 to 2 weeks later the parent's report of medication effects is used to determine whether adjustments in dose are war-

The first two authors contributed equally to the development and completion of this project. This research was supported in part by a grant from the National Association of School Psychologists to Gary Stoner. We gratefully acknowledge Brian K. Martens and the anonymous reviewers of an earlier version of this manuscript for their helpful comments and suggestions.

Reprints can be obtained from Gary Stoner, School Psychology Program, Division of Special Education and Rehabilitation, College of Education, University of Oregon, Eugene, Oregon 97403.

ranted. When prescribing physicians do gather information about dosage effects in a more objective manner, they typically use behavior rating scales that are completed by parents and/or teachers. It is estimated that rating scales are used for this purpose by only slightly more than half of prescribing physicians (Copeland, Wolraich, Lindgren, Milich, & Woolson, 1987). Behavior rating scales may be problematic for purposes of evaluating stimulant medication effects, especially when used in the absence of other outcome measures. For example, many of these instruments are technically inadequate, and in general they are subject to informant bias (Shapiro & Kratochwill, 1988). The most important problem in using behavior rating scales exclusively to evaluate stimulant medication effects is that most scales are capable of indicating behavioral improvement solely in terms of reductions in ratings of problem behaviors. Thus, a dose that results in a child being rated as less noncompliant and disruptive compared to baseline may be considered effective and beneficial, although that dose impairs the child's ability to complete academic tasks accurately.

An alternative to behavior rating scales for determining an effective dose of stimulant medication that appears primarily in the clinical research literature is the use of various performance tasks administered in clinic or office settings. Examples of these tasks include the Continuous Performance Test (Douglas, 1984; Rapport, DuPaul, Stoner, & Jones, 1986), the Matching Familiar Figures Test (Brown & Sleator, 1979; Rapport et al., 1988), the Paired Associates Learning Test (Swanson & Kinsbourne, 1975; Rapport, Stoner, DuPaul, Birmingham, & Tucker, 1985), and other tasks involving short-term memory (Sprague & Sleator, 1977). The primary problem with these methods of determining dose is that they have been shown to lack predictive validity with respect to a child's behavior in the classroom setting (see Rapport & Kelly, 1991, for a review).

Another alternative, also appearing primarily in clinical research, is the use of daily measures of a child's classroom performance. For example, percentage of language arts work completed and the percentage of this work completed accurately have been used as determinants of effective dose (Rapport et al., 1988). Although this approach is the most naturalistic of the methods used to date, it is problematic because of variability in task difficulty from day to day and the lack of standardization with respect to the conditions under which the student's work is produced. Thus, the validity of making decisions about dosage using comparisons of work completion and accuracy rates across days and conditions must be questioned.

A critical decision in evaluating the effects of methylphenidate on the academic performance and behavior of a student with ADHD is the selection of appropriate outcome measures. Further, there is a critical need in practice for readily usable, reliable, and valid measures of important educational outcomes that might be influenced by methylphenidate.

Basic principles of behavioral assessment, as outlined in a variety of sources (e.g., Barlow, Hayes, & Nelson, 1984; DuPaul & Barkley, 1993; Shapiro & Kratochwill, 1988), provide clear guidelines for choosing or developing outcome measures. Briefly, these principles suggest that to assess changes in behavior, measures should (a) use rate of responding as a primary dependent measure; (b) be socially valid; (c) be accurate (i.e., demonstrate sound psychometric properties); (d) be capable of repeated administration prior to, during, and following treatment; (e) provide standardized counts of behavior; and (f) demonstrate sensitivity with respect to precise measurement of critical effects.

One systematic approach to measuring educational behavior and outcomes that meets these criteria is curriculum-based measurement (CBM; Goodwin & Shinn, 1990; Shinn, 1989). CBM was developed from a behavioral-assessment perspective to provide a technology for systematic, formative evaluation of student academic outcomes in the basic skill areas of reading, spelling, writing, and math, and for evaluating intervention effectiveness using single-case study designs (Deno, Mirkin, & Chiang, 1982). Primary CBM data are derived from brief (1- to 3-min) fluency measures of student performance in reading, math, spelling, and written expression. CBM is content valid; the materials used to evaluate outcomes are sampled di-

Table 1
Standardized Tasks, Scoring Procedures, and Technical Adequacy Evidence for CBM Measures of Reading, Math, Spelling, and Written Expression

Academic area	Task	Scoring	Technical adequacy evidence
Reading	Students read passages aloud for 1 min.	Number of words read correctly; number of errors.	Deno, Mirkin, and Chiang (1982); Fuchs, Fuchs, and Maxwell (1988)
Math	Students write answers to written computation problems for 2 to 5 min.	Number of correct digits.	Fuchs and Fuchs (1987); Marston, Fuchs, and Deno (1986)
Spelling	Students write words dictated orally for 2 min.	Number of correct letter sequences; number of words spelled correctly.	Deno, Marston, Mirkin et al. (1982); Marston, Lowry, Deno, and Mirkin (1981)
Written expression	After being given a story starter or topic sentence, students write a story for 3 min.	Number of words written; number of words spelled correctly; number of correct word sequences.	Deno, Marston, and Mirkin (1982)

rectly from the student's curriculum. Also, the measures assess important and socially valid terminal behaviors (i.e., number of words read correctly, number of correct letter sequences written, number of correct math problems, or correct digits written).

