4-2003

An Experimental Analysis of the Effects of Reading Interventions in a Small Group Reading Instruction Context

Christine M. Bonfiglio
Western Michigan University

AN EXPERIMENTAL ANALYSIS OF THE EFFECTS OF READING INTERVENTIONS IN A SMALL GROUP READING INSTRUCTION CONTEXT

by

Christine M. Bonfiglio

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
April 2003

AN EXPERIMENTAL ANALYSIS OF THE EFFECTS OF READING INTERVENTIONS IN A SMALL GROUP READING INSTRUCTION CONTEXT

Christine M. Bonfiglio, Ph.D

Western Michigan University, 2003

The validation of specific reading intervention components in increasing oral reading fluency is evident within the literature for academic responding. However, much of the literature examines these components using single-case designs on an individual basis across students.

This study utilized response-guided experimentation in the context of a small group reading instruction with four 4th graders (three male and one female) while examining individual patterns of student responding for oral reading fluency. Acquisition, fluency, and consequential components were examined in a dismantling fashion. Sequential modifications of the treatment package and its components were utilized within a multiple-baseline design in order to identify the most effective, yet efficient reading intervention package.

The results indicated that all potential combinations of the package were effective in producing favorable effect sizes over the typical instruction control condition. However, an effective, more efficient package subtracting the reward component was identified and produced substantial increases in performance across all students.

ACKNOWLEDGMENTS

I would like to begin by acknowledging Dr. Edward Daly, III of the University of Nebraska-Lincoln. It is through his guidance, support, and belief in my work that I have been so credibly trained. It has been a pleasure working with him these past years in research that we find both rewarding and critical for positive student outcomes.

Secondly, I would like to thank Dr. Ruth Ervin, Dr. Linda LeBlanc, and Dr. Scott Gaynor of Western Michigan University for their instruction and support throughout the past two years. Moreover, I appreciate their work on this committee. It has been a joy working with them and I have received superior training throughout their instruction.

Third, I would like to give a special thanks to my husband, Dave, and my children, Jonathon and Ciana. They have endured and sacrificed many hours, as I have worked tirelessly to complete this manuscript and degree.

Finally, but most importantly, I would like to give a grateful thanks to the Lord, Jesus Christ. He has blessed me with so many gifts and allowed me the ability to use God-given talents. Furthermore, he has given me the strength each day to rise and drive the hour and fifteen minutes each day to Kalamazoo for this education.

Christine M. Bonfiglio

TABLE OF CONTENTS

Table of Contents—Continued

Table of Contents—Continued

Table of Contents—Continued

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Reading Statistics

There is growing concern in society about the increasing numbers of academic problems within the educational system. In 2000, 37% of fourth graders read below the "Basic" (i.e., instructional) level (National Center on Educational Statistics, 2001). Furthermore, the National Center on Educational Statistics (2001) indicates that the reading scores of the lowest 25th percentile of fourth graders have declined over the past eight years. To address these concerns, school psychologists must have substantial sophistication in both assessment and intervention to assist in remediating these problems. Indeed, the revised training standards of the National Association of School Psychologists require that school psychologists have competence in effective instruction and development of cognitive/academic skills (National Association of School Psychologists, 2000), reflecting demands for a new standard of proficiency in assessment and intervention skills (Reschly & Ysseldyke, 1995; Ysseldyke et al., 1997).

Descriptive Analysis

Functional assessment is the process of generating information on the events preceding and following a behavior of concern, in an attempt to identify which antecedents and consequences are reliably associated with the occurrence of the behavior (Miltenberger, 2001). One method of functional assessment is descriptive

analysis (Mace & Lalli, 1991). Descriptive analyses allow for the direct observation of problematic behaviors in the natural environment, as well as the identification of the natural covariation between the problematic behavior and specific environmental events (Lalli, Browder, Mace, & Brown, 1993). Information on antecedents and consequences and their relationship to target behaviors can be obtained via interviews, direct assessments (i.e., through traditional testing, curriculum-based measurement), and/or direct observations. Direct observation is regarded as the most important form of descriptive analysis, as it outlines variables in relation to behavioral and/or academic problems within the natural setting (Lentz & Shapiro, 1986). One fundamental principle is that problem behaviors have an essential linkage to the antecedent and ecological context in which they occur (Bijou, Peterson, & Ault, 1968). Bijou et al. provide four ingredients of a descriptive analysis. They include: (a) specifications of the situation in which the study is being conducted, (b) definitions of behavioral and environmental events in observable terms, (c) measurements of observer reliability, and (d) procedures for collecting, analyzing, and interpreting the data.

The outcome of descriptive analysis provides the evaluator with hypotheses about the function of antecedents and consequences in maintaining the problem behavior (Lalli et al., 1993). Descriptive studies provide information on events and their occurrence. They do not provide information on the functional properties of the events or the functional relationships (Bijou et al., 1968). A functional analysis (i.e., experimental analysis) is needed to verify that the naturally occurring variables are

functionally related to the maintenance of problem behaviors (Taylor & Romanczyk, 1994). A functional analysis involves exposing an individual to each of the possible maintaining conditions in a tightly controlled experimental design in an attempt to identify the function maintaining problematic behavior (Gresham, Watson, & Skinner, 2001).

Experimental Analysis and Its Importance

An experimental analysis is superior to descriptive assessment for identifying a causal relationship (Lerman & Iwata, 1993). Two advantages of conducting experimental analyses are its objectivity and quantitative precision (Iwata, Vollmer, & Zarcone, 1990). The purpose of conducting an experimental analysis is to identify operant or instructional contingencies maintaining troublesome behavior, be it behavioral excess or a behavioral deficit (Martens, Witt, Daly, & Vollmer, 1999). However, there may be good reason to conduct a formal descriptive analysis prior to an experimental analysis (Lalli et al., 1993; Mace & Lalli, 1991; Sasso et al., 1992). Doing so might result in shorter, more efficient experimental analyses because only the variables that appear to be relevant through descriptive analyses would be experimentally manipulated.

A number of researchers have been successful in treating problem behaviors with the paired use of descriptive and experimental analyses (Dunlap, Kern-Dunlap, Clarke, & Robbins, 1991; Lalli et al., 1993; Mace & Lalli, 1991). However, to date the research has mainly focused on behavioral problems, thus neglecting academic difficulties. Only a small number of studies have applied brief functional analyses to

problematic academic behaviors (Daly, Martens, Dool, & Hintz, 1998; Daly, Martens, Hamler, Dool, & Eckert, 1999; McComas et al., 1996). Ervin et al. (2001) report that only 6.8% of the functional assessment literature has involved academic performance as the target behavior.

Academic Engagement and Academic Responding

One criticism of the functional analysis literature is its dichotomous categorical measurement of "on-task" vs. "off-task" behavior (Hoge & Andrews, 1987; Lentz, 1988). Active engagement as a performance target has been shown to lead to greater generalized improvements in overall achievement than when the target behavior was "on-task" (Hoge & Andrews, 1987; Lentz, 1988). Active engagement is defined as the proportion of time that a student actively responds to relevant academic tasks and is a variable under the control of educators that is directly related to student achievement (Greenwood, Terry, Marquis, & Walker, 1994). Furthermore, Greenwood (1996) suggests that increasing the quantity of academic engagement can be expected to increase measured achievement.

There have been attempts to develop measurement systems for academic responding in order to assess academic performance. Two such systems are the Behavioral Observation of Students in Schools [BOSS (Shapiro, 1996b)] and the Code for Instructional Structure and Student Academic Response [CISSAR (Stanley & Greenwood, 1981)]. The BOSS is a measure designed specifically for direct observation of academic skills. It encompasses the recording of academic engaged and non-engaged time in which the interaction of the two provides a picture of the

student's behavior in context of meaningful academic outcomes (Shapiro, 1996b). The CISSAR seeks to combine the classroom ecology, student behavior and teacher behavior within a single taxonomy (Greenwood, Carta, Kamps, Terry, & Delquadri, 1994). It was validated in a preliminary study of the instruction and achievement of disadvantaged students attending Title 1 and shown to be sensitive to ecological differences in instruction and student behavior (Greenwood, Delquadri, & Hall, 1984; Stanley & Greenwood, 1983). Although these tools attempt to measure academic engagement in assessing academic performance, they provide little guidance about the selection of intervention components. Furthermore, they have not been incorporated in the analysis of instructional or motivational variables within a broader evaluation framework for intervention design in the way that behavioral assessment technology has been incorporated in functional assessment methodology for problem behaviors. Finally, these measures provide estimates of behavior under the broad construct of engagement without specifying which academic responses are occurring or not occurring.

A direct measure of academic engagement that is amenable to calculating rate is active student responding (Daly & Murdoch, 2000). Academic responding occurs when a student emits a detectable response to ongoing instruction (Heward, 1994). Some advantages to quantifying academic responding are: (a) the measures are easy to construct and can be used to describe rate; (b) it provides an index of the behaviors of interest in the academic environment under the effects of typical classroom factors, which allows teachers the assurance that appropriate responses are occurring and

opportunities to provide corrective feedback as needed; (c) it is sensitive to changes in the classroom; and (d) it can be readily incorporated into a functional assessment for students' academic performance problems (Daly & Murdoch, 2000). Thus, academic responding is a viable, measurable form of target behavior when assessing academic performance. For example, reading fluency can be measured as the number of correctly read words per minute (Shinn, 1989). When reading fluency estimates are obtained in curricular materials, oral reading fluency as a type of academic responding specifies not only relevant dimensions of reading behavior (i.e., accurate and fluent oral reading), but also the context in which those behaviors should be displayed (i.e., in curricular materials). These features make it a particularly useful form of active responding for measurement and instructional purposes. Indeed, the report by the National Reading Panel (National Center on Educational Statistics, 2002) stresses the critical role of reading fluency in achieving overall reading proficiency. Only recently has oral reading fluency been integrated into frameworks for conducting individualized experimental analyses of student responsiveness to treatment.

Brief Experimental Analyses in Academic Responding

Several studies have combined instructional and motivational variables in experimental analyses of academic responding. McComas et al. (1996) applied brief functional analysis procedures in a multi-element design to analyze academic responding in the areas of spelling and reading comprehension. Results of this study

suggested that brief experimental analyses may be helpful in identifying instructional components that may be effective in instruction.

Daly et al. (1998) applied test conditions to identify effective reading instructional strategies in order to increase oral fluency rates. This study utilized a multi-element design to implement baseline conditions and treatment conditions until a visible effect was identified, after which a mini-reversal was conducted repeating the last ineffective treatment followed by the effective treatment to demonstrate experimental control (Daly et al., 1998). The purpose of the study was to identify the most effective intervention (i.e., the one that produced the largest outcome) in the most efficient manner. Daly et al. (1998) reported through this process that although the results do not guarantee that an effective treatment will be identified, the data are useful for eliminating ineffective interventions that do not produce immediate effects.

Daly et al. (1999) utilized brief functional analyses wherein instructional reading components were sequentially stacked in an attempt to identify the most effective instructional package for increasing oral reading fluency. Using a brief multi-element design, Daly et al. applied the treatment package in a hierarchical manner to develop individualized treatment recommendations for students. When a treatment or combination of treatments produced an improvement in student responding, a mini-reversal was conducted to confirm the effects across passages. This study required fewer test sessions than Daly et al. (1998), thus making the process more efficient (Daly et al., 1999).

The results of these studies indicate that functional analyses can be conducted successfully with interventions targeting students' academic responding. However, these studies are among the few addressing academic performance in the functional assessment/analysis literature. Furthermore, the data from these studies were collected from individual students in the context of an analogue setting. I have been unable to find any studies that have been conducted in the natural educational environment (i.e., the classroom) amongst the naturally occurring variables (i.e., teacher, curriculum, and peers) that directly target academic responding.

Conceptual Models for Treatment Selection

One of the challenges to assessing academic responding within the instructional context in which they should appear is that an almost limitless number of variables exist within the curriculum and environment that may contribute to the occurrence or non-occurrence of such behaviors (Dunlap & Kern, 1996). Nonetheless, a finite number of robust principles has emerged that could be incorporated productively into a more comprehensive framework for assessing and intervening with academic responding targets. For example, increasing opportunities to respond increases fluency rates (Eckert, Ardoin, Daly, & Martens, 2002; Levy, Nicholls, & Kohen, 1993; Skinner, Ford, & Yunker, 1991) and academic achievement (Lentz, 1988). Providing more time to learn yields an increase in the opportunity and frequency of student responses (Paine, Radicchi, Rosellini, Deutchman, & Darch, 1983). Reducing the amount of time teachers and students engage in non-academic activities will provide more academic learning time (Gettinger, 1995). Thus, for

starters, evaluating the number of opportunities to respond, as well as the number of responses within instruction, will provide valuable information. As a potential intervention target, learning rates can be increased by providing more opportunities to respond within the same period of time (Berliner, 1984).

Selection of treatment components for academic analyses has been guided in part by conceptual models of instruction that are grounded in empirically derived behavioral processes. Reliance on robust principles of behavior as they apply to behavior deficits has been productive thanks to several models that have conceptualized how treatments interact with student responding in various ways. The most directly relevant models include the delineation of skill versus performance deficits (Lentz, 1988), the Instructional Hierarchy for conceptualizing instructional components (Daly, Lentz, & Boyer, 1996), the conceptualization of academic engagement and opportunities to respond (Greenwood, Hart, Walker, & Risley, 1994) and the emphasis of instructional practices within behavior analysis (Wolery, Bailey, & Sugai, 1988). These sources have provided investigators with ways to select and order treatment components in experimental analyses.

Lentz (1988) makes a distinction between skill-based versus performance-based deficits. Skill-based deficits occur when a student does not possess requisite skills for success. For example, a student may not be fluent, may not have adequate sight word vocabulary, or may not possess efficient word attack skills. On the other hand, if a child does possess sufficient skills, poor academic performance may be the result of performance-based deficits. That is, a child lacks sufficient motivation to

perform acquired skills. Gresham (1981a,b) conceptualized performance deficits in terms of an individual possessing a particular skill in the behavioral repertoire, but failing to perform it at acceptable levels. Witt, Daly, and Noell (2000) stated that probably as many as 25% of children who don't perform up to par in the classroom can actually do the work but prefer not to. This differentiation of skill- versus performance deficits may allow one to select a treatment based on hypothesized maintaining variables.

The Instructional Hierarchy is a conceptual framework for refining the idea of academic responding within a learning hierarchy (Haring, Lovitt, Eaton, & Hansen, 1978). Haring et al. (1978) describe learning as consisting of four stages: (a) acquisition [i.e., "the period between the first appearance of the desired behavior and the reasonably accurate performance of that behavior" (p.25)], (b) fluency (i.e., the rapid and proficient performance of the skill), (c) generalization (i.e., the process of displaying the skill in appropriate contexts and/or multiple settings), and (d) adaption (i.e., the modification of the learned skill in the face of novel environmental demands). When a new skill is being taught, the learner must first acquire it. The learner next becomes fluent in the use of this skill. If the learner is fluent and instruction is appropriate, he or she is more likely to generalize its use to novel contexts. Finally, the learner adapts the use of the skill to modify the response in order to accommodate its use to novel demands (Haring et al., 1978).

This framework encompasses different levels of academic responding and each level has corresponding instructional procedures that will efficiently lead to

mastery at that level (Haring et al., 1978). For example, modeling, cuing, and demonstration are particularly important during the acquisition phase (Haring & Eaton, 1978). Directly prompting accurate responses may be the most effective way to instruct students during the acquisition phase (Haring & Eaton, 1978). Moreover, demonstration and modeling may be the most common procedures used to increase accuracy because of their efficiency and effectiveness (Hendrickson & Gable, 1981). When specific levels of academic responding are accurately defined and corresponding instructional procedures are correctly implemented, increases in student responding are more probable (Daly et al., 1996). Thus, the use of the Instructional Hierarchy provides researchers and practitioners with a conceptual framework to accurately identify skill deficits more precisely by helping to identify the specific level at which academic responding is deficient.

Treatments for Reading Fluency Problems

Many treatments for reading fluency problems have been investigated. Some of these include Taped Passage Preview, Choral Reading, and Word or Phrase Drill. Taped Passage Preview (TPP) is a technique in which the passage is recorded by an adult or peer. The student follows along with his or her finger while listening to the pre-recorded passage. TPP is a form of modeling which has been shown to be the most effective way of instructing students during the acquisition phase (Haring & Eaton, 1978). Moreover, modeling has been shown to be effective in increasing the accuracy of responding (Espin & Deno, 1989; Lalli & Shapiro, 1990). In an

examination of TPP, Rose and Beattie (1986) found listening preview to be effective in increasing oral reading fluency rates in reading passages.

Choral Reading (CR) is a technique that allows all students to read the specified passage together aloud (Heward, 1994). Increasing opportunities to respond has been shown to increase fluency rates (Eckert, et al., 2002; Levy, et al., 1993; Skinner, et al., 1991). Research on choral responding has produced similar results: higher rates of responding, more accurate responding, and better on-task behavior (Carnine, 1976; Sainato, Strain, & Lyon, 1987).

There is evidence that word drill (WD) increases accuracy and fluency rates relative to an absence of an error correction procedure (Rosenberg, 1986; Tan, Moore, Dixon, & Nicholson, 1994). WD is a procedure utilized immediately following an error, whereby the student listens to the teacher model the pronunciation correctly and then repeats it three times. Leper (1985) stated that immediate feedback is an essential component of the learning trial in which the goal is to increase accuracy. This is important because students often make incorrect responses during the acquisition phase. Furthermore, WD encompasses the technique of modeling. Modeling aids in the acquisition phase whereby the behavior of reading is not stable and occurring less frequently than what is expected of a mastered skill (Daly et al., 1996).