CBM is distinct from other behaviorally oriented approaches to academic assessment, such as subskill mastery measurement or precision teaching, in that it is an example of a general outcome measurement approach to instructional decision making (see Fuchs & Deno, 1991, for an extended discussion of this distinction). Using CBM, student performance is typically assessed with respect to broad cumulative behaviors (such as reading unfamiliar text aloud rather than reading practiced material or word lists), with an emphasis on prescriptive assessment. A goal of CBM is to provide teachers with assessment information that can be used to plan instructional programs and to evaluate overall student growth. Another distinctive feature of CBM is that an extensive body of research has accumulated to support the technical adequacy of the principal measures from both behavioral and traditional psychometric perspectives (Fuchs & Deno, 1991). Further, the standardized procedures for conducting CBM probes are designed for simple, low-cost, repeated administration (Knutson & Shinn, 1991).

A systematic program of research was initiated in the late 1970s to investigate the technical adequacy of the standardized academic tasks used in CBM (Marston, 1989). Since then, an extensive body of empirical evidence supports the reliability and validity of CBM for educational decision making. For example, in the CBM reading task, the number of words read correctly has been validated as a reliable and accurate measure of students' general reading skills, including reading comprehension (Shinn, Good, Knutson, Tilly, & Collins, 1992). A summary of the standardized tasks and scoring procedures, along with references to several studies documenting the technical adequacy of CBM reading, math, spelling and written expression measures, is presented in Table 1.

A number of researchers have demonstrated that CBM data are sensitive to changes in student performance as a result of various instructional interventions (Deno et al., 1982; Marston, Fuchs, & Deno, 1986). For example, CBM has been used to evaluate the effects of computer-assisted instruction (Fuchs, 1988), classwide peer tutoring (DuPaul & Henningson, 1993), and goal-setting strategies (Fuchs, Fuchs, & Hamlett, 1989) on students' academic achievement. Given these features, CBM is a potentially useful technology for assessing the effects of medication on the academic performance of children with ADHD. To test this possibility, we conducted two case studies using CBM reading and math measures to evaluate the influence of different doses of methylphenidate on the academic performance of 2 students with ADHD. The first study focused on the issue of the sensitivity of CBM to medication effects. If methylphenidate influences current academic performance in a positive or negative manner, then that influence should be reflected by changes in level of the CBM data. Therefore, an increased (or decreased) level of reading or math performance relative to baseline or placebo would indicate a beneficial (or deleterious) dose of methylphenidate. Such short-term shifts in level of performance would be indicative of methylphenidate effects on a student's ability to demonstrate previously acquired skills. Thus, the research question being addressed was whether the student's CBM reading and math scores would covary with changes in the dose of methylphenidate.

Sensitivity to treatment effects alone is not sufficient to demonstrate the adequacy of a behavioral outcome measure. As Power and Franks (1988) stated, behavioral assessment has a "two-fold purpose: (a) to provide predictive information with respect to the potential efficacy of one intervention over another, and (b) to monitor and evaluate the effects of the intervention once it is implemented" (p. 17). Therefore, in the second case study, we focused not only on sensitivity but also on the issues of prediction, monitoring, and evaluation. Following a baseline phase and after the student was given several trials at different doses of methylphenidate, his CBM scores were used to select the dose that was likely to have optimal effects on his academic performance. The student was monitored on this follow-up dose for 2 weeks. At the end of this period, his performance during the follow-up phase was compared to baseline to evaluate the effects of the methylphenidate dose. This comparison was intended to investigate the utility of CBM for predicting the student's response to methylphenidate. If the student's reading and math performances were improved relative to baseline, higher levels of performance and/or higher rates of growth would provide evidence to support the utility of CBM for predicting methylphenidate effects.

In both studies, assessment instruments that have been developed specifically for evaluating medication effects on children with ADHD were used in conjunction with CBM. The data from these additional measures were compared with the CBM data to investigate their degree of congruence with respect to changes due to medication dose.

METHOD

Subjects

Dan was 9 years old and in third grade at a rural elementary school. Teachers reported that he had problems in his social interactions with peers and had difficulty remaining on-task and paying attention during academic instruction. Bill was 13 years old and in the eighth grade at a rural junior high school. His teachers reported that he had difficulties in the areas of work completion, remaining on-task, and following directions.

Both students were referred by their family physicians to a university clinic, where they were evaluated by a licensed psychologist and pediatrician as meeting the DSM-III-R (APA, 1987) criteria for diagnosis of ADHD. Based on the diagnostic information, both students' parents and physicians decided that trials of stimulant medication would be appropriate. After making this decision, the students and their parents were invited to participate in a controlled evaluation of the effects of different levels of stimulant medication on the students' behavior. In each case, the families agreed, and written informed consent to participate was obtained from the students and their parents.

Measures

Changes in student behavior in response to the stimulant medication were monitored using CBM math and reading probes. In addition, two standardized behavior rating scales were used—one that provides information about a child's classroom academic performance and behavior, and one that provides for a brief evaluation of child inattention and overactivity in the classroom. Finally, a rating scale was used to evaluate potential medication side effects.

CBM probes. Standardized procedures for preparing, administering, and scoring the CBM reading and math probes were followed. Reading passages and math computation problems were sampled from the textbooks in which the students were being instructed (Shinn, 1989). Bill's reading passages were chosen from *Arrangement in Literature* (Scott, Foresman, and Company), and Dan's were selected from *Each New Day, Book 3-1* (Scribner). Following standardized CBM procedures, passages of at least 250 words were selected randomly, using a table of random numbers to choose page numbers. If the text on a selected page was written as a poem or play, included many unusual proper names, or had extensive dialogue, it was not included. Each passage was retyped in a font of approximately the same size and type as the original text to maintain a standard presentation format. These selection and presentation procedures are intended to control for relative difficulty of reading passages (Shinn, 1993b). Students read aloud for 1 min from a randomly selected passage. Data collectors recorded the number of words read correctly and the number of errors made. Each reading was tape recorded so that reliability observations could be conducted.