Other utilized techniques in the reading literature include the use of praise and contingent reinforcement. Praise is a form of teacher attention that has been documented to increase student attentiveness (Cossairt, Hall, & Hopkins, 1973; Hall,

Lund, & Jackson, 1968). Praise can be utilized as social positive reinforcement, which is intended to increase the probability of accurate responses, whether they are discrete academic responses or appropriate forms of academic engagement. Contingent reinforcement (i.e., positive reinforcement) is a type of reinforcement in which, contingent on a specified behavior, a stimulus is presented and the probability of the response increases (Miltenberger, 2001). Delivering a reward based on the achievement of a pre-specified goal (e.g., a specified number of correctly read words in 1 minute) would be one example. Utilizing this procedure has been shown to be effective in increasing reading fluency rates (Billingsley, 1977; Jenkins, Barksdale, & Clinton, 1978). Furthermore, there is evidence that providing reinforcement contingent on a specified goal improves reading performance (Billingsley, 1977; Jenkins et al., 1978).

Design Issues

Single-case designs constitute an experimental tool that utilizes objective information, continuous assessment of performance over time, and the reliance on stable levels of performance before and after treatment wherein each participant serves as his/her own control (Kazdin, 1982). It evolved because of the need to understand patterns of *individual* behavior in response to specific variables to promote the implementation of effective interventions (Morgan & Morgan, 2001). While there is many options in design tactics, two basic sequential and/or hierarchical tactics that fit practices related to special service delivery decisions are the increasing intensity design (i.e., adding intervention components sequentially) and the

decreasing intensity design [i.e., dismantling an intervention by removing components (Barnett, Daly, Jones, & Lentz, 2003)].

Increasing intensity design. The increasing intensity design involves step-by-step decisions wherein interventions are attempted and become more intensive and intrusive only as necessary (Fuchs & Fuchs, 1998). This design begins with a baseline and through the evaluation of changing conditions; the least amount of intervention that will bring about desired change is selected (Barnett et al., 2003).

For example, following earlier work of Cooper and Wacker and colleagues (Cooper et al., 1992), Harding and his colleagues utilized brief experimental analysis wherein they directly applied treatment to behaviors of concern. This research evaluated a large sample of young children with aberrant behavior. The assessment conditions were arranged in a least-to-most intrusive hierarchical manner (Harding, Wacker, Cooper, Millard, & Jensen-Kovalan, 1994). Harding et al. (1994) had parents implement treatments while they evaluated interventions through a series of rapidly changing assessment conditions [i.e., free play, general directions and discussion (both of which served as control conditions), specific directions, choice making, differential reinforcement of appropriate behavior, differential reinforcement of communication, preferred activities, and time-out and guided compliance]. In this study, Harding et al. (1994) correctly identified a treatment that improved behavior for all 7 children. Furthermore, follow-up results substantiated the continued use of the treatment, treatment integrity, and parental satisfaction. As previously stated, Daly, his colleagues, and students have been refining a model for evaluating potential

components for instructional interventions building upon this earlier work (Daly et al., 1998; Daly et al., 1999). One of the principle changes they have made to the analysis is to add intervention components one-by-one without removing instructional or motivational components.

Dismantling procedure. In contrast to the stacking procedure, the dismantling procedure, also known as a decreasing intensity design (Barnett et al. 2003), begins with an intense intervention design (usually a multi-component strategy) that has been proven to be effective and components are withdrawn as soon as possible (Barnett et al., 2003). The withdrawal of intervention components allows for intervention designs to become more ecologically sustainable (e.g., Dooley, Wilczenski, & Torem, 2001).

Dooley et al. (2001) conducted a study utilizing a dismantling procedure (i.e., decreasing intensity design) wherein a two-component intervention package was implemented for a child with Pervasive Developmental Disorder (PDD) exhibiting disruptive behaviors. Following baseline and the implementation of the reinforcement package, the more intrusive component was withdrawn and levels of disruptions and compliance were examined (Dooley et al., 2001). This type of design allows for decisions about necessary services to be empirically validated (Barnett et al., 2003).

Data analysis. In applied investigations utilizing single-case designs, experimental criteria (i.e., determination of whether the intervention has had an effect) and therapeutic criteria (i.e., determination of whether the effects are important or of clinical or applied significance) are used to evaluate data (Risley, 1970). The

experimental criterion refers to a comparison of performance during the intervention to that of baseline data (Kazdin, 1982). To evaluate the experimental criteria, data were analyzed via visual inspection. According to Kazdin (1982), "Visual inspection refers to reaching a judgment about the reliability or consistency of intervention effects by visually examining the graphed data" (p. 232). Visual analysis is conducted to answer two questions: (a) Did a meaningful change in behavior take place? and (b) To what extent can that change be attributed to the experimental manipulation?

Analysis of data is an ongoing task, rather than a process that occurs at the end of a study (Wolery et al., 1988). Wolery and Harris (1982) suggest the following four-step process in analyzing the effects of data utilizing visual inspection.

1) Periodically assess the reliability with which data are collected and the reliability with which the intervention is implemented.
2) Determine whether changes occurred in the data series.
3) Determine whether changes occurred between instructional conditions.
4) Determine whether the changes were useful or worthwhile.

Trend, level, and variability are the three characteristics analyzed through visual inspection (Kazdin, 1982). Cooper, Heron, and Heward (1987) describe these characteristics as the following. Trend refers to the overall direction or path taken by the data. The value on the vertical axis around which a set of behavioral measures converges is called the level. Variability refers to the extent to which measures of behavior differ from one another under the same environmental conditions. These characteristics should be evaluated both within phases and between phases. Within

and between phases, stable patterns of data are preferred over variability (Kazdin, 1982). Sidman (1960) describes this pattern as a steady state. It is needed to ensure that manipulation of the independent variable is responsible for changes in measured behavior [i.e., the dependent variable (Cooper et al., 1987)]. Shifts or discontinuity of performance from the end of one phase to the beginning of the other (i.e., level changes) between phases indicate that the independent variable was responsible for change in behavior (Kazdin, 1982). Within phases, if a trend occurs during baseline, it is preferred that it be in the opposite direction of the desired behavior change. If implementation of treatment produces a trend in the desired direction, it can be assumed that it was the independent variable that was responsible for behavior change (Kazdin, 1982).

Cooper et al. (1987) outline the following major advantages of graphic display and visual analysis of behavioral data.

1) The investigator has ongoing access to a complete record of the subject's behavior because it is plotted immediately after the observational period.
2) The investigator can explore interesting variations in behavior as they occur because he/she has direct, continuous contact with it.
3) Little time or special equipment is needed to utilize graphing, unlike statistical analysis. Furthermore, it is relatively easy to learn and no predetermined or arbitrary level for evaluating significance of behavior change is needed.

4) Visual inspection is a conservative method of determining the significance of behavior change, as variables that produce only weak or unstable effects, which are obtained with statistical methods declaring significance, are not likely to be reported.
5) The graph allows and encourages independent judgments and interpretations of meaning and significance.
6) The graph can be an effective source of feedback for the person whose behavior is represented.

Despite the many advantages, one should caution the use of visual inspection when phases are highly variable, baselines are unstable or have a trend in the undesired direction, and/or weak changes occur as a result of manipulation (Kazdin, 1982).

Curriculum-based measurement (CBM) is a systematic set of procedures designed for progress monitoring which employs repeated and frequent administration of skill probes taken from the curriculum. They are designed to be reliable, provide direct assessment of skills, are repeatable and sensitive to student growth, and moreover, can assist in deriving appropriate strategies for academic performance (Shapiro, 1996).

CBM offers educators of both special and general education classrooms an alternative to traditional standardized achievement testing (Marston & Magnusson, 1985). Marston & Magnusson (1985) demonstrated that CBM may be successfully utilized in making decisions regarding screening, identification, program planning, progress monitoring, and program evaluation. Moreover, the use of CBM to predict

performance on state tests, as well as to identify students at-risk for low reading skills and failing such tests has been investigated and a moderately strong relationship between oral reading fluency and performance on state tests was established (McGlinchey & Hixson, in press).

CBM is used to assess oral reading fluency by means of measuring correct read words per minute (CRW), as well as errors per minute (ER). Data are graphed and are examined via visual inspection (Shinn, 1989). Fuchs and Deno (1982) provide normative standards for levels of reading fluency (i.e., frustration, instructional, and mastery). Fourth grade standards include the following:

1. Frustration: <70 CRW with >6 errors.
2. Instructional: 70-100 CRW with 6 or less errors.
3. Mastery: >100 CRW with 6 or less errors.

Marston and Magnusson (1985) provide proposed reading standards for 4th graders at 110 CRW/min. Additional standards are provided by others including Howell and Nolet (2000) and the University of Oregon (i.e., Dibels for early literacy assessment; grades K-3). McGlinchey and Hixson (in press) report that 72% of students who read at least 100 CRW/min passed the state reading test and 76% of students who read 110 CRW/min passed. Thus, CBM is a useful measurement tool and provides valuable information, as well as a possible ability to predict performance on standardized, mandated state tests.

Data may also be analyzed via examination of effect sizes. An effect size is the separation or lack of overlap between phases due to the independent variable; it

increases with greater differences between means and decreases with greater standard deviations in the phases (Aron & Aron, 1997). Effect sizes can be viewed as an index of individual treatment outcomes and can be used in both research and practice-based evaluations (Busse, Kratochwill, & Elliott, 1995). Cohen (1965) defines qualitatively effect sizes as small (.25), medium (.50), and large (1.00). Effect sizes are used to obtain a standard measure when different outcomes are used within or across studies (Moncrieff, 1998). There have been several methods proposed for the calculation of effect sizes within single-case research studies. They can be grouped according to non-regression vs. regression approaches (Faith, Allison, & Gorman, 1997).

1. Non-regression approaches: These approaches are an appropriate selection when researchers are investigating changes in level differences in which a measure of central tendency is needed (Busk & Serlin, 1992). They are methods for obtaining a description of the data. Furthermore, they can be used when the intent is to discover the effect of a particular behavioral intervention and to accurately summarize the data on this treatment (Busk & Serlin, 1992). The following are a few of the more popular techniques with a description of their strengths and weaknesses.

   a. Standardized Difference Approaches: The formula $d = (M_t - M_b)/s$ (where M_t is the mean during treatment phases, M_b is the mean during the baseline phase, and s is the standard deviation) has been used to calculate effect size (Gingerich, 1984; Jayaratne, Tripodi,

& Talsma, 1988). It has been borrowed from methods used in group designs and is comparable to that used in group meta-analysis (Faith et al., 1997). Advantages to this method are its compatibility with group analysis, as well as its convenience of implementation. However, Faith et al. (1997) report that this method does not appear to bypass the autocorrelation problem, which can result in biased testing. Furthermore, it only compares levels of behavior across phases ignoring other parameters such as slope.

b. Percent of Nonoverlapping Data (PND): The number of data points in the treatment phase that exceeds the highest number of data points in the baseline phase is divided by the total number of data points in the treatment phase and this number is multiplied by 100 yielding a percentage (Faith et al., 1997). This method is easy to compute and interpret and moreover, is readily understood by most readers (Faith et al., 1997). However, PND can yield misleading results when outliers are present. Furthermore, PND is strongly affected by sample size (i.e., as the number of baseline data points increases, the statistic systematically decreases regardless of actual effect (Allison & Gorman, 1994).

c. Assumption of equality of variances: Busk and Serlin (1992) present an approach which assumes the homogeneity of variances

across baseline and treatment phases and the use of pooled within-phase variances as the denominator in the calculation, which provides a better estimate. With this approach, confidence intervals can be constructed using the Wilcoxon model or the one-sample *t* test (Busk and Serlin, 1992).

2. Regression approaches: These approaches offer a more sophisticated, flexible method for research synthesis by fitting statistical models to observed data. Furthermore, they offer better descriptions of the trends/slope in the data, as well as the effects of the intervention on the behavior (Faith et al., 1997). The following are a couple examples within the literature.

   a. Trend analysis: Gorsuch (1983) reports that trend analysis is the most appropriate analysis in terms of minimizing both Type-I and Type-II errors. It is conducted by evaluating the effects of the intervention in a regression after entering time as a covariate. White, Rusch, Kazdin, and Hartmann (1989) provide a conceptually similar model. The major advantage of these methods is that they statistically adjust for the effects of time. However, these only account for and measure differences in level; differences in slope are not detectable (Faith et al., 1997).

   b. Regression for computing effect size that accounts for trend and slope: Center, Skiba, and Casey (1985-1986) developed a more

sophisticated model in which the effects of level, trend, and slope can be calculated. The major advantage is its flexibility for modeling additional parameters of behavior, such as slope (Faith et al., 1997). However, one must be cautious when using this model, as under certain circumstances the effects of trend may be overestimated, presenting bias in the data. Moreover, changes in level and slope in opposite directions also present biases.

c. Allison and Gorman (1993) have modified the above method and present a 10-step algorithm in order to address the concerns of overestimation and the problem posed by changes in level and slope that occur in opposite directions. However, there are concerns with this model as well. First, at times, baseline regression equations make out of bounds projections into subsequent treatment phases (Faith et al., 1997). Faith et al. (1997) explains that percentage scores have a finite range (i.e., 0 to 100). It is, however, possible for baseline regressions to project values that exceed these limits. Second and more importantly, single-case designs typically contain only a small number of baseline data points. In order to allow for a stable regression model, primary sources ideally should provide a large number of baseline data points (Faith et al., 1997).

Based on the numerous considerations, strengths, and weaknesses regarding regression vs. non-regression approaches, effect sizes were calculated across all subjects and treatment conditions using the first method (i.e., the standardized difference approach). It was chosen since there were a small number of baseline data points and treatment was implemented once a steady state was obtained in the baseline phase. A difference in level was sought since the intervention was not implemented without the steady state in baseline. Furthermore, due to the nature of the multiple-probe design, the intervention was implemented at the very *latest* within 5 weeks of introduction of the passage. Thus, the effects of time were minimal. Moreover, this method was convenient and provided standardized information regarding the effectiveness of treatment intervention across all passages and participants.

Variables Influencing the Selection of Treatment

Treatment acceptability. Treatment acceptability refers to perceptions of whether interventions are fair, reasonable, and appropriate for use with specific target behaviors, as well as consistent with expectations of what treatment should be (Kazdin, 1980a). Assessment of acceptability is critical in that it affects intervention usage and outcomes (Elliott, 1988). Through research, many factors affecting acceptability have been identified and include treatment efficacy, presence of adverse side effects, and the use of jargon in describing treatment procedures (Kazdin, 1980b, 1981). Additional variables include time involvement, behavior problem severity, and type of intervention (Witt, Martens, & Elliott, 1984).

Witt et al. (1984) explored the influence of teacher time involvement, intervention type, and behavior problem severity on teachers' judgments pertaining to acceptability utilizing two written case studies. Results revealed that the amount of time for teacher involvement significantly affected the judgments of intervention acceptability. That is, the higher the time requirement for implementation the lower the acceptability.

A follow-up study investigated the influence of behavior problem severity, interventionist, and modality of case presentation on teachers' judgments of the acceptability of school-based interventions (Martens, Witt, Elliott, & Darveaux, 1985). Martens et al. (1985) manipulated modality of presentation (written vs. videotape) to assess intervention acceptability with two interventions (i.e., missing recess in the principal's office or a response-cost procedure implemented by the teacher) for two problem behaviors (i.e., daydreaming or destruction of property). All combinations of these variables were manipulated and teacher acceptability was measured using the Intervention Rating Profile-15 (Witt, & Martens, 1983). Results indicated that the intervention requiring moderate amounts of time to implement was rated as more acceptable than the less time-consuming alternative because of ratings of effectiveness or severity of punishment. Consequently, the intervention designed to increase appropriate behavior was more acceptable than the reductive intervention (Witt et al., 1984).

Witt and Elliott (1985) summarized the results of behavioral intervention acceptability research. Research in this area has led to the following findings.

1) Interventions differ markedly in acceptability ratings.
2) The severity of the behavior affects treatment acceptability.
3) The presence of side effects as a result of an intervention may negatively affect acceptability ratings.
4) Positive behavioral interventions (e.g., token economies and positive reinforcement) are viewed as more acceptable than reductive-type interventions (e.g., time-out).
5) Acceptability ratings are affected by the effort, resources, and amount of time needed for implementation. That is, the more demanding the intervention, the lower the acceptability rating.

Based on the above findings, acceptability of interventions becomes a critical component in the intervention design process. Teacher acceptance may significantly affect whether an intervention will be implemented with integrity or not. Thus, when devising behavioral strategies within the classroom, it is important to incorporate the above findings. Moreover, assessment of acceptability prior to implementation will alleviate complications with treatment integrity.

The Intervention Rating Profile is an assessment instrument that has been used to evaluate the acceptability of classroom interventions (Witt et al., 1984). It is a series of items presented on a 6-choice Likert-type scale ranging from "Strongly Disagree" to "Strongly Agree". It appears to be sensitive to a variety of variables that influence teachers' perceptions and the potential utilization of interventions (Witt et al., 1984). The original IRP consisted of 20 items; however, it has been reduced to 15

items (IRP-15). The IRP-15 was designed to yield a unitary measure of acceptability, that is, "general acceptability" (Martens et al, 1985). Reliability was assessed using Cronbach's alpha and yielded results of .98 (Martens et al., 1985). Thus, the IRP-15 is a strong measure for assessing teachers' acceptability and utilization of interventions. Although the IRP-15 is a reliable measure of acceptability, it is not without limitations. Tarnowski and Simonian (1992) have reported that a subset of subjects report dissatisfaction with the IRP-15 because of its time-intensiveness when used with multiple treatment comparisons. Furthermore, subjects with limited educational backgrounds have experienced difficulties completing the scales.