Standardized CBM procedures were also used to prepare the worksheets that served as math probes. First, the scope and sequence charts of the students' math curriculum were examined (a) to determine the range of math computation skills covered by the curriculum (e.g., single- or double-digit addition and subtraction, etc.) and (b) to estimate the proportion of each problem type included in the curriculum (e.g., 10% single-digit addition, 15% double-digit addition, etc.). Second, a pool of computation problems that covered the range of math skills in each curriculum was created. Finally, a series of math probes was constructed in which each problem type was represented in proportion to its occurrence in the curriculum, and in which the same sequence of problem types was followed on each probe. Thus, each probe was a worksheet containing 10 rows and 10 columns of math computation problems. These procedures were followed to ensure that probes were of comparable difficulty.

During math probes, the students were given 2 min to answer as many problems as possible on a randomly selected worksheet. Using scoring templates, data collectors scored the number of correct digits written by each student in his answer to each problem on a probe (e.g., $18 + 9 = 27$ is scored as 2 correct digits; if the student wrote 26 in answer to the problem $18 + 9 =$ __, this would be scored as 1 digit correct). After scoring each problem attempted by the student, the data collectors added the number of correct digits written on each probe and recorded the total on a summary data sheet (Shinn, 1993a).

Teacher ratings. At the end of each medication trial, students' primary classroom teachers were asked to complete ratings of the students' in-class behavior during the trial period, using the Academic Performance Rating Scale (APRS; DuPaul, Rapport, & Perriello, 1991) and the Child Attention Problems scale (CAP; Barkley, 1990). The APRS is a 19-item rating scale that was designed to assess teachers' perceptions of children's academic skills and performance (e.g., amount and accuracy of work completed), as well as behavior relevant to classroom learning and performance. The first four items pertain to a student's work completion and accuracy in language arts and mathematics, with ratings corresponding to estimates of amount of work completed or accuracy during the rated time frame. For work completion, the ratings and their corresponding estimates were: 1 = 0% to 49%, 2 = 50% to 69%, 3 = 70% to 79%, 4 = 80% to 89%, and 5 = 90% to 100%. In the areas of work accuracy, the ratings and their corresponding estimates were: 1 = 0% to 64%, 2 = 65% to 69%, 3 = 70% to 79%, 4 = 80% to 89%, and 5 = 90% to 100%.

The remaining items on the APRS focus on educationally relevant behaviors (e.g., how frequently does this child require assistance to accurately complete academic work) using 5-point Likert-type scales to indicate quality of work or degree of behavior observed. The APRS yields a total score, as well as scores entitled Academic Success, Impulse Control, and Academic Productivity. Available data suggest that the APRS total score and factor scores possess adequate internal consistency and test–retest reliability characteristics and yield valid information regarding the classroom performance and behavior of children in Grades 1 through 6 (DuPaul et al., 1991). Dan's primary teacher completed the entire APRS once per phase, and Bill's homeroom and language arts teacher completed only the two items from the APRS specific to language arts work completion and accuracy for each phase.

The CAP is a 12-item rating scale that was designed primarily to assess stimulant drug effects on the classroom performance of children diagnosed with ADHD, and it has been shown to be sensitive for this purpose (Barkley, McMurray, Edelbrock, & Robbins, 1989). The CAP was derived from the 126-item Child Behavior Checklist—Teacher Report Form (Edelbrock & Achenbach, 1984) and consists of items that purportedly yield information regarding a child's inattentive or overactive behavior (e.g., fidgeting, daydreaming, failing to carry out assigned tasks, talking out of turn). Respondents check off each item as it describes the student's behavior during the preceding week as not true (scored as 0), somewhat or sometimes true (scored as 1), or very or very often true (scored as 2). These ratings are summed to give a total score (range, 0 to 24). Normative data are available for boys and girls aged 6 to 16, as is evidence for the internal consistency of the instrument and its validity (reported in Barkley, 1990). The teachers completed the entire CAP scale once per phase for their respective students.

Side effects scale. Both students, their parents, and Bill's teacher completed the Stimulant Drug Side Effects Rating Scale (SDERS; Barkley, 1981) at the end of each medication trial. On the SDERS, respondents rated whether the students had displayed any of the side effects (e.g., stomachaches, insomnia) sometimes associated with stimulant medications on a scale from 0 (= absent) to 9 (= serious).

Interobserver Agreement

Interobserver agreement was assessed throughout the evaluations for both students on CBM. Agreement on the reading scores was assessed by having a second observer listen to and score a sample of each student's recorded reading probes. Agreement was assessed for 30% of probes for Dan and 25% of probes for Bill. Agreement on the math scores was assessed by having a second observer score a sample of the completed math worksheets. Interscorer agreement was assessed for 30% of math worksheets completed by Dan and 25% of the worksheets completed by Bill. Percentage agreement for reading and math was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100%. Interobserver agreement was 100% for each measure.

Procedures

After a baseline period of 1 week for Dan and 2 weeks for Bill, the medication trials were begun. Each student received four levels of medication: 5 mg, 10 mg, and 15 mg of methylphenidate, and a placebo. Bill's medication trials consisted of three consecutive daily doses at each level. Dan received three consecutive daily doses of placebo and five consecutive daily doses at each level of methylphenidate. The order of medication level was determined randomly. The number of days at each dose was influenced by the combination of parents' concerns regarding the overall length of the trial, school schedules, student attendance, and research design factors. All medications were prepared by a licensed pharmacist. Each dose of methylphenidate and placebo was ground into a powder, mixed with an inert compound, and sealed in a small colored drug capsule so that doses were identical in appearance and taste. The pharmacist packaged the doses for each level of medication in separate envelopes that were coded for later identification. One of the authors then arranged the coded envelopes according to the randomly determined order of trials. Thus, neither the pharmacist, the data collectors, nor the subjects and their families knew the order of drug administration.

Parents were provided with their child's set of envelopes along with the sequence in which they were to be opened. To assess the integrity of drug administration, parents were instructed to initial a monitoring form each time they gave a dose to their child immediately after breakfast. CBM probes were administered to the students at school approximately 1 to 2 hr later, when the behavioral effects of methylphenidate are expected to peak (Donnelly & Rapoport, 1985). Three observers who had extensive training in CBM shared responsibilities for data collection and reliability observations. After all of the medication trials for a student were completed, the code for each dose was revealed. The CBM data were then graphed and labeled accordingly. A report that summarized the results of the medication evaluation, including the graphed CBM data, teacher ratings, and side effects scales, was prepared for each student. Copies of the report were given to the students' families and physicians.

Design

In keeping with the recommendations of Sprague and Werry (1971) for conducting drug evaluations, a double-blind, placebo-controlled, crossover experimental design was used. Fixed doses (5 mg, 10 mg, 15 mg) were given to reflect typical pediatric prescription practices and because a clear relationship between methylphenidate effects and blood levels has not been established (DuPaul et al., 1991).

RESULTS

Dan's performance data on the CBM probes across phases are displayed in Figure 1. The order of drug trials and Dan's mean level of reading performance, in words read correctly (WRC) per minute, were baseline, 29; placebo, 41; 10 mg, 38; 15 mg, 52; and 5 mg, 45. Dan's best average level of reading performance occurred during the 15-mg phase. Overall, there was considerable variability in his reading performance within each phase as well as across all conditions (range, 13 to 69 WRC). Dan's mean math performance scores, expressed in digits correct (DC), were baseline, 23; placebo, 31; 10 mg, 29; 15 mg, 36; and 5 mg, 33. As with his CBM reading performance, Dan's best average level of math performance occurred during the 15-mg phase. Again, there was a large degree of variability in his math performance across conditions (range, 9 to 45 DC).

Dan's academic performance in the classroom, as reflected in the APRS rating, was rated best for the 15-mg phase of the trial. This APRS rating was the only one obtained that was within the average range for third-grade boys. His attention-related behavior in the classroom, as indicated by the CAP scores, was rated best for the 10-mg phase. CAP ratings also indicated improvement in the area of attention at the 15-mg and placebo phases, compared to baseline. Thus, for Dan there was congruence between the CBM reading and math measures and the teacher ratings on the APRS, indicating that his best performance occurred during the 15-mg trial phase. In addition, Dan and his mother reported potential side effects of the medication (dizziness, irritability, stomachaches, and difficulty sleeping) during the 10-mg and 15-mg trials. These problems reportedly diminished as the trials proceeded.

These results were summarized and reported to Dan's parents and physician. They decided to continue with the 15-mg dose. Anecdotal reports from Dan's mother suggested that with the 15-mg dose of medication at follow-up, he was "better behaved and more focused" in school and in community activities (e.g., karate class), and that complaints regarding his academic and social behavior in school diminished considerably.

Bill's performance data on the CBM probes across phases are displayed in Figure 2. For Bill, the order of trial phases and mean level of reading performance, in WRC per minute, were baseline, 149; 10 mg, 146; 15 mg, 165; no-medication probe, 120; placebo, 154; no-medication probe, 125; and 5 mg, 181. Thus, Bill reached his highest reading level during the 5-mg trial. Bill's performance on the CBM math probes across phases ranged from 35 to 107 digits correct. His mean levels of CBM math performance were baseline, 64; 10 mg, 59; 15 mg, 59; no-medication probe, 86; placebo, 88; no-medication probe, 100; and 5 mg, 96. As with his CBM reading performance, his highest mean math performance score was recorded for the 5-mg phase, with his next best performance occurring