Ecological validity. Ecological validity has been described in several places (Bronfenbrenner, 1977, 1979; Martens & Witt, 1988; Neisser, 1976, as cited in MacMann & Barnett, 1999). Applying this concept to intervention research, the basic principles include the following: (1) begin assessment planning at the points that guide the behaviors of parents, teachers, and children; (2) address the significant concerns of parents, teachers, and children; (3) develop logical generalizations from intervention research through comparisons to individual children, behaviors, caregivers, settings, resources, and carefully described interventions; (4) assure that intervention outcomes are consistent with setting expectations; and (5) evaluate the effects of interventions in accord with ecological principles, including the analysis of planned and unplanned outcomes over long time periods (MacMann & Barnett, 1999).

For example, upon identification of a specific intervention that promotes positive changes in behavior (e.g., increases in oral reading fluency), will the intervention produce the same behavior change within the natural environment (i.e., the classroom) upon implementation? That is, does the analogue setting closely resemble that of the natural environment (i.e., the classroom) so that evaluation of the intervention and probability of generalization may occur?

To increase ecological validity of behavioral assessments and interventions, a broad diagnostic perspective and scientifically sound procedures that focus on the interaction of individual behavior with the surrounding environment must be adopted (Carlson, Scott, & Eklund, 1980). Martens and Witt (1988) outline the major assumptions of the ecological perspective in the application of behavioral intervention.

First, behavior systems strive toward homeostasis. That is, to the extent that elements within the ecological system are able to function in a harmonious fashion on a continuous basis, the system is said to be homeostatic (Martens & Witt, 1988). Thus, the classroom can be viewed as an ecology system made up in part of the behavior of students and the teacher; moreover, the interaction thereof, strives for homeostasis. Therefore, change in behavioral regularity is likely to require change in other regularities in order for homeostasis to be maintained (Martens & Witt, 1988).

Second, behavior systems as ecologies maintain finite tolerance to stress. Within the classroom, behavioral regularities often change. When such change exceeds the tolerance level, existing regularities in behavior decay (Martens & Witt,

1988). That is, when one child's behavior changes, others' behavior is affected and thus, variability in the regularity can break down and cause disruption in the classroom.

Finally, behavioral ecologies are closed systems, and what has been done cannot be undone. This assumption suggests that once a behavioral intervention has been implemented, remnants of that innovation will be reflected in the system for some time (for good or for bad). Therefore, once treatment has been applied within the classroom, the effects will not only sustain, but any further treatment implementation will have to be distinguishable from the latter in order for discrimination to occur and for it to be effective (Martens & Witt, 1988).

Martens and Witt (1988) describe the importance of determining the behavioral regularities and the interrelationships between the environment and its key players (in this instance, the classroom and its teacher and students). In order to obtain both efficiency and effectiveness, these issues will have to be addressed in order to restore the ecological balance of the system. Ecological validity is achieved only through examination of the above-described assumptions.

Clarke, Dunlap, and Peck-Stichter (2002) conducted a meta-analysis across 10 journals published between 1980 and 1999, wherein behavior modification for students with emotional behavioral disabilities (EBD) and/or developmental disabilities (DD) was evaluated. The percentage of articles addressing measurement of ecological validity (among other variables) was assessed. Clarke et al., (2002) report that research conducted in typical physical contexts was represented in both the

EBD and DD studies at approximately the 25% level. Thus, the majority of research was conducted in the analogue setting.

Ecological validity is an important factor when assessing and intervening for academic and behavioral concerns. The context in which the intervention may be implemented must inform any and all analyses that are to be conducted, and moreover, should closely resemble the natural environment. When the context of analyses is incorporated, you get greater ecological validity. Because of the vast complexity of the classroom environment and the interrelationships exhibited between individuals and the environment, the regularities play an integral role in the future success of an intervention.

The utilization of functional assessment is an integral piece in assuring that powerful ecological validity is attained within an analogue setting. Obtaining information related to events preceding and following behavior allows the researcher to establish the analogue setting so that it closely resembles the natural environment. The hope is to obtain results from the analogue setting that when implemented in the natural environment will produce similar results. Thus, functional assessment when conducted accurately may be a vital contributor to the intervention process because behaviors obtained during experimental manipulation may be more generalizable to and maintained in the natural environment.

Purpose of This Study

The purpose of this study was to integrate descriptive and experimental analyses within an elementary class reading group in an analogue setting in order to

identify potential variables that may serve to improve academic responding, thus increasing academic engagement and oral reading fluency. The small group reading instruction (i.e., the analogue setting) was designed so that it closely resembled key components of the natural environment (i.e., the classroom), in order that results could be generalized more readily. While incorporating elements of functional assessment, its intended strong internal validity was obtained, which in turn, increased ecological validity.

The experimental manipulation occurred with the experimenter implementing a treatment package of empirically validated techniques (i.e., TPP, CR, WD, CTR) within a multiple-baseline design. These techniques rely on key components within the Instructional Hierarchy (i.e., acquisition, fluency, and generalization), which serves as the conceptual framework of academic responding (Haring et al., 1978). Furthermore, these techniques also address and allow for remediation for both a skill-based and/or a performance-based deficit. The treatment package was dismantled until the most efficient, yet effective, package was identified. Because of the nature of a dismantling procedure [i.e., components of a powerful multi-component intervention are withdrawn so that the intervention becomes more ecologically sustainable (Dooley et al., 2001)], results of increased oral reading fluency were expected with the full package, and withdrawal of a specific component would provide information as to its utility. The goal was to determine the simplest treatment package (i.e., requiring the least effort for the teacher), albeit a package that would produce increased oral reading fluency (i.e., the most effective) across all students.

Results were evaluated through visual inspection of the data, as well as through examination of effect sizes and the number of CRW and errors. As the theory of the Instructional Hierarchy provides, students learn in different stages and moreover, different skills are necessary throughout these stages in order to produce results (Haring et al., 1978). Thus, careful consideration was taken to identify students of equal ability (i.e., similar rates of oral reading fluency within their instructional level) and to create an environment that targeted students' levels of learning within the Instructional Hierarchy via the functional assessment.

A variety of student, teacher, and peer variables were measured utilizing both direct observation and curriculum-based measurement. Active responding, quantified as fluency within the curriculum, was targeted. Measurement throughout the experimental manipulation was conducted within an evaluation framework, as it guided decisions regarding withdrawal and/or re-implementation of specific reading remediation techniques, in order to identify the treatment package that successfully increased oral reading fluency rates with the most efficiency across all participants. Moreover, careful attention was given to the teacher's reading technique preferences, as the intervention package for reading instruction was implemented by the classroom teacher following the experimental manipulation, during small group reading instruction. Thus, the teacher was given the opportunity to examine the treatment package's efficacy and efficiency firsthand. Finally, acceptability was evaluated with the IRP-15, yielding favorable acceptability results.

It was hypothesized that at least one effective treatment would be identified based on the rigorous analysis of the variables. Through the functional assessment of student characteristics, classroom variables, and reading techniques, as well as the continuous visual inspection and examination of the data (i.e., effect sizes, number of CRW's and errors), a multi-component treatment package of reading remediation techniques could be experimentally manipulated to identify the "ideal package", based on effectiveness and efficiency. It was also hypothesized that the treatment package selected as the most effective and efficient would be so, quite possibly for not all four of the students, but for the majority of them. Furthermore, any package selected would have to produce increases in oral fluency rates across all participants, but would not have to be the "best" package for each student. This would allow for meeting the criteria of efficiency (i.e., it would be simpler in regards to effort and time for the teacher). Subsequently, with an efficient package, it was hypothesized that it would be more acceptable to the teacher, which might lead to high treatment integrity.

The results of this study provide evidence as to the applicability of descriptive and experimental analyses in the context of the classroom setting addressing deficiencies in academic responding (i.e., reading fluency). Furthermore, external and ecological validity were increased as treatment occurred within the context of the reading group, as well as the implementation of the effective package by the classroom teacher during the final week of the study.

Method

Participants and Setting

One reading group, consisting of four 4th grade children from the same elementary classroom, served as participants in this study. Three of the children were males (Blake, Cody, and Devon) and the fourth child a female (Karla). The ethnic background of the group was diverse, comprising two Caucasians, one African-American, and one Hispanic. The students were identified as instructionally low readers (i.e., reading at instructional level below his/her grade level) by their elementary teacher; however, none received special education services. Participants' gender, age, grade, and screening results are displayed in Table 1.

Table 1

Participant Information

Student	Gender	Age	Grade Level	Instructional Level
Blake	Male	9 yrs 10 mos	4	4th grade
Cody	Male	9 yrs 10 mos	4	3rd grade
Devon	Male	9 yrs 7 mos	4	3rd grade
Karla	Female	10 yrs 5 mos	4	3rd grade

A Human Subjects Institutional Review Board (HSIRB) approved the contents of this study prior to obtaining informed consent. Informed consent was collected via two means. First, a phone call to each parent or guardian was made explaining the nature of the study, possible intervention strategies that would be

utilized, and the right to withdraw for any reason at any time. Furthermore, parents were informed that a permission form was being sent home with the student explaining the study in more detail and each parent was asked to sign and return the consent with the student. Participants received the consent forms to have signed by a parent or guardian and returned them after signatures were obtained. All participants were advised of the study and assent was gathered prior to the beginning of any data collection. Following consent and assent, screening was conducted using curriculum-based measures of reading fluency to determine instructional level.

The direct observation and intervention procedures were carried out in a classroom during the reading group. The experimenter implemented the reading group four days per week, typically Monday through Thursday. Curriculum-based assessment was conducted individually with each student at a small table in the school psychologist's office four days per week as well, typically Tuesday through Friday.

Materials

Instructional reading passages. Reading passages were obtained from the Houghton Mifflin Reading Series. Only narrative and expository texts were used. Readability formulas were used to identify the difficulty level of passages. The Spache formula was used to calculate readability scores for the passages (Spache, 1953). Eight passages were identified and assigned to a specific week in a random fashion (six initially and two additional passages halfway through the study, as it appeared that more would be needed to complete the experimental analysis). These

passages were used for instructing small group reading. The readability and length (number of words) of the passages are presented in Table 2.

Table 2

Readability Level and Length of Passages

Instructional Passages	**Instructional Level**	**Length of Passage (in words)**	**Readability Level**
1	4	162	4.3
2	4	232	4.3
3	4	208	4.4
4	4	237	4.7
5	4	173	4.2
6	4	229	4.6
7	4	243	4.5
8	4	247	4.0
Generalization Passages	**Instructional Level**	**Length of Passage (in words)**	**Readability Level**
1	4	203	4.7
2	4	210	4.4
3	4	239	4.3
4	4	184	4.6
5	4	175	4.9
6	4	234	4.4
7	4	217	4.3
8	4	217	5.0
9	4	291	4.4
10	4	216	4.3
11	4	180	4.6
12	4	275	4.6

Generalization reading passages. Generalization reading passages were obtained from the Houghton Mifflin Reading Series. These passages were texts to which participants had not been previously exposed. Only narrative and expository texts were used. Readability formulas were used to identify the difficulty level of passages. The Spache formula was used to calculate readability scores for the passages (Spache, 1953). Twelve passages were identified and randomly assigned to each day of assessment (i.e., Tuesday through Thursday). Each passage was placed back into the pool of passages following its assignment. Thus, specific passages were utilized more than once. These passages were used for measuring generalized outcomes. The readability and length (number of words) of the passages are presented in Table 2.

Measured Variables

Oral reading fluency. Curriculum-based measurement was used to monitor progress for treatment analysis in instructional passages and in the generalization passages. Correctly read words (CRW) and errors per minute were used to assess reading fluency in both the instructional passages and the generalization passages.

Moreover, these measures provided an index as to the rate of fluency when specific reading interventions were implemented, thus assessing the effects of baseline and treatment conditions. A CRW was defined as a word that was pronounced correctly within a 3-s interval with no aid or correction by the experimenter. If more than 3-s elapsed, the word was counted as an error. Other errors included omissions, substitutions, transpositions, and mispronunciations. The

experimenter scored CRW and errors while the student read the passage for 1 min. All sessions were taped using an audiocassette recorder to assess interobserver agreement and treatment integrity.

Direct observation data. Target behaviors were recorded during observations conducted prior to, during, and following intervention implementation. Categories of behaviors included target student behaviors, teacher behaviors, and peer behaviors. Operational definitions are provided in Table 3.

The focus of measured student variables was academic engagement (i.e., opportunities to respond), opportunities to benefit from modeling, task management behaviors, and disruptive and off-task behaviors. Target student behaviors included the following:

1. Academic Engagement: Academic engagement was recorded if the student was observed producing a verbal or written response, attending to instruction, and/or reading aloud or following along while others were reading.
2. Academic Non-Engagement: Academic non-engagement (i.e., exhibiting distracted behaviors, attending to other individuals and/or materials, and staring in another direction) by the target student was also recorded. This measure provided an index of the amount of time spent in behaviors that distracted from academic learning time.
3. Task Management: Task management behaviors (e.g., sharpening pencil, getting out materials, etc.) by the target student were recorded.

Table 3

Operational Definitions for Direct Observations in the Classroom

TARGET STUDENT BEHAVIORS:

Academic Engagement (ACENG):	Student is observed engaging in any behavior relevant to academic instruction. This includes producing a verbal response (answering or asking a question or making a comment towards instruction), attending to instruction (listening and visually attending), reading aloud during group, reading silently (following along while another reads), and/or writing a response.
Academic Non- Engagement (ACANON):	Student is observed as distracted, daydreaming, handling materials other than reading materials being instructed during group, and/or being disruptive (i.e., talking out, crying, throwing things, destroying materials).
Task Management (TASK):	Student is observed participating in an activity that aids in academic instruction such as raising hand, signaling for help, looking for/taking out materials, moving to a new task, and/or sharpening pencil.

Table 3—Continued

Operational Definitions for Direct Observations in the Classroom

TEACHER/PEER BEHAVIORS:

Instruction (INSTR):	Teacher verbally directs a question or an instructional prompt to target student; verbally presents lesson materials, provides instructions, or describes contingencies for completion to target student, models how to respond to an instructional item and/or corrects an incorrect verbal/written response.
Praise (PRAISE):	Teacher is observed making positive statements to the target student.
Disapproval (DISAPP):	Teacher is observed expressing disapproval of student behavior either verbally (e.g., "You shouldn't be doing that!") or physically (e.g., frowning at student).
Deliver Reward (DLVREW):	Teacher or another student is observed allowing access to tangible item, an activity (e.g., extra recess, game, time w/adult), and/or a token to the target student contingent on the target student performing another behavior.
Peer Attention (PEERAT):	Student is observed speaking to, laughing with, watching, and listening to the target student.

The focus of measured student variables was academic engagement (i.e., opportunities to respond), opportunities to benefit from modeling, task management behaviors, and disruptive and off-task behaviors. Target student behaviors included the following:

4. Academic Engagement: Academic engagement was recorded if the student was observed producing a verbal or written response, attending to instruction, and/or reading aloud or following along while others were reading.
5. Academic Non-Engagement: Academic non-engagement (i.e., exhibiting distracted behaviors, attending to other individuals and/or materials, and staring in another direction) by the target student was also recorded. This measure provided an index of the amount of time spent in behaviors that distracted from academic learning time.
6. Task Management: Task management behaviors (e.g., sharpening pencil, getting out materials, etc.) by the target student were recorded.

The emphasis for teacher behaviors was on antecedent and consequent instructional strategies that provide models, prompts, opportunities to respond, feedback, and possible forms of contingent reinforcement. Teacher behaviors included the following:

1. Instruction: Instruction included asking questions, telling/explaining, and modeling by the teacher, correcting an incorrect verbal or written response, and/or describing contingencies for completion.

2. Praise: Verbal praise to the student provided by the teacher was also recorded.
3. Disapproval: Verbal and non-verbal disapproval from the teacher was recorded. The amount of time spent redirecting a student's problematic behaviors provided an index for the amount of time engaged in non-academic time. Examples of disapproval would include the teacher asking a student to return to the task, the teacher shaking her head no and/or the teacher placing her finger to her lips to "shshsh" a child for talking.
4. Delivers Reward: Delivering material rewards (e.g., pencils, baseball cards, candy, pens, stickers, etc.) by the teacher to students contingent on academic responses were also recorded.

Peer behaviors included the following:

1. Peer Attention: The amount of peer attention was recorded in order to obtain a potential function of inadequate academic behaviors (i.e., low academic responding may be due to behaviors that distract from academic engagement). A distinction between positive and negative was not made.

A behavioral observation format developed for this study was used to collect the data (refer to Appendix A). Teacher and peer behaviors were recorded using 10-s partial-interval recording. Student behaviors were recorded via momentary time sampling at 10-s intervals. Each student's behavior was recorded every 4th interval. That is, the student's behavior in seat 1 was recorded during intervals 1, 5, 9, etc. The student's behavior in seat 2 was recorded during intervals 2, 6, 10, etc. and in the

same fashion for students 3 and 4, respectively. Participants were randomly assigned to each of the four seats on a daily basis. The purpose of random assignment was to assure that one student did not sit in the same position each day, thereby avoiding reading the same paragraph/sentences of the passage each day. It was the position of the student that determined which portion of the passage was read (e.g., seat 1 read the first paragraph, seat 2 read the second paragraph, etc.). However, due to some behavioral problems between two students (Cody and Karla) halfway through the study, the investigator randomly assigned a student to seat one and then seated the remaining three students in a specific order so as not to seat these two students together. Although the remaining students were specifically placed in certain seats, the investigator was careful to rotate the students so as to ensure each participant had the opportunity to sit in all of the different positions throughout the course of the study.

Independent Variables and Treatment Conditions

Treatment strategies were devised for the treatment condition to reflect preview (i.e., Taped Passage Preview), practice (i.e., Choral Reading), modeling (i.e., teacher modeling words that are read incorrectly), error correction (i.e., Word Drill), praise for correct reading, and contingent reinforcement (i.e., the chance to obtain tangible rewards for meeting a goal). The treatment package included all of these components, thus maximizing potential effects of reading group interventions. A description of the variables is provided below.