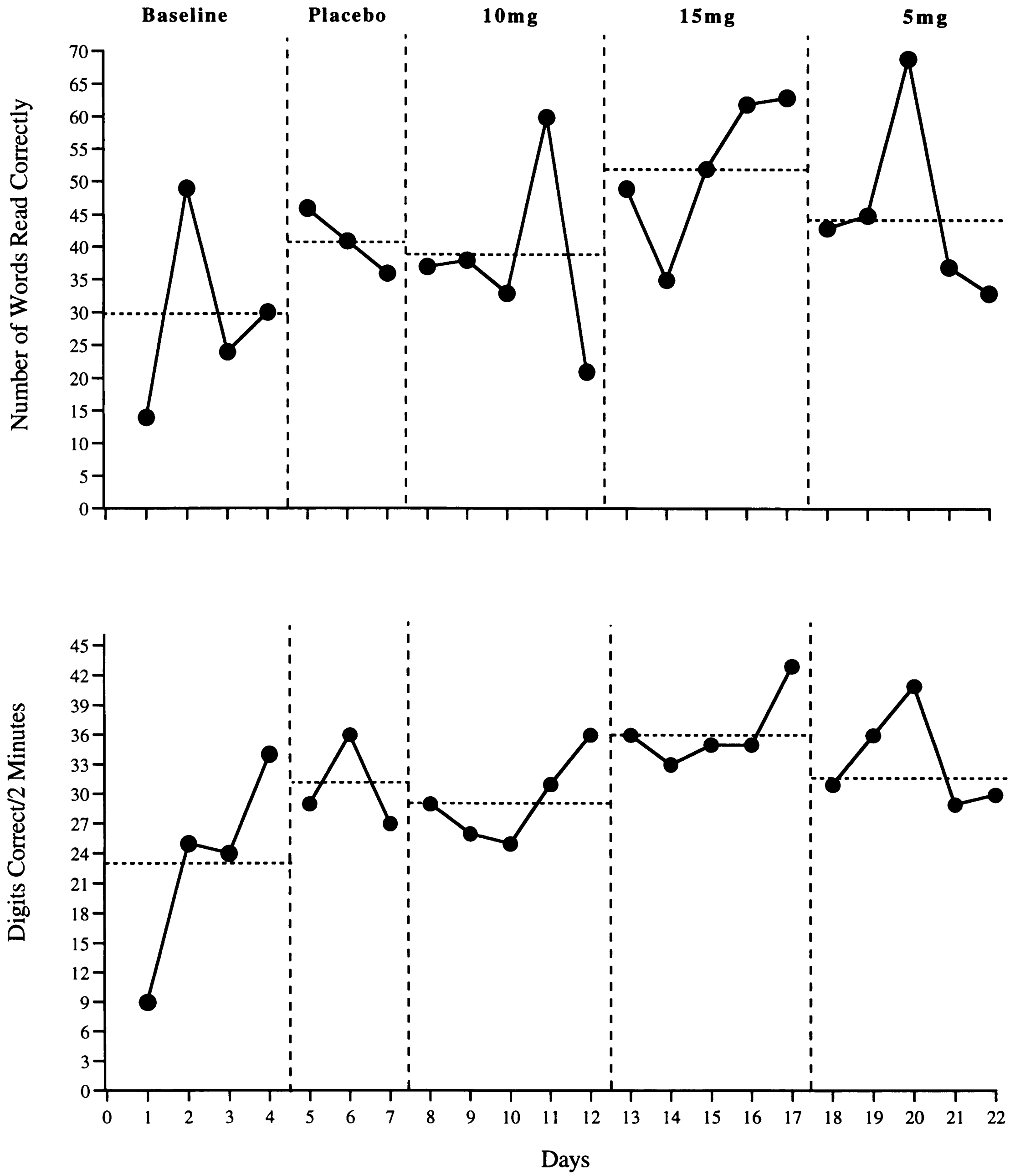

Figure 1. Number of words read correctly per minute and number of math digits correct per 2 minutes by Dan across conditions. Horizontal dashed lines denote mean level of performance during each condition.

during the placebo phase. On the 2 nondrug days, Bill's scores remained relatively high (86 and 100) compared with the baseline level.

For the trial phases, Bill's academic performance in the classroom, as reflected by teacher-completed items from the APRS, was rated as slightly and similarly improved, compared to baseline, across the 15-mg, placebo, and the 5-mg phases. The best teacher rating for Bill on the CAP occurred during the placebo phase. Thus, during Bill's trial phases, congruence with respect to optimized outcomes was obtained at the 5-mg phase across the

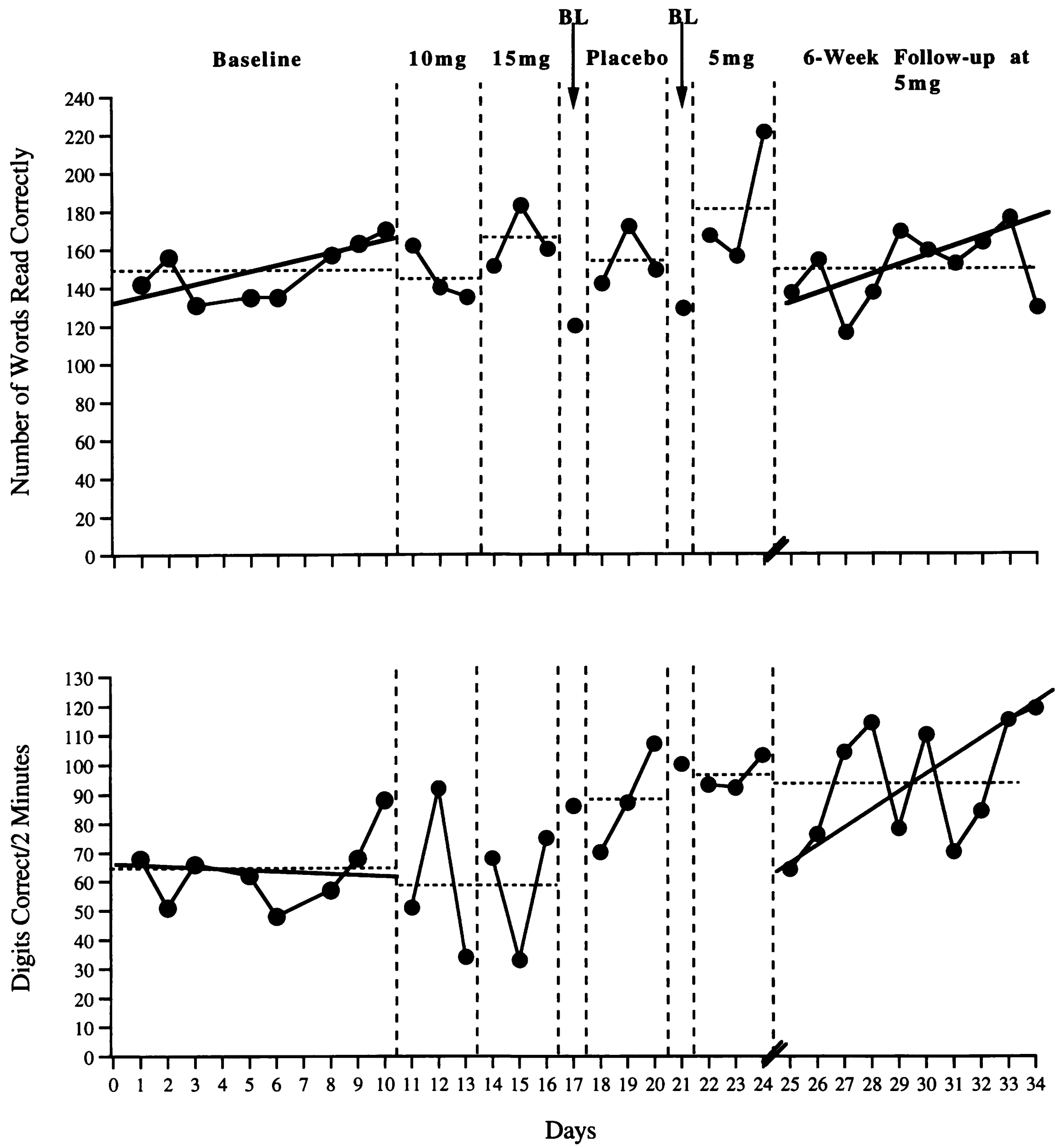

Figure 2. Number of words read correctly per minute and number of math digits correct per 2 minutes by Bill across conditions. Horizontal dashed lines denote mean level of performance during each condition. Solid lines represent slope of change over time, determined using a split-middle technique.

CBM reading and math outcomes. The teacher rating on the APRS at 5 mg was the same as each other nonbaseline phase. In addition to academic and behavioral outcomes, Bill's self-report indicated several potential side effects (difficulty sleeping, nightmares, dizziness) during the 10-mg phase only. These results were reported to Bill's parents and physician, who decided to continue with the 5-mg dose.

For Bill, a follow-up assessment phase took place 6 weeks later, when he was receiving 5 mg methylphenidate twice daily. In the follow-up phase, Bill's mean reading rate was 149 WRC, the same as his baseline level. His mean math score was 93

DC, which was considerably higher than the baseline level. To permit baseline versus follow-up comparisons of Bill's rates of performance in reading and math, a split-middle technique (Barlow & Hersen, 1984; White & Haring, 1981) was used to fit a line representing rate of change within each of these two phases. The plot values determined from this technique were also used to generate slopes for each phase, using the formula $(Y_2 - Y_1)/(X_2 - X_1)$. The slopes for Bill's reading were 3.67 words correct improvement per week for the baseline phase and 4.4 words correct improvement per week for the follow-up phase. In math, the slopes were −0.83 digits correct decline per week for baseline and 6.4 digits correct improvement per week for the follow-up phase. The rates of improvement at follow-up are consistent with improvement rates of typical students in general education classrooms (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993). Finally, teacher ratings on the APRS and CAP indicated that Bill's classroom performance and behavior were improved, compared to baseline conditions. Thus, during the follow-up phase, four of the five outcome measures were congruent (CBM reading being the only exception) in indicating optimized behavior and academic performance, and slopes of both CBM reading and math indicated improved rates of change at follow-up.

DISCUSSION

The case studies presented here represent an initial attempt to develop an innovative method for determining an optimal dose of methylphenidate for students being prescribed this medication. The results suggest that CBM for reading and math holds promise as outcome measures for use in stimulant medication trials, especially when the student's academic performance is a concern.

The sensitivity of CBM for reading and math to the effects of methylphenidate was the principal issue addressed in the study with Dan. If the CBM data were sensitive indicators of the medication's influence on Dan's academic performance, it was expected that his performance in reading and math would covary with changes in the dose of methylphenidate that he received. His results provided evidence to support this hypothesis. With each change in dose, there were concomitant changes in his reading and math scores, and his highest levels of both reading and math performance were observed during the same 15-mg medication trial. Further, the highest ratings by Dan's teacher of his general academic performance on the APRS were also obtained during the 15-mg phase. Thus, the covariation of CBM reading and math scores with medication doses, together with the congruence between these scores and the teacher's ratings in identifying the dose at which Dan's academic performance was optimized, suggest that CBM for reading and math may indeed be sensitive to methylphenidate effects.

In addition to the question of sensitivity, the study with Bill addressed the issue of the utility of CBM for selecting an optimal dose of methylphenidate for ongoing treatment. Following a 2-week baseline in which both the level and slope of his reading and math performance were monitored, Bill was given several brief trials at different levels of methylphenidate. Again, there was congruence between the reading and math data, indicating an optimal 5-mg dose of methylphenidate that was then prescribed for long-term treatment. After 6 weeks of daily methylphenidate doses at this level, his reading and math performance was monitored again for a 2-week period. In math, both level and slope were improved relative to baseline, whereas in reading his level was unchanged but the slope had improved. These results suggest that CBM data collected during short medication trials can be used to select a dose of methylphenidate that is likely to be beneficial to a student's ongoing academic growth.

Although the results from both case studies are promising, they need to be interpreted with great caution, because various methodological and design features introduced a number of threats to internal validity. For example, it would be more in keeping with standard practice of clinical research trials of medication to include a no-medication day between trial phases. Even though methylphenidate has a relatively brief half-life (4 to 6 hr) and is thought to be completely eliminated from the body within 24 hr of ingestion, it is possible that there was a carryover of effects from one trial to another. Further, the fact that the students experienced only one phase at each dose raises the potential for order and history effects (Cook & Campbell, 1979) as alternative explanations for changes in the students' performance from trial to trial.