Taped Passage Preview (TPP). During weeks when the condition included TPP, the passage was pre-recorded on an audiocassette player by the experimenter. At the beginning of reading group instruction, the entire group and the experimenter first listened to the story on tape.

Choral reading (CR). Following TPP during group, if CR was a part of the specified package, the reading group read together the passage aloud with the experimenter. This provided additional opportunities to respond.

Praise (P). Praise (P) was delivered to reinforce accurate academic responses and academic engagement. Examples would include "good job", "great", etc. Praise was delivered on an intermittent schedule at a rate of about once every minute to minute and a half.

Error correction. Word Drill (WD) is a technique that requires the student to read the word three times following the correct modeling of the errors. WD was used with the entire reading group, regardless of who made the initial error each time an error was made. This procedure provided a model for correct responding and three additional opportunities to respond for all group members.

Contingent token reward. Contingent Token Reward (CTR) involves the presentation of preferred stimuli contingent upon oral reading fluency or accuracy. If, for example, CTR was a part of the treatment package identified for that specific week, the experimenter provided the student with the opportunity to obtain rewards on a daily basis during the reading group. The experimenter first asked the student if he or she wished to attempt to exceed the goal. If so, the experimenter told the

student that if he or she read the passage at a rate that matched or exceeded the previous number of CRW while matching or decreasing the number of errors, he or she could choose an item from the "goodie bag." Goals were developed on an individual basis in which each participant's best score on the specific passage being used served as the goal (i.e., the highest number of CRW and the lowest number of errors for the passage). Goals were revised daily throughout the study when CR was a part of the intervention package.

Experimental Design

A variation of the multiple-probe design across tasks (reading passages) was used to examine the efficacy of treatment across multiple passages (Wolery et al., 1988). Experimental control was demonstrated when the introduction of treatment increased responding in stimulus materials above and beyond any existing trends or patterns in the data during baseline conditions. Visible differences in level, trend, and/or variability were sought for individual passages in which treatment was carried out versus as-yet-untreated passages. Continuous measurement of already treated passages throughout the course of the study identified whether increased responding was maintained.

Procedures

Overview. Training graduate students in direct observation preceded student identification and data collection. Simultaneous with the training, a survey was administered to elementary school teachers to obtain an index of typical reading group instruction for elementary aged students. A teacher interview was conducted to

identify four participants for the study-- students with difficulty in reading fluency. Once consent and assent were obtained, data collection began and included daily observation sessions whereby treatment/interventions were manipulated by the experimenter (acting in a small group setting as the teacher) within an experimental design. Measurement of oral reading fluency was obtained the following morning after each session, including both instructional passages and generalization passages. Upon identification of the most effective, efficient reading intervention package, the experimenter trained a special education teacher to implement the reading group procedures and the teacher conducted the reading group for the final week of the study. Interobserver agreement and treatment integrity were assessed as well as treatment acceptability.

Training for observations. Graduate students were trained in direct observation using the format created for this study. Operational definitions were memorized and videotapes of classroom instruction were used while trainees practiced using the specified observation format (Refer to Appendix A). Moreover, trainees obtained further practice in a classroom environment during academic instruction. Interobserver agreement was computed for all practice sessions. Prior to participation in the study, observers were expected to demonstrate interobserver reliability of at least 80%. Reviewing and quizzing of operational definitions were utilized approximately every 10 days to ensure reliability throughout the study.

Survey of common instructional practices. A survey was distributed to approximately 15 teachers in three local elementary schools describing four steps of

typical reading instruction for small groups. It elicited responses and opinions as to its similarity with teachers' instruction styles. The survey began with a hypothesized first step of typical reading instruction and proceeded through three additional steps. A 3-point Likert scale was utilized to obtain this information (i.e., 1- Not Typical, 2- Somewhat Typical, 3- Typical). The information gathered from these surveys was used to establish a control condition, as typical instruction was constructed based on teachers' responses. Refer to Appendix B for a copy of the survey.

Teacher Interview. A semi-formal interview was conducted using the procedures from Witt et al. (2000) in order to identify low performance readers and obtain information regarding current levels of academic responding (i.e., reading fluency). Four children from the classroom were identified via the interview and teacher recommendation. Furthermore, information was collected regarding teaching instruction within the reading group, logistics of reading group (e.g., time that it meets), and curricular materials utilized within the students' general education classroom.

Direct Observation. Direct observations by trained observers were conducted daily during the analogue reading group. Four trained observers used the specified format (Appendix A) created for this project. One observer conducted the direct observation for each session, except for approximately every third session (approximately 33% for interobserver agreement), during which two observers conducted the observation. Observers were positioned off to the side of the reading group so that they could see both the teacher and the four students. Sessions were

between 6 minutes and 20 minutes in length depending on the condition in effect. Observers listened to an audiotape via two earpieces and independently recorded target behaviors for the teacher, target students, and peers using the definitions found in Table 3. The purpose of the observations was to identify absolute levels of student and teacher behaviors in the context of the reading group. Direct observation continued throughout the baseline and intervention phases of the study. Treatment integrity for the implementation of such procedures was evaluated during the observation by the observers, whereby observers followed a treatment protocol and recorded whether each step was properly implemented. Interobserver agreement and treatment integrity were computed for 34.4 % of the observation sessions.

Direct measurement of oral reading fluency. On a daily basis, students were individually assessed using the instructional CBM passages. Assessment occurred following every observation session and reflected performance in the story taught the previous day. The CBM probe was administered the following day after instruction in that passage. CRWs and errors were recorded as the student read the passage for 1 min. All data were graphed by the experimenter for visual inspection to identify any changes in level, trend, and/or variability. Direct measurement of reading fluency continued throughout the study.

Progress monitoring. Students were administered non-curricular generalization reading passages daily throughout the study. These data were continuously graphed by the experimenter for visual inspection to identify any changes in level, trend, and/or variability. The purpose of these data was to identify

whether students were generalizing the effects of treatment conditions to passages in which they were not receiving instruction.

Treatment Analysis. An experimental analysis was conducted whereby a variation of the multiple-probe design across tasks was used to evaluate the treatment package as a whole and to identify which component(s) of the proposed intervention package were effective in treating problematic academic responding (i.e., low reading fluency). Six passages were initially chosen for baseline and experimental conditions. Additional passages were added midway through the study so that additional package combinations could be examined. Prior to the study, passages were numbered, and randomly assigned to an order in which instruction was carried out. CRWs and errors were recorded and graphed for each passage. Data were collected for four days to establish baseline utilizing the typical instruction package obtained via the survey information. Following baseline, treatment conditions were carried out. Instruction with the complete treatment package (i.e., TPP, CR, P, WD, CTR) was carried out in the first passage of the treatment condition. Measurement of oral reading fluency was conducted in the treatment passage the following day, each day during that week, as well as in the remaining instructional passages one time that week. That is, all passages to be used at some point as instructional passages were measured for oral reading fluency even if they were not yet treated. An additional treatment phase in the second passage was conducted using the full treatment package to determine whether the treatment effects were consistent across passages.

The complete treatment package was effective at improving reading in instructional passages. Thus, modifications were made to the treatment package in an attempt to simplify it. Individual treatment components were withdrawn in clusters according to whether they were acquisition components, fluency components, or reinforcement components within the multiple-probe design across tasks. That is, the acquisition components were subtracted together leaving fluency and reinforcement components in the third week of treatment. The fluency and reinforcement components were subtracted in the same fashions during weeks four and five of treatment, respectively. All data from these conditions were continuously graphed throughout the experimental analysis phase. Visual inspection was used to identify the most salient treatment components (i.e., the most effective package with the fewest number of treatment components) to be implemented within the reading group during teacher instruction. If a subtracted component did not yield favorable results (i.e., fluency decreased), it was placed back in the package and another component was subtracted. This process continued until the most efficient and effective treatment package was identified.

In the second to last passage (week 7 of the study), the identified efficient, effective treatment package was instructed. During this week, the special education teacher was provided with the protocol to implement treatment. She also observed implementation for training for two sessions, while the experimenter conducted the reading group. The following week (i.e., the final week of the study), the teacher implemented the treatment package with the reading group as the experimenter

observed. The protocol was utilized throughout the reading group so that the teacher could rely on it for use during implementation.

Teacher training. Once the most effective package was identified for the group as a whole, the teacher was trained to implement the package for the following week. A training session consisted of explanation, modeling, practice, and feedback conducted between the teacher and the experimenter. A treatment protocol (Appendix C) was provided for the teacher for training. Furthermore, as noted earlier, the teacher had observed the experimenter implement the package during the reading group for two days prior to her implementation, while following along with the treatment protocol.

Interobserver Agreement

Curriculum-based measurement data. An independent observer listened to the audiotape recorded sessions and scored the passages for CRW and errors. To compute interobserver agreement, the total number of agreements for CRW and errors was divided by the total number of words in the passage, which represents all possible agreements plus disagreements. A total of 34.4 % of all sessions (i.e., 11 of the 32 sessions) was assessed for interobserver agreement. The mean agreement was 99.2% (range, 95.1% to 100%) across all participants.

Direct observation Data. To compute interobserver agreement, the total number of agreements for occurrences was divided by the total number of agreements plus disagreements. Agreement was assessed on an interval-by-interval basis. A total

of 34.4% of the direct observation sessions (11 of the 32 sessions) was assessed for interobserver agreement. The mean agreement was 91.5% (range, 79% to 100%).

Treatment Integrity

Experimenter integrity. Independent observers also assessed treatment integrity. Observers were provided with a protocol that outlined the maximum treatment package (i.e., TPP, CR, P, WD, and CTR). Refer to Appendix D for a copy of the baseline and treatment condition protocols. While observing the reading group, observers recorded whether each step was completed and in the specified order. The total number of steps completed was divided by the total number of steps in the condition to yield the percentage of steps completed for each session. Treatment integrity was assessed for 34.4% of all sessions (i.e., 11 of the 32 sessions). The mean percentage of correctly implemented steps was 99.3% (range, 88% to 100%).

Teacher integrity. Treatment integrity was assessed during implementation by the teacher. Seventy-five percent of the week's sessions (i.e., 3 of 4) were observed by two independent observers. The mean percentage of correctly implemented steps was 100%.

Treatment Acceptability

The teacher was given a questionnaire (i.e., IRP-15) evaluating treatment acceptability. She was given the opportunity to fill out the questionnaire and offer any feedback deemed necessary. The IRP-15 has been statistically analyzed and is a valid survey for intervention acceptability ratings (Martens et al., 1985). Refer to Appendix

E for a copy of the IRP-15 form that was used. A 6-point Likert Scale was utilized with 1 equaling "strongly disagree" and 6 equaling "strongly agree".

Results

Oral Reading Fluency

Instructional passages. CRW/min for all participants (i.e., Blake, Cody, Devon, and Karla) are displayed in Figures 1, 3, 5, and 7, respectively. In general, all participants had increases in CRW/min from baseline across all conditions. Effect sizes obtained are large and range from 1.37 to 6.99 across all treatments for all participants. ER/min data for all participants (i.e., Blake, Cody, Devon, and Karla) are displayed in Figures 2, 4, 6, and 8, respectively. In general, all participants had decreases in ER/min from baseline across all conditions. Results further suggest evidence of maintenance, as CRW/min are maintained throughout the study when treatment was withdrawn from each specified passage and another treatment was implemented in successive phases of the multiple-probe design across tasks. Means for CRW/min, errors/min, and effect sizes per condition can be found in Table 4.

Blake obtained higher increases of CRW/min from baseline rates in the full package (i.e., 65 and 63.90) followed by the effective package [Taped Passage Preview, Choral Reading, Word Drill, and Praise (i.e., 59.75 and 44.90)]. Effect sizes for the effective package are substantial at 5.33 and 5.2 (i.e., experimenter implemented and teacher implemented, respectively). Error rates were cut in half across all treatment conditions, with decreases in errors for the effective package at 1.75 and 3.50, respectively. Thus, along with increases in fluency, accuracy also

improved substantially. Blake's fluency rates increased by 54.85 CRW/min in the TPP, WD, P, CTR (Contingent Token Reward) package, wherein a tangible reward was offered for obtaining a specified goal. However, indications of fluency rates for the effective package, which did not contain a tangible reinforcer component, appear that reward is not needed to sustain increases. Furthermore, the presence of the tangible reward in CR, P, CTR, which lacks the acquisition and fluency-building pieces (i.e., TPP and WD), indicates that tangible reward is more effective for Blake only when he has had enough opportunity to effectively learn the passage and become fluent. Maintenance of effects is observed, as CRW/min are maintained at a level either commensurate to treatment or slightly higher. ER/min are slightly variable but are at a level below that of pre-treatment. Fluency and error rates, as well as effect sizes (Table 4) are consistent with the visual display in Figures 1 and 2.

Table 4

Means for CRW/min, ER/min, and Effect Sizes on Oral Reading Fluency

	Blake		Cody		Devon		Karla	
Condition	**CRW**	**Errors**	**CRW**	**Errors**	**CRW**	**Errors**	**CRW**	**Errors**
Typical								
Baseline	111.00	3.25	93.25	4.50	103.25	6.50	101.50	8.50
Full Package								
Baseline	108.00	1.75	91.00	4.25	98.75	8.75	98.25	6.50
Treatment	173.00	1.50	112.25	1.50	161.75	4.25	170.00	3.50
Discrepancy	65.00	-0.25	66.00	-2.75	63.00	-4.50	71.75	-3.00
Effect Size	**3.19**		**2.21**		**4.08**		**2.95**	

Table 4—Continued

Means for CRW/min, ER/min, and Effect Sizes on Oral Reading Fluency

Condition	Blake CRW	Blake Errors	Cody CRW	Cody Errors	Devon CRW	Devon Errors	Karla CRW	Karla Errors
Full Package								
Baseline	93.60	4.80	77.00	4.25	89.40	7.40	74.20	10.60
Treatment	157.50	1.25	111.25	0.25	138.00	1.50	150.00	7.00
Discrepancy	63.90	-3.55	34.25	-4.00	48.60	-5.90	75.80	-3.60
Effect Size	**3.62**		**2.35**		**3.38**		**3.56**	
CR/P/CTR								
Baseline	114.33	5.00	80.80	4.00	74.00	6.80	85.40	7.80
Treatment	162.75	2.25	105.75	1.50	122.25	6.00	123.75	5.50
Discrepancy	48.42	-2.75	24.95	-2.50	48.25	-0.80	38.35	-2.30
Effect Size	**2.67**		**2.22**		**6.99**		**6.69**	
TPP/WD/P/CTR								
Baseline	109.90	5.00	89.40	4.00	101.30	5.10	91.70	13.10
Treatment	164.75	1.25	141.25	0.75	156.75	2.25	142.00	5.25
Discrepancy	54.85	-3.75	51.85	-3.25	55.45	-2.85	50.30	-7.85
Effect Size	**3.34**		**4.66**		**5.08**		**2.82**	
TPP/WD/CR/P								
Baseline	118.75	2.10	100.00	3.25	101.80	4.80	104.40	8.10
Treatment	164.75	1.25	128.00	2.00	146.50	1.75	144.50	4.00
Discrepancy	46.00	-0.85	28.00	-1.25	44.70	-3.05	40.10	-4.10
Effect Size	**2.51**		**3.69**		**3.44**		**3.68**	
Effective: EX								
Baseline	120.50	4.00	91.00	1.00	90.50	1.20	104.60	4.40
Treatment	180.25	2.25	113.00	0.33	159.75	2.00	197.75	5.00
Discrepancy	59.75	-1.75	22.00	-0.67	69.25	0.80	93.15	0.60
Effect Size	**5.33**		**1.37**		**4.68**		**4.10**	
Effective: TCH								
Baseline	89.90	4.75	84.00	1.40	80.30	5.30	81.30	8.40
Treatment	134.75	1.25	127.50	0.25	132.00	2.50	126.80	5.00
Discrepancy	44.85	-3.50	43.50	-1.15	51.70	2.80	45.50	-3.40
Effect Size	**5.20**		**2.42**		**2.43**		**4.28**	

Figure 1

Measure of Oral Reading Fluency for Blake: CRW/min

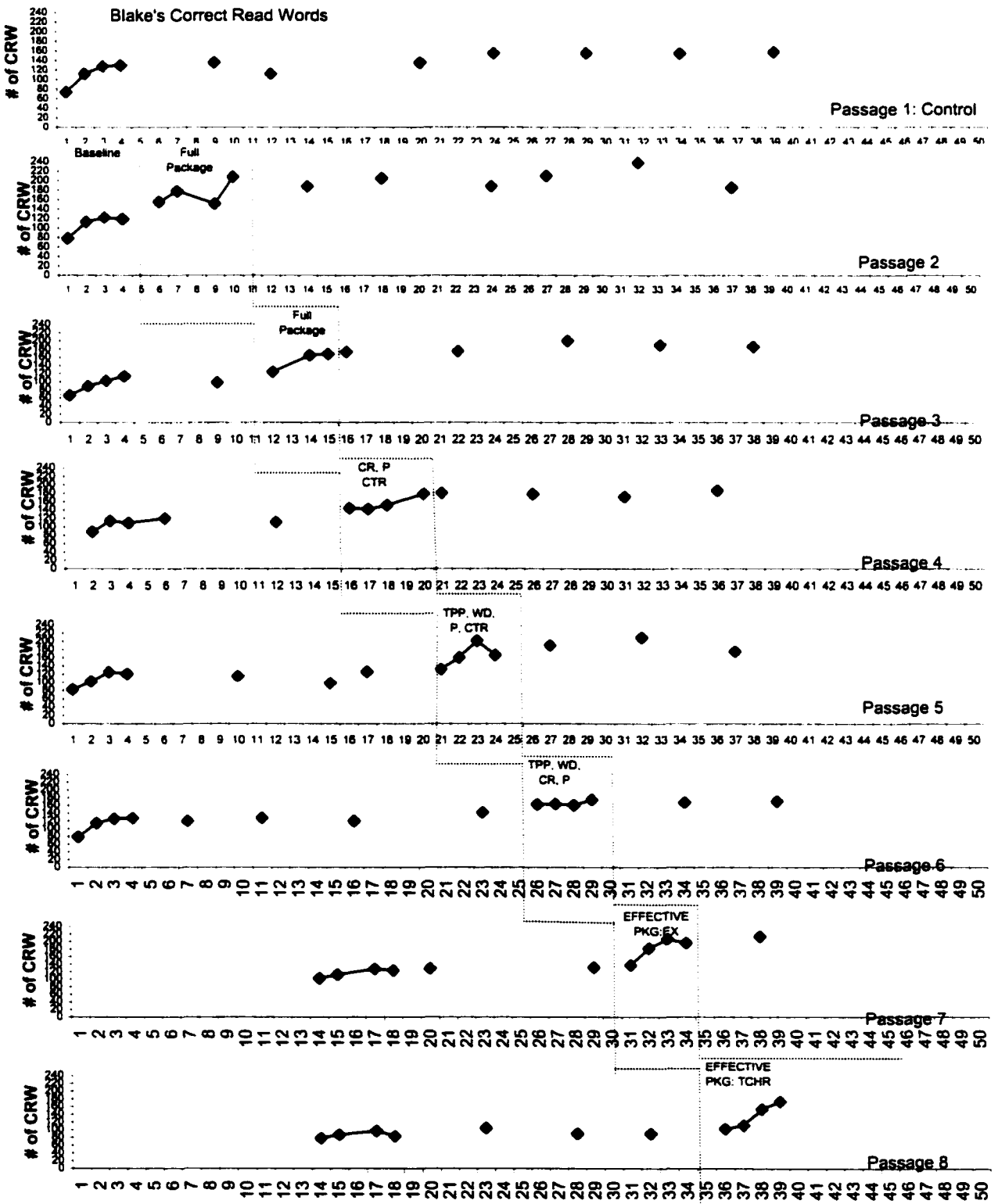
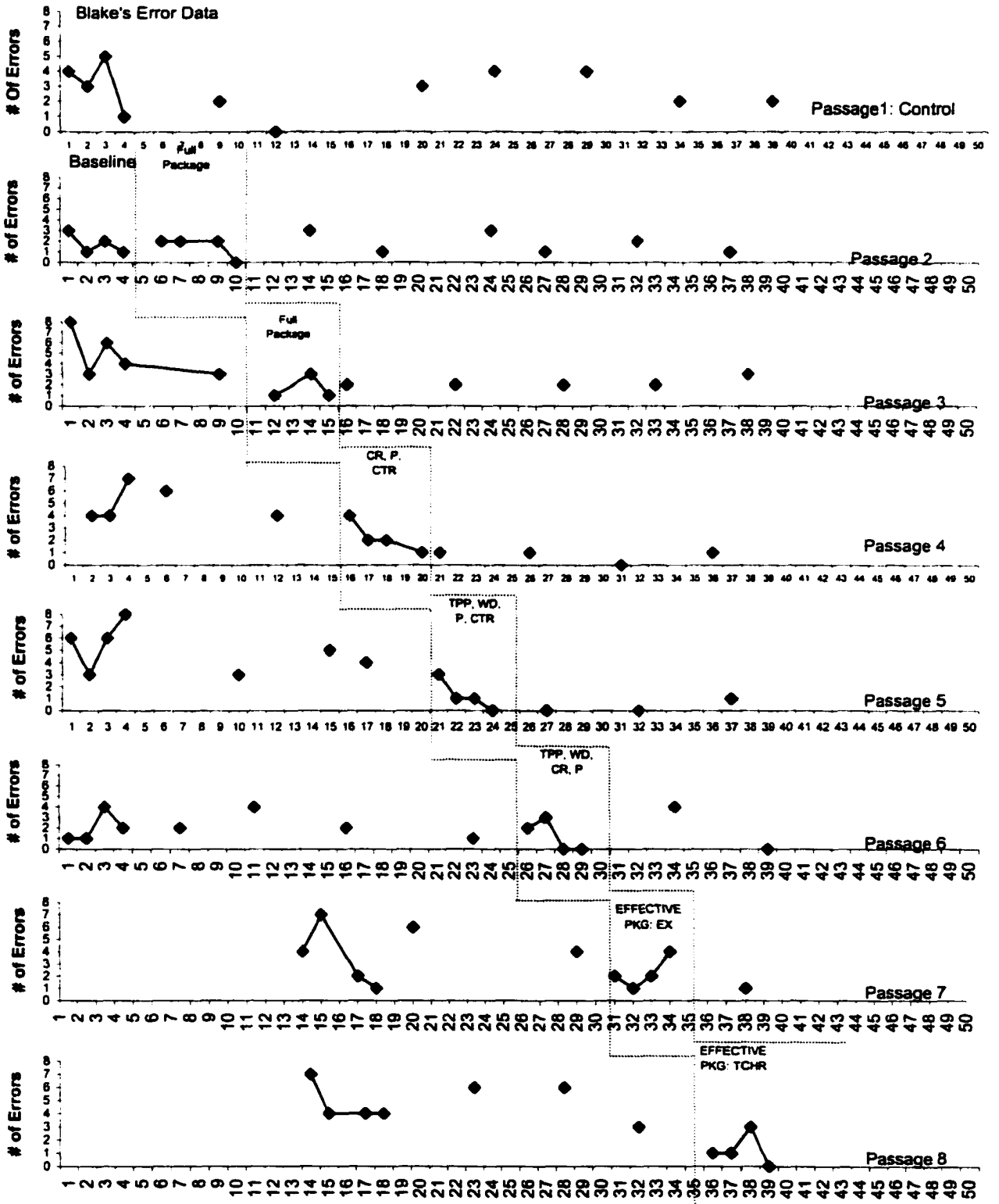


Figure 2

ER/min for Blake



Cody obtained higher increases of CRW/min in the TPP, WD, P, CTR package (i.e., 51.85), than in any other intervention package including the full package. The effective package (TPP, CR, WD, and P) produced fluency rate increases of 22.00 and 43.50 CRW/min, respectively. Effect sizes for this package are 1.37 and 2.42 (i.e., experimenter implemented and teacher implemented, respectively) and are considered "large" (Cohen, 1965). Effect sizes, however, are the lowest across all four participants. Cody's fluency rates did not vary significantly across other treatment packages, as increases in CRW/min were within 4 words of each other. It appears that his oral fluency rates were affected by his desire to earn a tangible reward and his competitive nature, as every day he would ask whether he could try to "beat his score". Cody's fluency rates were lower, on average, than the other participants; however, his accuracy was amongst the best. Cody had the lowest error rates during treatment conditions, with a mean of .94 ER/min (range .25 to 2.00) across all treatment conditions. Like Blake, Cody's error rates were cut in at least half for all intervention packages. The biggest decrease in errors (3.25 words) was in the TPP, WD, P, and CTR condition, again attesting to his desire to earn the reward. Maintenance is evident by the level of CRW/min and errors/min following treatment change. Cody's errors maintain at levels near zero across most instructional passages. Fluency rates, error rates, and effect sizes (Table 4) are consistent with the visual display in Figures 3 and 4.

Figure 3

Measure of Oral Reading Fluency for Cody: CRW/min

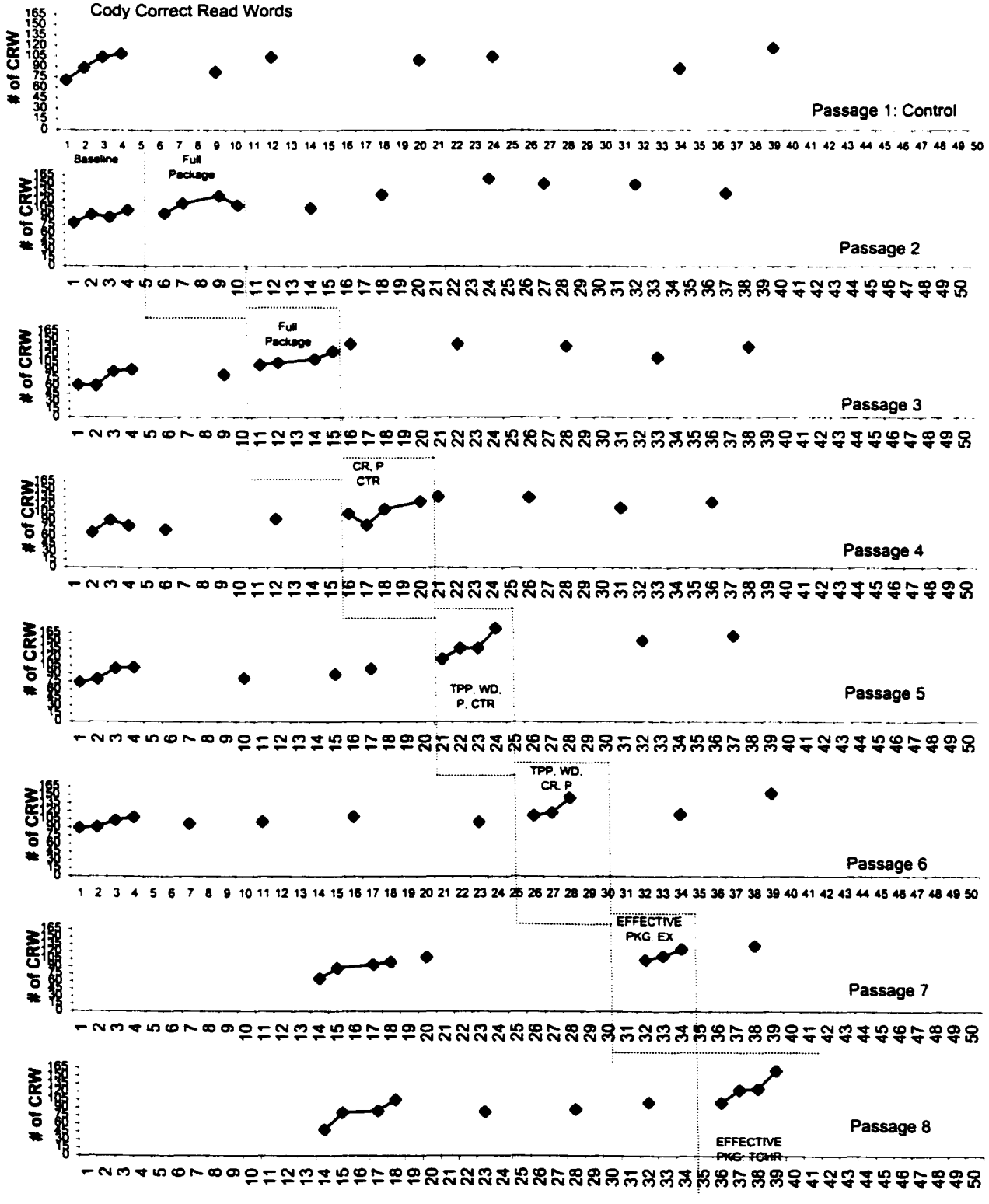
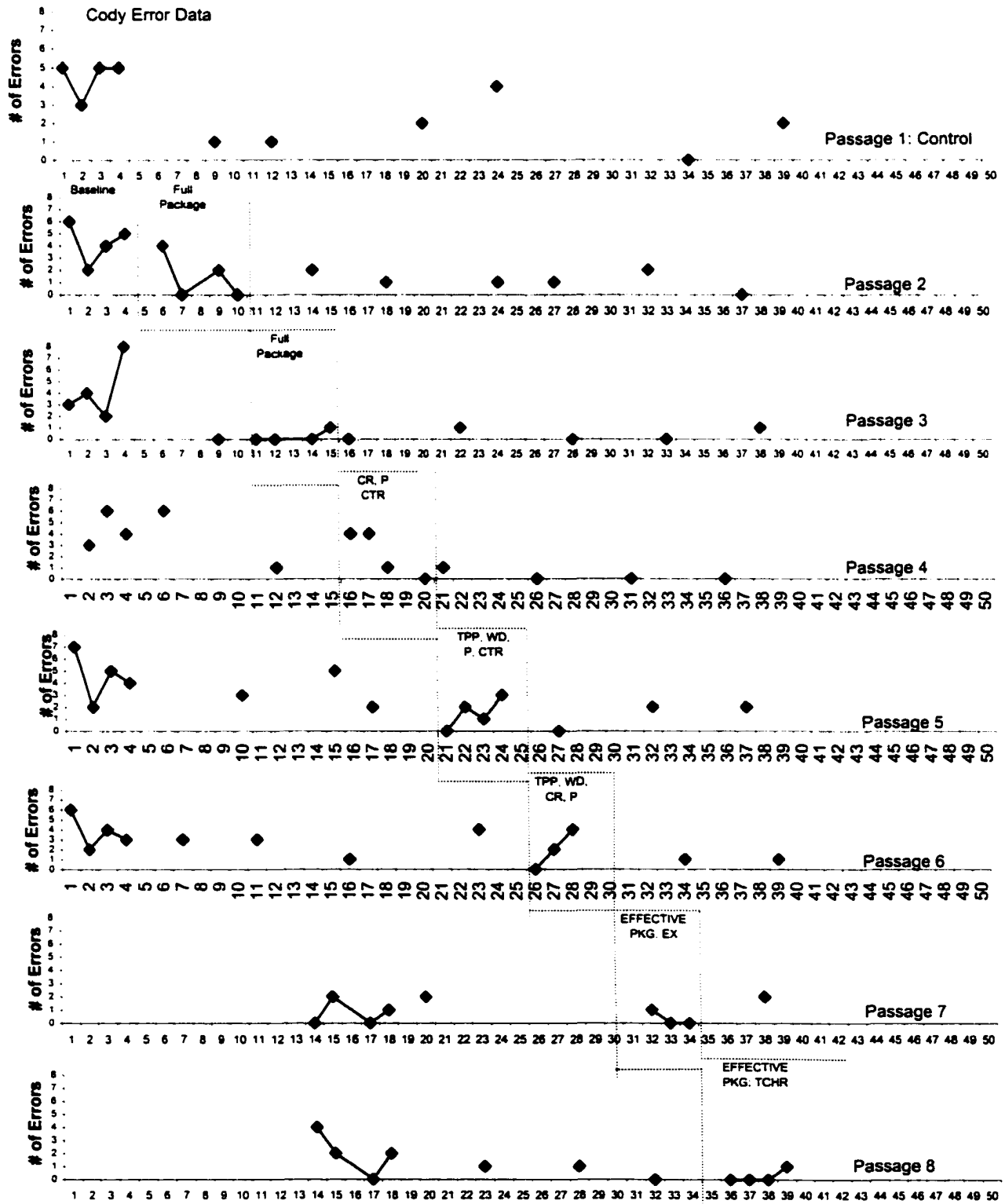


Figure 4

ER/min for Cody



Devon obtained higher increases of CRW/min in the effective package (i.e., 69.25 and 51.7) followed by the full package (i.e., 63 and 48.6). His fluency rate increases were substantial across all treatment packages, as the lowest increase was 48.25 CRW/min (CR, P, CTR). Decreases in ER/min were substantial across all treatment conditions (range 2.80 to 5.90) excluding one effective package in which his errors increased by .80 (the mean error count at baseline started at 1.20) and CR, P, CTR, wherein decreases in errors were only .80. Because this latter package was his weakest performance (i.e., the lowest increase in CRW/min and lowest decrease in ER/min) and moreover, did not contain acquisition and fluency-building components, it appears that Devon benefits from repeated exposure to the passage, and furthermore, a tangible reward does not appear to be warranted (as it was not a part of the effective package and this was his strongest display of oral reading fluency). Evidence of maintenance can be seen by the steady state of CRW/min, which is at a level commensurate to that of CRW/min during treatment. Furthermore, although there is a slight increase in the level of errors/min following phase changes, errors do not increase to a pre-treatment level. These results are consistent with the other participants. Again, effect sizes were large ranging from 2.43 to 6.99 across all treatment conditions. Fluency and error rates, as well as effect sizes (Table 4) are consistent with the visual display in Figures 5 and 6.