Additional doubts about attributing changes in the CBM data to variations in methylphenidate dose are raised by the students' elevated performance relative to baseline during the placebo phase on some occasions. However, this apparent placebo effect may be an artifact of the relatively short medication trials. It is questionable whether the improvements in performance associated with taking the placebo would have been maintained over time. For both students, optimal levels of performance occurred when they actually were receiving a dose of methylphenidate, and Bill's follow-up data strongly suggest that the medication produced a sustained improvement in his reading and math achievement. Thus, although the potential influence of a placebo effect should not be minimized, neither should the observed positive effects of the drug.

Another threat to the internal validity of the results is the high degree of variability in performance both within and between phases. At least in part, this variability may be attributed to within-subject factors, because it has been suggested that exceptional behavioral variability is a distinguishing feature of children with ADHD (Barkley, 1990). Nevertheless, from a behavior-analytic perspective, it would be more appropriate to search for and control the sources of such variability in current or recent stimulus conditions.

For example, further attention to variability in measurement materials may be warranted. The method used to select reading passages may have influenced the variability in reading performance. Because passages were selected randomly from the current reading curriculum, it was possible that the students were asked to read, on some occasions, passages that they had been exposed to previously in class and, on other occasions, passages that they would not cover until later in the academic year. Further, it is possible that text occurring later in a reading curriculum may be more complex than text that occurs earlier. Thus, variability in both the familiarity and difficulty of reading passages may have contributed to the variability in students' reading data. The method used to select these reading passages is in keeping with standard CBM practice that focuses on measurement of general outcomes (Fuchs & Deno, 1991). However, for relatively short medication trials, it may be important to restrict the range of a reading curriculum from which passages are chosen. Limiting the range of materials may reduce the influences of familiarity and variable difficulty on reading performance. Alternatively, a long-term CBM approach may be appropriate for evaluating the effects of stimulant medication, particularly over extended periods of time. Long-term CBM is characterized by selection of measurement materials within which the student will be expected to perform in 1 year's time.

The most important recommendation for future research is for systematic replications of the procedures described here with a wider range of children and with extended follow-up phases. Research of this type is necessary to examine thoroughly (a) the sensitivity of CBM measures to medication effects and (b) the utility of these measures for making decisions regarding stimulant medication for children with ADHD. Such future work holds promise for contributing to improved outcomes for the hundreds of thousands of children who are prescribed stimulant medication annually.

REFERENCES

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

August, G. J., & Garfinkel, B. D. (1990). Behavioral and cognitive subtypes of ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry, 28,* 739–748.

Barkley, R. A. (1981). *Hyperactive children: A handbook for diagnosis and treatment.* New York: Guilford.

Barkley, R. A. (1990). *Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment.* New York: Guilford.

Barkley, R. A., McMurray, M. B., Edelbrock, C. S., & Robbins, K. (1989). The response of aggressive and nonaggressive ADHD children to two doses of methylphenidate. *Journal of the American Academy of Child and Adolescent Psychiatry, 28,* 873–881.

Barlow, D. H., Hayes, S. C., & Nelson, R. O. (1984). *The scientist-practitioner: Research and accountability in clinical and educational settings.* New York: Pergamon.

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change.* New York: Pergamon.

Brown, R. T., & Sleator, E. K. (1979). Methylphenidate in hyperkinetic children: Differences in dose effects on impulsive behavior. *Pediatrics, 64,* 408–411.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin.

Copeland, L., Wolraich, M., Lindgren, S., Milich, R., & Woolson, R. (1987). Pediatricians' reported practices in the assessment and treatment of attention deficit disorders. *Developmental and Behavioral Pediatrics, 8,* 191–197.

Deno, S. L., Marston, D., & Mirkin, P. K. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children, 48,* 368–371.

Deno, S. L., Marston, D., Mirkin, P. K., Lowry, L., Sindelar, P., & Jenkins, J. (1982). *The use of standardized tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study.* Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49,* 36–45.

Donnelly, M., & Rapoport, J. L. (1985). Attention deficit disorders. In J. M. Wiener (Ed.), *Diagnosis and psychopharmacology of childhood and adolescent disorders* (pp. 179–197). New York: Wiley.

Douglas, V. A. (1984). The psychological processes implicated in ADD. In L. M. Bloomingdale (Ed.), *Attention deficit disorder: Diagnostic, cognitive, and therapeutic understanding* (pp. 147–161). New York: Spectrum.

DuPaul, G. J., & Barkley, R. A. (1993). Behavioral contributions to pharmaco-therapy: The utility of behavioral methodology in medication treatment of children with attention deficit hyperactivity disorder. *Behavior Therapy, 24,* 47–65.

DuPaul, G. J., Barkley, R. A., & McMurray, M. B. (1991). Therapeutic effects of medication on ADHD: Implications for school psychologists. *School Psychology Review, 20,* 203–221.

DuPaul, G. J., & Henningson, P. N. (1993). Peer tutoring effects on the classroom performance of children with attention deficit hyperactivity disorder. *School Psychology Review, 22,* 134–143.

DuPaul, G. J., Rapport, M. D., & Perriello, L. M. (1991). Teacher ratings of academic skills: The development of the academic performance rating scale. *School Psychology Review, 20,* 284–300.

Edelbrock, C. S., & Achenbach, T. A. (1984). The teacher version of the Child Behavior Profile: I. Boys aged 6–11. *Journal of Consulting and Clinical Psychology, 52,* 207–217.

Frick, P. J., & Lahey, B. B. (1991). The nature and characteristics of attention-deficit hyperactivity disorder. *School Psychology Review, 20,* 163–173.

Fuchs, L. S. (1988). Effects of computer-managed instruction on teachers' implementation of systematic monitoring programs and student achievement. *Journal of Educational Research, 81,* 294–304.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57,* 488–500.