Figure 5

Measure of Oral Reading Fluency for Devon: CRW/min

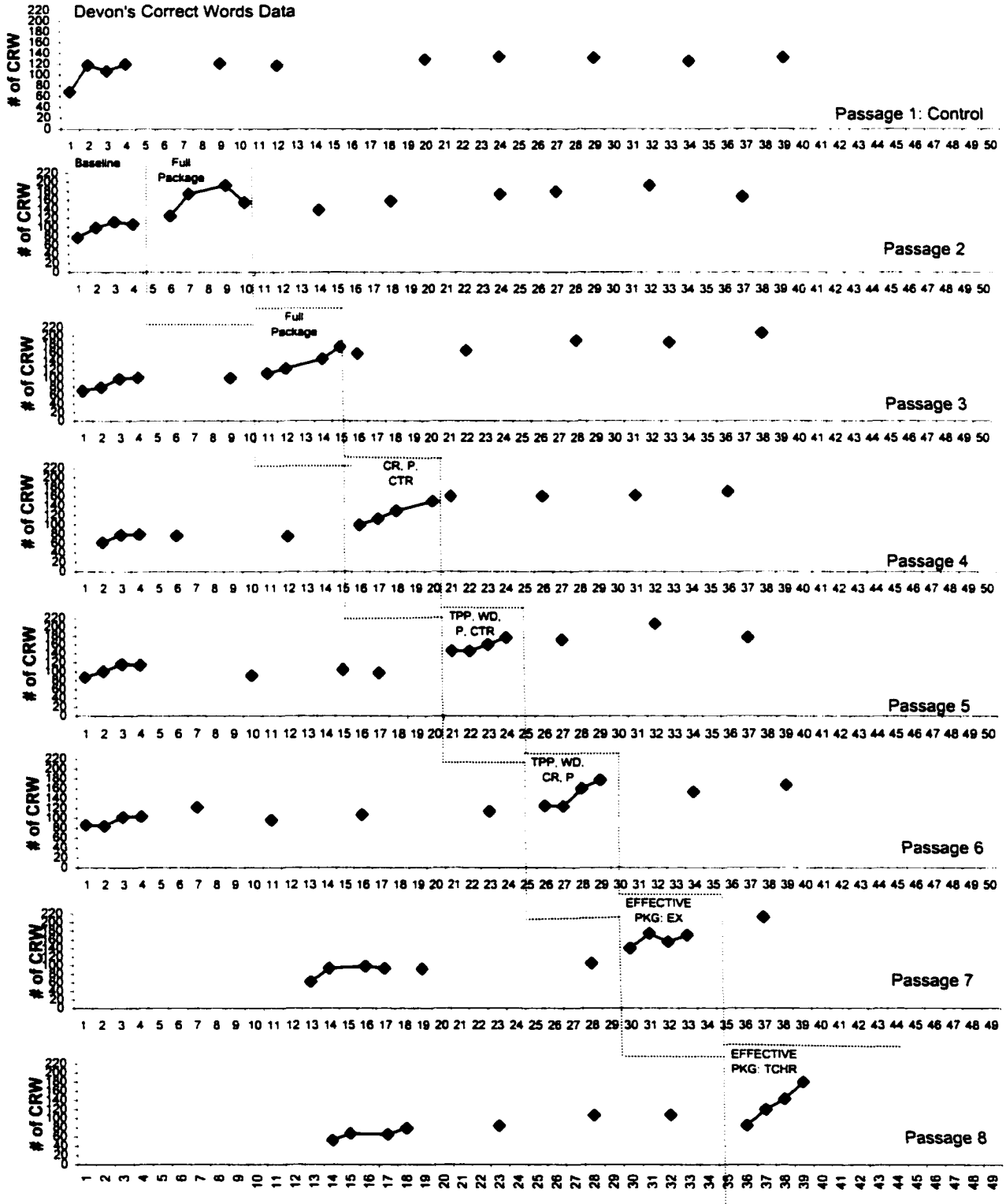
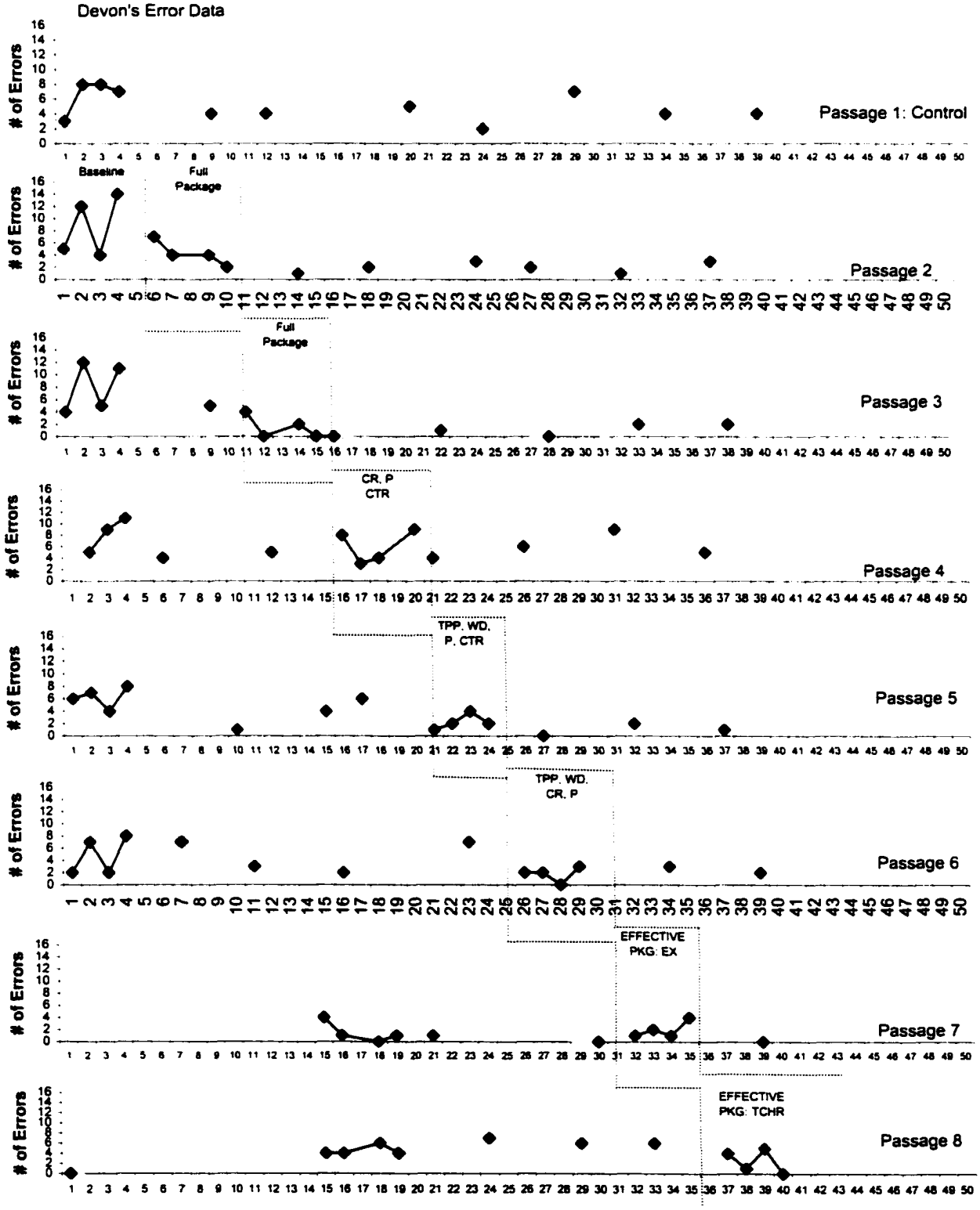


Figure 6

ER/min for Devon



Karla obtained higher increases of CRW/min in one implementation of the effective package (i.e., 93.15) than in any other of the intervention packages. The next highest was the full package (i.e., 71.75 and 75.8). Much like Blake and Devon, Karla had large increases in CRW/min across all treatment conditions (mean = 59.28; range 38.35 to 93.15). Her oral reading fluency rate was nearly double a mastery level (Shapiro, 1996a) in the effective package with CRW/min at 197.75 in grade level materials. Moreover, maintenance is evident in that Karla's CRW/min tend to level off, and at times, increase to post-treatment levels. Unlike the other students, Karla had a large numbers of errors during baseline (range = 4.40 to 13.10), as well as during treatment (range = 3.50 to 7.00) across all conditions, although decreases occurred (mean decrease = 3.38). Therefore, her number of CRW/min is relative to accuracy as well. Much like Blake and Devon, it appears that Karla benefits from repeated exposure and opportunities to respond. The treatment condition lacking such qualities (CR, P, CTR) was her weakest performance resulting in increases of CRW/min of 38.35 and decreases in errors of 2.30. Following withdrawal of treatment, Karla's error rates tend to increase slightly and then level off prior to reaching pre-treatment numbers. Effect sizes are again substantial, ranging from 2.82 to 6.69 across all treatment conditions. Fluency and error rates, as well as effect sizes (Table 4) are consistent with the visual display in Figures 7 and 8.

Figure 7

Measure of Oral Reading Fluency for Karla: CRW/min

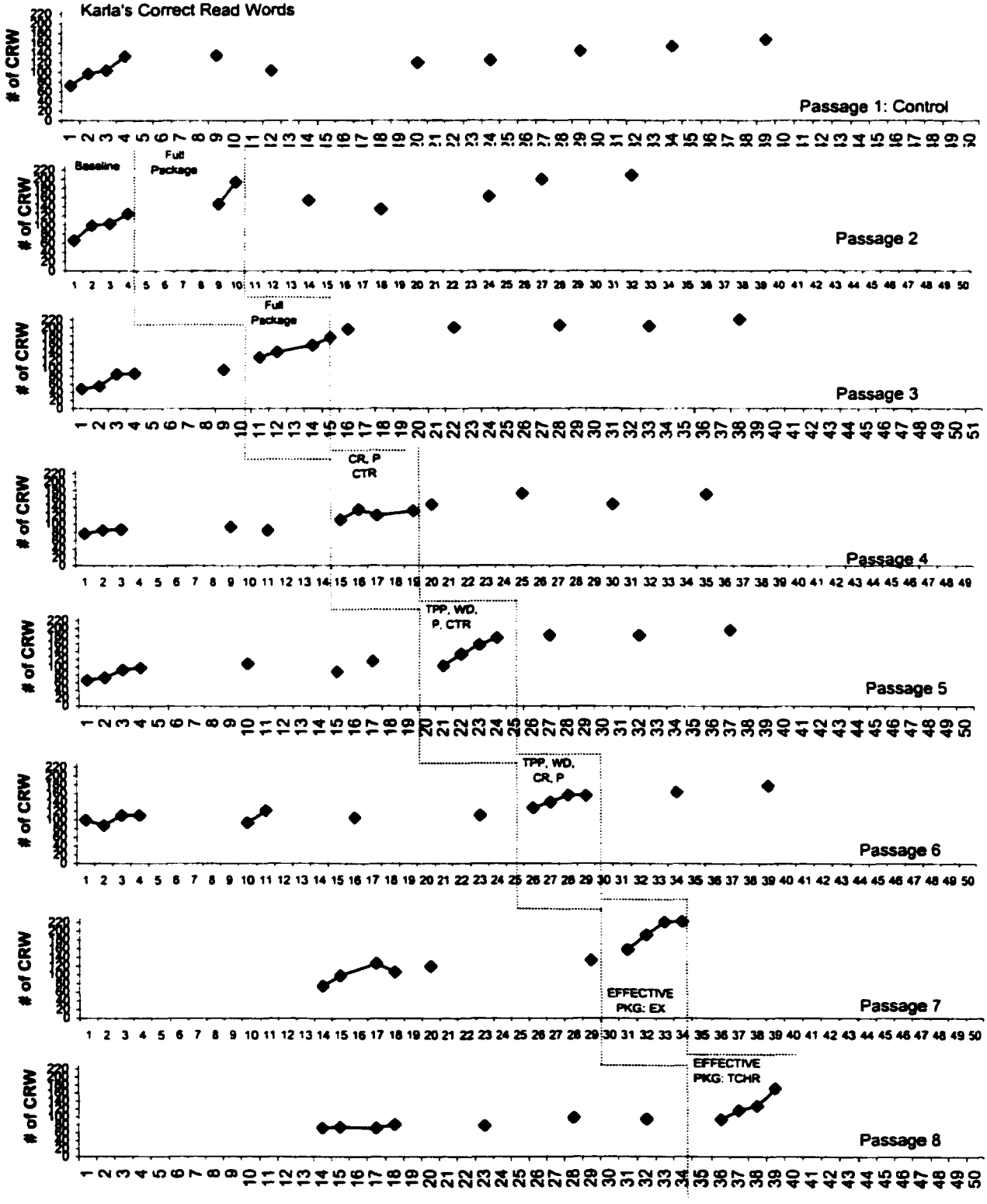
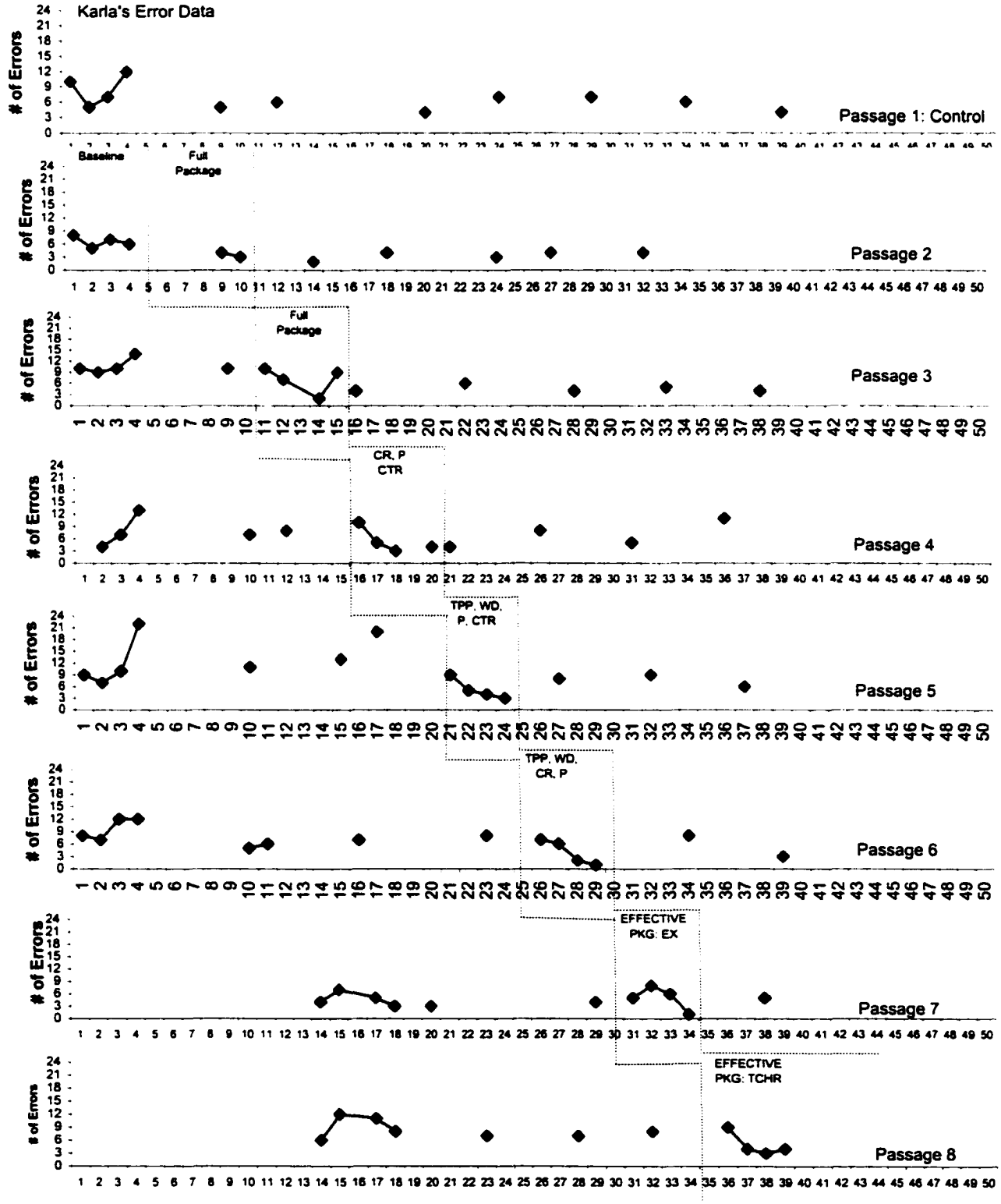


Figure 8

ER/min for Karla



Overall, all participants had increases in CRW/min and decreases in errors/min across all intervention packages. For three of the four participants (Blake, Devon, and Karla), the effective package (TPP, CR, WD, and P) produced the highest increases in CRW/min and the largest decreases in errors/min. Effect sizes for the effective package were amongst the most substantial for Blake, Devon, and Karla as well. For the fourth participant (Cody), the effective package was the second most effective in maintaining the highest increases in CRW/min. An earlier package (i.e., TPP, WD, P, and CTR) was the most effective for Cody. Maintenance of CRW/min is evident for all participants following the withdrawal of intervention packages, as levels of CRW/min are commensurate or higher than treatment levels and errors/min are lower than during baseline and commensurate to treatment levels.

Generalization passages. Generalization data for all participants (i.e., Blake, Cody, Devon, and Karla) are displayed in Figures 9 through 12, respectively. In general, all students had increases in CRW/min from baseline (i.e., typical instruction) across all treatment conditions. An increasing trend for CRW/min is displayed for all participants. (Figures 9 through 12). However, the trend for CRW/min for Cody (Figure 10) is less steep and appears to be more level, although visible. Despite the increasing trend, data are highly variable for all participants across all treatment conditions. In general, errors/min are substantially less variable across treatment conditions for three of the four participants (i.e., Blake, Cody, and Devon). As with the instructional passages, the number of CRW/min is higher for Blake (mean = 100.3), Devon (mean = 79.9), and Karla (mean = 77.4), than for Cody

(mean = 72.8), although to a lesser degree. Additionally, Cody's errors/min (mean = 2.8) across all passages are lower than the other participants. Thus, Cody appears to read at a slower, but more accurate rate than the others across both the instructional and generalization passages. As with the instructional passages, Karla has the highest number of errors/min across all passages (mean = 9.4). Based on the variability of the data, the increasing trend, and the short length of time each treatment condition was implemented, it is difficult to draw any conclusions relative to which treatment produced the least generalization (if any).

Figure 9

Generalization of Oral Reading Fluency for Blake

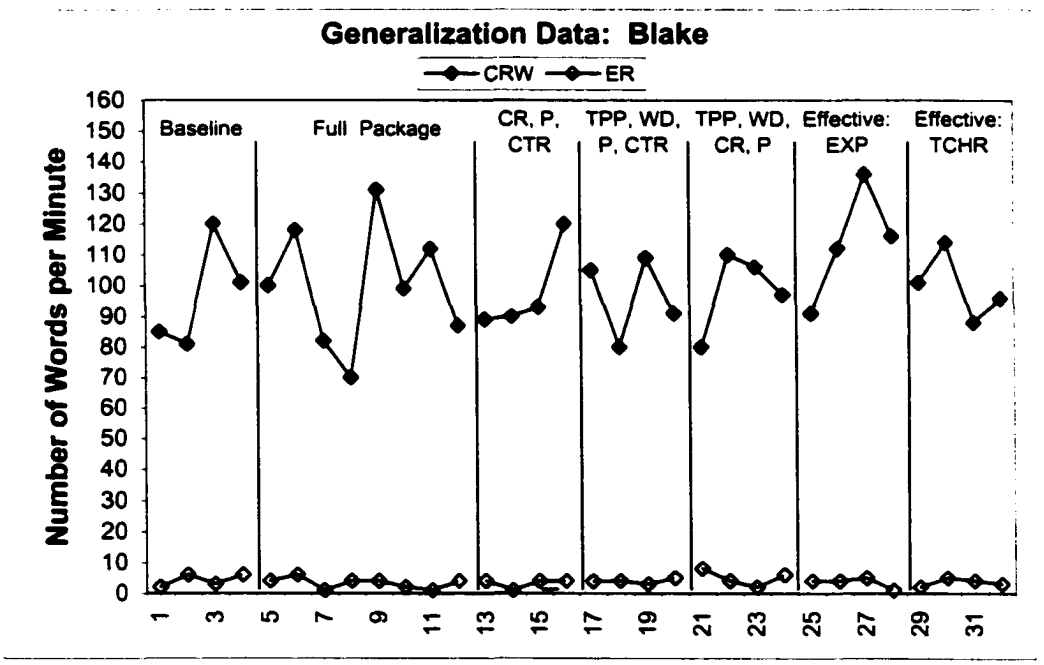


Figure 10

Generalization of Oral Reading Fluency for Cody

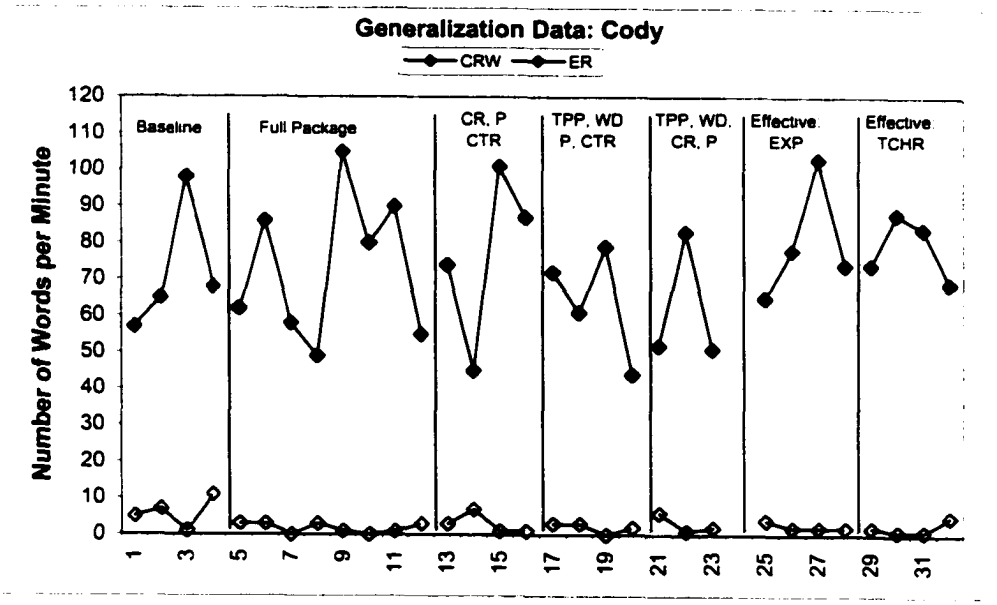


Figure 11

Generalization of Oral Reading Fluency for Devon

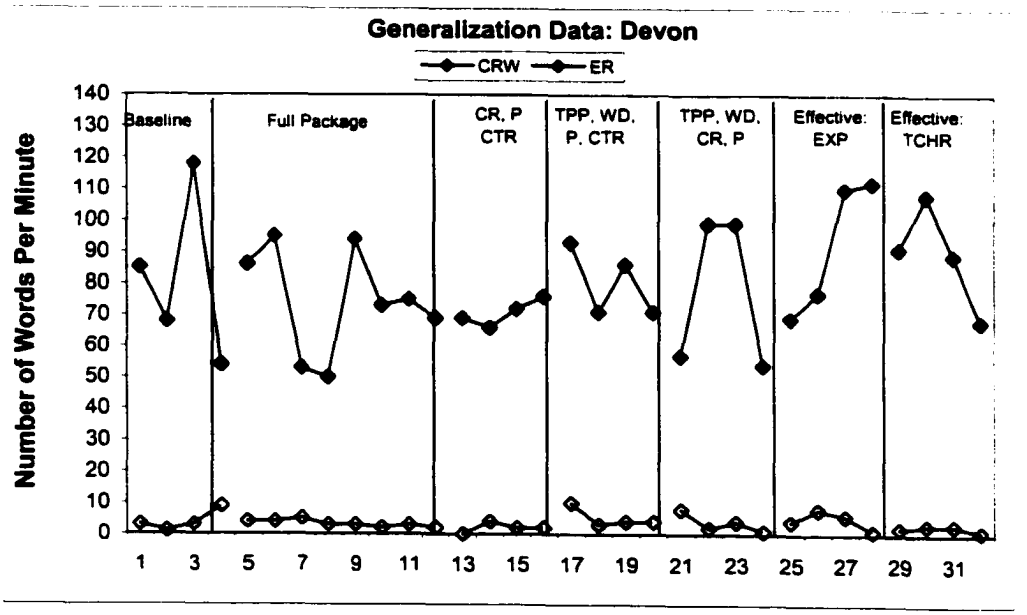
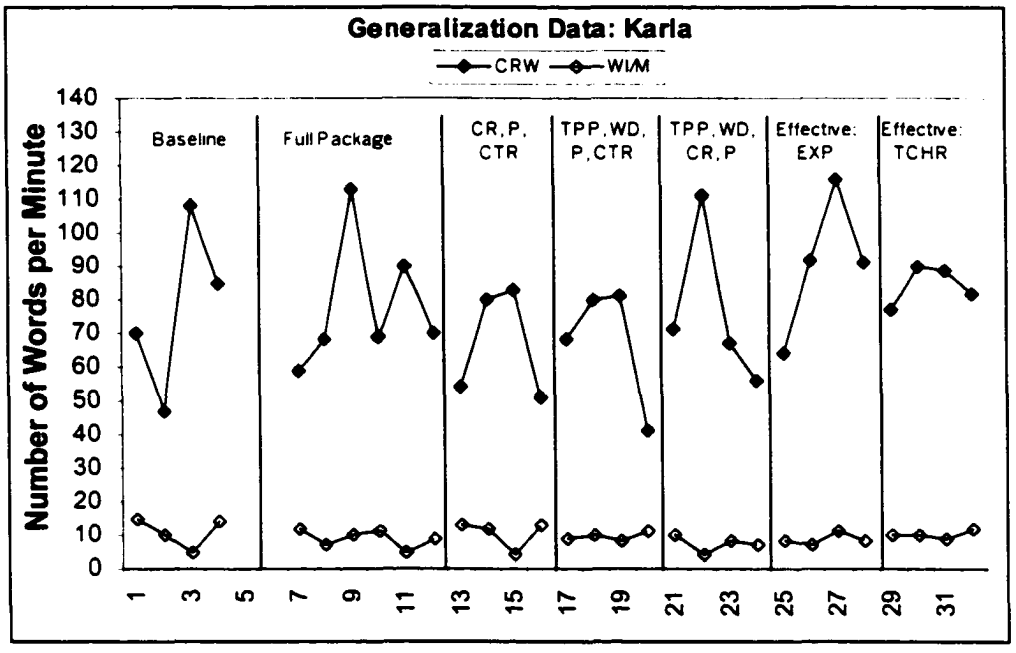


Figure 12

Generalization of Oral Reading for Karla



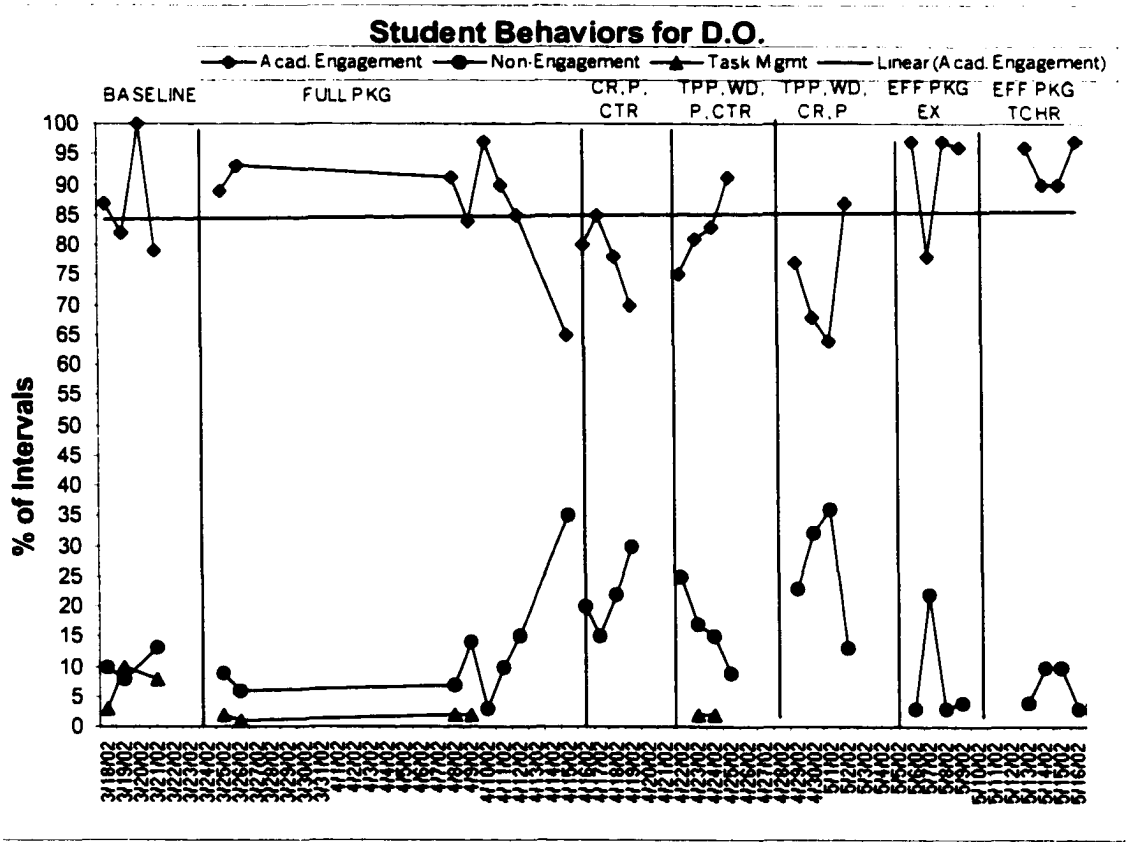
Direct Observation Data

Student engagement. Direct observation data of student behavior for all participants are displayed in Figure 13. Percentage of time of students' academic engagement during small group reading instruction ranges from 64% to 100% across all conditions. These data indicate that there is a slight increasing trend in the percentage of time spent academically engaged from baseline throughout the treatment phases. As is evident through visual inspection, there is a significant decline in the percentages of engagement during the full treatment package. From this phase, academic engagement begins to increase throughout treatment changes and upon completion of the study and the implementation of the effective package, the average time spent in academic engagement increased by 6.25% [from 87% during baseline to

93.25% during the effective package (TPP, CR, WD, and P)]. That is, students spent more time on average academically on-task/engaged during the intervention package that was most effective (i.e., obtained the highest increases of CRW/min) than during any other intervention package.

Figure 13

Measure of Student Behaviors During Direct Observation of Small Group Reading Instruction



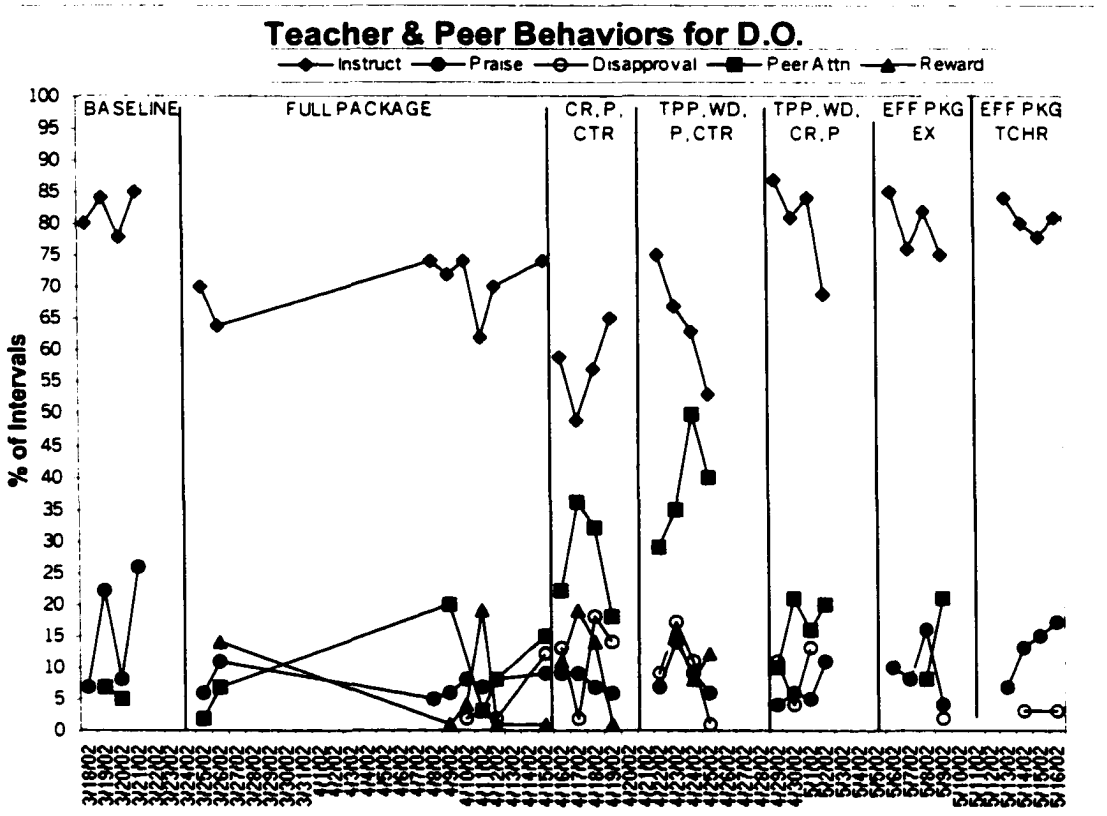
Teacher behaviors. Direct observation of teacher behaviors are displayed in Figure 14. These data indicate that the average time spent instructing students

decreased slightly from baseline (i.e., typical instruction) to the effective package from 81.75% to 80.13%, respectively. Typical instruction consisted of a story preview, students taking turns reading the story aloud to one another, students reading the story quietly to himself/herself, and students answering comprehension question about the story's content. During instruction, the teacher spent on average nearly 82% of the small group instructing students. Throughout the effective package condition (TPP, WD, CR, and P), the teacher spent nearly 2% less time instructing. Percentage of time spent in instruction is substantially lower for other treatment packages [e.g., CR, P, and CTR (57.5%) and TPP, WD, P, CTR (64.5%)]. However, levels of praise decrease from 15.75% to 6.5% and levels of disapproval increase (i.e., redirecting students to be on-task and academically engaged) from 0% to 9.3% during typical instruction and TPP, WD, P, and CTR (i.e., the package prior to the effective package), respectively. Percentage of time praising students and redirecting students during the effective package was 11.25% and 1%, respectively.

Overall, the total percentage of teacher time spent in all interactions with students during small group reading instruction decreased from 97.5% (typical instruction) to 92.38% (effective treatment). Thus, the most effective treatment package for increasing oral reading fluency and improving accuracy (i.e., TPP, WD, CR, and P) was less time consuming than everyday typical instruction.

Figure 14

Measure of Teacher and Peer Behaviors During Direct Observation of Small Group Reading Instruction



Other student behaviors. Peer attention is displayed in Figure 14. Levels of peer attention are present throughout all treatment phases and range from 0% to 38.5%. Percentage of peer attention was nearly equal during typical instruction and the effective treatment package as results indicate 3% and 3.6%, respectively. Thus, time spent with others providing attention to the target student, both positively with social praise (e.g., "Good job, Blake."), or negatively (e.g., "Come on Karla, read

right.") is about equal for typical instruction versus the effective intervention package. Peer attention was highest during treatment packages including CTR (contingent token reward), as students were more verbally expressive to each other about goals, obtaining rewards, etc.

IRP-15

The mean rating across all items was 4.86. The teacher did express concern about how a teacher would implement this package with an entire class. Thus, a lower score on "Would you be willing to implement this in the classroom setting" was obtained (i.e., a 4). It was explained that the procedures were designed for small groups and she then agreed that this intervention could and would be productive if implemented with a small reading group. The teacher further added verbally that the intervention procedures were valuable and that she implemented them into the classroom with two small reading groups. Moreover, she is presenting the treatment package and her success with increasing oral reading fluency rates at the Michigan Council for Exceptional Children in February of 2003. Therefore, the teacher fully accepted and was pleased with the intervention procedures and their outcomes.

Summary of Results

Overall, results obtained included increases in CRW/min and decreases in ER/min across all students within every treatment package. Furthermore, effect sizes were large across all conditions and participants as well. The effective package produced the greatest results in oral reading fluency for three of four participants (i.e., Blake, Devon, and Karla) and produced the 2nd greatest results for Cody.