Fuchs, L. S., & Fuchs, D. (1987). *Effects of curriculum-based measurement procedures in spelling and math.* (Unpublished manuscript available from L. S. Fuchs, Box 328, Peabody College, Vanderbilt University, Nashville, TN 37203)

Fuchs, L. S., Fuchs, D., & Hamlett, C. (1989). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children, 55,* 429–438.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much can we expect? *School Psychology Review, 22,* 27–48.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial & Special Education, 9,* 20–28.

Gadow, K. D. (1981). Drug therapy for hyperactivity: Treatment procedures in natural settings. In K. D. Gadow & J. Loney (Eds.), *Psychosocial aspects of drug treatment for hyperactivity* (pp. 325–378). Boulder, CO: Westview.

Goodwin, M. S., & Shinn, M. R. (1990, April). *Assessment of educational outcomes: Are we using appropriate strategies to demonstrate effects?* Paper presented at the National Association of School Psychologists convention, San Francisco, CA.

Knutson, N., & Shinn, M. R. (1991). Curriculum-based measurement: Conceptual underpinnings and integration into problem-solving assessment. *Journal of School Psychology, 29,* 371–393.

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.

Marston, D., Fuchs, L. S., & Deno, S. L. (1986). Measuring pupil progress: A comparison of standardized achievement tests and curriculum-related measures. *Diagnostique, 11,* 77–90.

Marston, D., Lowry, L., Deno, S. L., & Mirkin, P. K. (1981). *An analysis of learning trends in simple measures of reading, spelling, and written expression: A longitudinal study* (Research Report No. 49). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Power, M. D., & Franks, C. M. (1988). Behavior therapy and the educative process. In J. C. Witt, S. N. Elliott,

& F. M. Gresham (Eds.), *Handbook of behavior therapy in education* (pp. 3–36). New York: Plenum.

Rapport, M. D. (1987). Attention deficit disorder with hyperactivity. In M. Hersen & V. B. Van Hasselt (Eds.), *Behavior therapy with children and adolescents: A clinical approach* (pp. 325–361). New York: Wiley.

Rapport, M. D., DuPaul, G. J., & Kelly, K. L. (1989). Attention-deficit hyperactivity disorder and methylphenidate: The relationship between gross body weight and drug response in children. *Psychopharmacology Bulletin, 25,* 285–290.

Rapport, M. D., DuPaul, G. J., Stoner, G., & Jones, J. T. (1986). Comparing classroom and clinic measures of attention deficit disorder: Differential, idiosyncratic, and dose-response effects of methylphenidate. *Journal of Consulting and Clinical Psychology, 54,* 334–341.

Rapport, M. D., Jones, J. T., DuPaul, G. J., Kelly, K. L., Gardner, M. J., Tucker, S. B., & Shea, M. S. (1987). Attention deficit disorder and methylphenidate: Group and single-subject analyses of dose effects on attention in clinic and classroom settings. *Journal of Clinical Child Psychology, 16,* 329–338.

Rapport, M. D., & Kelly, K. L. (1991). Psychostimulant effects on learning and cognitive functioning: Findings and implications for children with attention deficit hyperactivity disorder. *Clinical Psychology Review, 11,* 61–92.

Rapport, M. D., Stoner, G., DuPaul, G. J., Birmingham, B. K., & Tucker, S. (1985). Methylphenidate in hyperactive children: Differential effects of dose on academic, learning and social behavior. *Journal of Abnormal Child Psychology, 13,* 227–244.

Rapport, M. D., Stoner, G., DuPaul, G. J., Kelly, K. L., Tucker, S. B., & Schoeler, T. (1988). Attention deficit disorder and methylphenidate: A multilevel analysis of dose-response effects on children's impulsivity across settings. *Journal of the American Academy of Child and Adolescent Psychiatry, 27,* 60–69.

Safer, D. J., & Krager, J. M. (1988). A survey of medication treatment for hyperactive/inattentive students. *Journal of the American Medical Association, 260,* 2256–2258.

Shapiro, E. S., & Kratochwill, T. R. (1988). *Behavioral assessment in schools: Conceptual foundations and practical applications.* New York: Guilford.

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

Shinn, M. R. (Ed.). (1993a). *Administration and scoring of curriculum-based measurement.* CBA Training Institute Module. (Available from the School Psychology Program, College of Education, University of Oregon, Eugene, OR 97403)

Shinn, M. R. (Ed.). (1993b). *Creating measurement materials.* CBA Training Institute Module. (Available from the School Psychology Program, College of Education, University of Oregon, Eugene, OR 97403)

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory factor analysis of its relationship to reading. *School Psychology Review, 21,* 459–479.

Sprague, R. L., & Sleator, E. K. (1977). Methylphenidate in hyperkinetic children: Differences in dose effects on learning and social behavior. *Science, 198,* 1274–1276.

Sprague, R. L., & Werry, J. S. (1971). Methodology of psychopharmacological studies with the retarded. In N. R. Ellis (Ed.) *International review of research in mental retardation* (Vol. 5, pp. 147–219). San Diego, CA: Academic.

Swanson, J., & Kinsbourne, M. (1975). Stimulant-related state-dependent learning in hyperactive children. *Science, 192,* 1354–1357.

Weiss, G., & Hechtman, L. (1986). *Hyperactive children grown up.* New York: Guilford.

White, O. R., & Haring, N. G. (1981). *Exceptional teaching* (2nd ed.). Columbus, OH: Merrill.

Received July 7, 1993
Initial editorial decision September 14, 1993
Revisions received October 19, 1993; November 29, 1993
Final acceptance November 29, 1993
Action Editor, Brian K. Martens