Maintenance of effects occurred across all treatments, as CRW/min are commensurate or higher than treatment levels and errors/min are lower than during baseline and/or commensurate to treatment levels. Within generalization passages, all students had increasing trends in CRW/min from baseline through treatment phases despite the variability in the data. Percentage of student engagement during small group reading instruction increased from baseline to the effective package by 6.25%. Total percentage of teacher time spent in interaction with students (i.e., instruction, praise, and disapproval) decreased by 5.12% from baseline to the effective package. Therefore, the effective package produced higher rates of oral reading fluency and higher average rates of academic engagement. The treatment, however, was more efficient for the teacher, as less time was spent in small group instruction. Moreover, it was rated acceptable and was implemented within the classroom following the study.

Discussion

This study sought to identify the variables that serve to improve academic responding, thus increasing oral reading fluency rates by identifying the most effective, efficient reading intervention package for four 4th grade students who were all participating in a single, small reading group. A multiple-probe design across tasks was utilized in examining the effects of acquisition, fluency, and consequential components (e.g., praise and contingent token reward) on oral reading fluency and generalization, active engagement, and teacher and peer behavior. The small group reading instruction was designed so that it closely resembled key components of the

natural environment (i.e., the classroom), which were identified via functional assessment.

It was hypothesized that at least one effective treatment would be identified based on the rigorous analysis of the variables. It was also hypothesized that the treatment package selected as the most effective and efficient would be so for most of the participants, but that it would also produce substantial increases in oral reading fluency for all students. Furthermore, the identification of the effective, efficient treatment package would provide higher acceptability ratings; thus, increasing treatment integrity and implementation.

The treatment packages were analyzed and the "effective package" was identified utilizing a dismantling procedure (Barnett et al., 2003) in an attempt to create an equally effective package that was more efficient (i.e., easier to do). Through an experimental analysis utilizing reading intervention components that (a) have been empirically validated [e.g., TPP, Rose & Beattie, 1986; CR, Eckert et al., 2002; WD, Rosenburg, 1986; P, Cossairt, 1973; and CTR, Billingsley, 1977)], (b) correspond to the conceptual framework of academic responding within a learning hierarchy [i.e., the Instructional Hierarchy (Haring et al., 1978)] and (c) address the issues of skill-based versus performance-based deficits presented by Lentz (1988), an effective package was identified. The results of this study indicate that all treatments were effective at increasing responding for all four participants, with the identified "effective package" being most successful in increasing responding for three of the four.

In most cases, students nearly doubled their reading fluency rates by the end of the study. Immediate effects were observed in most instances and there was clear evidence that students maintained these effects once treatment was withdrawn. That is, CRW/min did not decrease substantially and/or ER/min did not increase substantially with the withdrawal of treatment. Active student engagement, which can be measured through active student responding [in this study, direct observation data, (Daly & Murdoch, 2000) increased from typical instruction to the effective package by 6.25%. This finding is consistent with increases in CRW/min, as both are overlapping response classes. That is, if a student is reading aloud, (s)he is actively engaged. Total percentage of teacher time spent in interaction with students during small group instruction decreased by 5.12% with the effective package. Therefore, performance and academic engagement increased, while teacher effort decreased; thus, identifying an effective, yet more efficient reading intervention package (the primary goal of this study).

Based on the results, it is reasonable to assume that treatment acceptability would be favorable. Although a formal survey of acceptability was administered (i.e., the IRP-15) following the study, the teacher had the luxury of implementing the procedures herself prior to evaluating them. Therefore, verbal consultation occurred throughout the study with the teacher, teacher training was delivered, and the teacher experienced the intervention package firsthand. The acceptability ratings were favorable, as she was pleased with student outcomes and with the efficiency of the

small group reading instruction. Since that time, she has implemented the procedures within her classroom for two small reading groups.

These data are encouraging, as they speak to the efficacy of the procedures and the effectiveness of the treatment package selected in promoting maintenance following instruction. They lend further empirical support to the robustness of the antecedent and consequent strategies employed, and furthermore, increase our understanding of the likelihood of achieving maintenance when instruction occurs repeatedly, as is typically the case during small group reading instruction. This finding is similar to the results of another study conducted by Bonfiglio, Daly, Martens, Lin, and Corsaut (in press).

Although the "effective package" (i.e., TPP, CR, WD, and P) was only the most effective treatment for three of four participants (Blake, Devon, and Karla), it was selected as the most effective, efficient package of the study. Cody obtained slightly higher increases in CRW/min in another treatment condition; however, one goal of the study was to determine the most effective, yet efficient package for the small reading group as a whole. Therefore, the combination of TPP, CR, WD, and P was identified and implemented as the most effective package for the final phase of the study (i.e., the phase wherein the teacher implemented the package during the reading group instruction). The decision appeared to be the correct decision, as Cody's CRW's/min increased by 43.5 words. Thus, when the teacher implemented this package, it became the most effective for Cody as well. Based on these data, it

appears reasonable to assume that the correct "effective package" was selected and implemented by the classroom teacher.

It might be that different results would have been obtained with students who were reading at different fluency levels (e.g., frustrational level at the fourth grade or instructional at a lower grade level, first or second). The advantage of the procedures employed in this study, is that although the treatment package selected for these students in this small reading group was effective, other combinations of variables might be more effective for other groups just as a function of prior baseline and grade levels. Thus, a package can be custom-made to meet the needs of different students with varying characteristics.

There are several limitations to the study that should lead the reader to exercise caution in interpreting the results. First, participants were frequently and repeatedly instructed and probed on the same passages, which increased opportunities to respond and might have augmented treatment effects above and beyond what would have been obtained if such repeated probing not been done. Second, the passages utilized in this study were shorter in length, averaging 212 words (range = 162-237), than an average story obtained from a curricular reading book, again producing increased opportunities to respond. Third, because there were only four students in the small group for reading instruction, these students might have had higher rates of opportunities to respond than occur in a classroom where small groups are large. As mentioned earlier, as the opportunities to respond increase, the rate of oral reading fluency also increases (Eckert et al., 2002; Levy et al., 1993; Skinner et

al., 1991). In a typical classroom environment, the number of opportunities to respond may be limited based on the number of students in a reading group. Finally, the methods employed in this study may not be readily adapted to the school setting. The analysis was time consuming and employed a plethora of data collection procedures, not easily amenable to practice. However, this study was an initial effort in the area of analysis of academic responding within a small reading group context and provides valuable data that can be applied to future investigations.

The replication of these results is vital to continued evaluation of the procedures, as well as combinations of other variables and modified designs that may enable more efficient application to applied settings. Furthermore, analysis implemented initially by the teacher may alleviate the time restrictions consultants typically face.

The examination of generalization [i.e., a process in which the behavior occurs in the presence of similar antecedent stimuli to the discriminative stimulus present when the behavior was reinforced (Miltenberger, 2001)] across reading passages and moreover, intervention strategies utilized to increase generalization, should also be investigated as results of this study show generalization data that are highly variable. Evidence of generalization for oral reading fluency has been examined through a procedure such as Instructional and Generalization [I&G (Daly, Martens, Dool, & Hintze, 1998)], a form of multiple exemplar training (Stokes & Baer, 1977). The process by which students come to generalize reading and the variables that lead to improved generalization are poorly understood (Daly,

Fountaine, Bonfiglio, Lin, & Corsaut, in submission) based on the lack of analytical investigations of generalized responding. Initial research indicates that the combination of instructional match and a high degree of word content overlap leads to greater generalized increases in reading accuracy and fluency (Daly, Martens, Kilmer, & Massie, 1996).

In spite of the above limitations, the results of this study should encourage future study of experimental analyses of academic performance. The positive treatment effects and the close assimilation of the natural environment (i.e., small reading group instruction) suggest their utility in the classroom. Furthermore, the use of single-case experimental design allowed the investigators to detect changes in performance for each child within the small group. Moreover, this study approximated the natural environment yielding greater external or ecological validity than previous research conducted (e.g., Daly et al., 1998; Daly et al., 1999).

There are several implications for school psychologists as a result of this study. First, because of the nature of this study (i.e., the identification of an effective, efficient reading intervention package for small groups), implementation to the classroom setting is viable. Within a small reading group context, it is valuable to identify a package that may not be the most effective package for *every* child, but for most, in that positive effects may be obtained for every child (i.e., an increase in oral reading fluency rates). Second, the utilization of the Instructional Hierarchy and grouping intervention components accordingly for brief experimental analysis can prove to be fruitful. The dismantling procedure can be utilized in a brief experimental

analysis, making the process more efficient with perhaps the same results. Finally, directly applying the treatment to the target behaviors (i.e., oral reading fluency) in a classroom environment will be beneficial to all students whether difficulties with reading are present in all students (i.e., low fluency rates and/or below instructional level oral fluency).

All in all, this study is an initial but integral piece to the literature of experimental analysis and academic responding. It approximates a natural classroom setting and therefore, suggests higher degrees of ecological validity and utility than previous studies in this area. Further replication and investigation utilizing other variables and combinations of the procedures described (e.g., utilizing the teacher in the classroom to conduct the experimental analysis) would add substantial validity evidence to the literature and would be an extension of the already valuable data described in this study.

Appendix A

Observation Format

Appendix B

Reading Instruction Survey

Reading Instruction Survey

Please rate whether the following sequence is typical for small group reading instruction with elementary school children when a new story is instructed.

1. **Teacher presents an overview of the setting and plot of the story.**

   1 – Not Typical 2 – Somewhat Typical 3 – Typical

   Comments:_____

   _____

2. **Students take turns reading story aloud.**

   1 – Not Typical 2 – Somewhat Typical 3 – Typical

   Comments:_____

   _____

3. **Students read story to self.**

   1 – Not Typical 2 – Somewhat Typical 3 – Typical

   Comments:_____

   _____

4. **Students answer comprehension questions.**

   1 – Not Typical 2 – Somewhat Typical 3 – Typical

   Comments:_____

   _____

Are any typical steps missing? Please provide any suggestions or comments regarding typical reading instruction in your classroom.

_____

_____

_____

Appendix C

Protocol for Effective Treatment

Effective Treatment

Taped Passage Preview:

Say:

Today we are going to read a story. Before we read this story, I would like you to listen to the story on tape. Please follow along as you listen. ☐

Listen to tape together ☐

Choral Reading:

Say:

Now, we are all going to read the story together. I will read it with you. Everyone ready...let's begin. ☐

Teacher and group read the entire passage together. ☐

Students Taking Turns w/Word Drill & Praise:

Say:

Now, we are going to take turns reading the story. Let's begin with (person sitting in chair #1) and go around the group. Ok (name of #1), please begin. I'll tell you when to stop. ☐

Teacher signals for each student to read after a paragraph. ☐

Each person was given the opportunity to read. ☐

Error correction procedure occurs during each student's turn. ☐

For correction, teacher stops the student and says:

The word is _____. Everyone say the word 3 times. ☐

All students say the word 3 times. ☐

As students are reading, teacher uses praise for good behavior and reading. Examples: "Good job!" "Great!" ☐

Session is over:

Say:

That is all for today. Tomorrow we'll have another chance to work again on reading. Thanks. ☐

Appendix D

Protocol for Baseline and Treatment Phases

Baseline Phase

Pre-session:

Draw name of student in chair # 1 ☐

Have student (name drawn) sit in chair 1; the rest can sit where they wish. ☐

Turn on tape recorder ☐

Pass out passage to students ☐

Overview of setting and plot of the story:

Say:

Today we are going to read a story. Before we do, I would like to tell you what it is going to be about. The story _____. ☐

Students take turns reading aloud:

Say:

Now, we are going to take turns reading the story. Let's begin with (person sitting in chair #1) and go around the group. Ok (name of #1), please begin. I'll tell you when to stop. ☐

Signals for each student to read after a few sentences. ☐

Each person was given the opportunity to read. ☐

Students read story to self:

Say:

Now, I would like everyone to take a few minutes and read the story to yourself. Please look at me when you are finished. ☐

Students answer comprehension questions:

Say:

Now we are going to answer a few questions about the story that you just read. ☐

Teacher asks 4 questions and allows each student to answer 1 (if possible). ☐

Session is over:

Say:

That is all for today. Tomorrow (or next week) we'll have another to work again on reading. Thanks. ☐

Treatment Phases

Pre-session:

Draw name of student in chair # 1 ☐

Have student (name drawn) sit in chair 1, the rest can sit wherever they wish. ☐

Turn on tape recorder and say: **Let's begin** ☐

Pass out passage to students. ☐

Withdrawal of components (if this is not the 1st week of treatment):

As the study draws on, the components will be subtracted to develop the most efficient, effective treatment package. Thus, treatment may not consist of every component. ☐

Taped Passage Preview:

Say:

Today we are going to read a story. Before we do, I would like you to listen to the story on tape. Please follow along as you listen. ☐

Choral Reading:

Say:

Now, we are all going to read the story together. I will read it with you. Everyone ready...let's begin. ☐

Teacher and group read the entire passage together. ☐

Students Taking Turns w/Word Drill & Praise:

Say:

Now, we are going to take turns reading the story. Let's begin with (person sitting in chair #1) and go around the group. Ok (name of #1), please begin. I'll tell you when to stop. ☐

Teacher signals for each student to read after a few sentences. ☐

Each person was given the opportunity to read. ☐

Error correction procedure occurs during each student's turn. ☐

For correction, teacher stops the student and says:

The word is _____. Everyone say the word 3 times. ☐

All students say the word 3 times. ☐

As students are reading, teacher uses praise for good behavior and reading. ☐
Examples: *"Good job!" "Great!"*

Contingent Token Reinforcement:

After everyone reads and the word drill and praise are provided, each student will be given the opportunity to earn a reward for meeting a specified goal.

Teacher says:

Who would like to earn a chance to earn a reward for reading this story and beating a goal? It is your choice. You will have 1 minute to do your best reading. If you can read _____ words with _____ or less errors, you can choose a reward from the goodie bag. Would anyone like to try? ☐

If someone wants to try, teacher says:

Ok, I'd like you to read this passage again. You will have 1 minute to do your best reading. Try to read each word, If you come to a word that you don't know, I'll tell it to you. If you can read _____ words with _____ or less errors, you can choose a reward from the goodie bag. Do you have any questions? Ok, begin. ☐

After 1-minute, teacher says:

You read _____ words correctly with _____ errors. ☐

If met the goal, teacher says:

Great work! You met the goal and earned the reward! ☐

If did not meet the goal, teacher says:

That's ok. Next time if you want to try again, you will have another opportunity. ☐

Session is over:

Say:

That is all for today. Tomorrow (or next week) we'll have another chance to work again on reading. Thanks. ☐

Withdrawal of components if an appropriate treatment week:

One component was withdrawn from study. ☐

Appendix E

Intervention Rating Profile-15 (IRP-15)

Intervention Rating Profile-15 (IRP-15)

The purpose of this questionnaire is to obtain information that will aid in the selection of classroom interventions. These interventions will be used by teachers of children with behavior/academic problems. Please circle the number that best describes your agreement or disagreement with each statement using the scale below.

1=Strongly Disagree 2=Disagree 3=Slightly Disagree 4=Slightly Agree 5=Agree 6=Strongly Agree

1.	This would be an acceptable intervention for the child's problem.	1 2 3 4 5 6
2.	Most teachers would find this intervention appropriate for academic problems in addition to the one described.	1 2 3 4 5 6
3.	This intervention should prove effective in changing the academic problem.	1 2 3 4 5 6
4.	I would suggest the use of this intervention to other teachers.	1 2 3 4 5 6
5.	The child's academic problem is severe enough to warrant use of this intervention.	1 2 3 4 5 6
6.	Most teachers would find this intervention suitable for the academic problem described.	1 2 3 4 5 6
7.	I would be willing to use this intervention in the classroom setting.	1 2 3 4 5 6
8.	This intervention would not result in negative side effects for the child.	1 2 3 4 5 6
9.	This intervention would be appropriate for a variety of children.	1 2 3 4 5 6
10.	This intervention is consistent with those I have used in classroom settings.	1 2 3 4 5 6
11.	This intervention was a fair way to handle the child's academic problem.	1 2 3 4 5 6

12. The intervention is reasonable for the academic problem described.	1 2 3 4 5 6
13. I liked the procedures used in this intervention.	1 2 3 4 5 6
14. This intervention was a good way to handle this child's academic problem.	1 2 3 4 5 6
15. Overall, this intervention would be beneficial for the child.	1 2 3 4 5 6

Appendix F

Protocol Clearance From the Human Subjects
Institutional Review Board

Human Subjects Institutional Review Board

Kalamazoo, Michigan 49008-5456
616 387-8293

WESTERN MICHIGAN UNIVERSITY

Date: 28 September 2001

To: Edward Daly, Principal Investigator
Ruth Ervin, Co-Principal Investigator
Christine Bonfiglio, Student Investigator for thesis
Nicole Malzahn, Student Investigator for thesis
Patricia Chin, Miriam Garcia, Laura Krause, Kristy LaFluer, Katherine Lovell, Christine McCaffrey, Michael Persampieri, Tracy Nitz, Jennifer Robinson, and Molly Zakrajsek, as Student Investigators

From: Mary Lagerwey, Chair

Re: Changes to HSIRB Project Number: 00-12-08

This letter will serve as confirmation that the changes to your research project "Academic Assessment and Intervention Project" requested in your memo dated 27 September 2001 have been approved by the Human Subjects Institutional Review Board.

The conditions and the duration of this approval are specified in the Policies of Western Michigan University.

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.

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

Approval Termination: 20 December 2001

**This is a created approval; original is on file at the Graduate College.

BIBLIOGRAPHY

Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behavior, Research, and Therapy, 31,* 621-631.

Aron, A. & Aron, E. N. (1997). Statistics for behavioral and social sciences: A brief course. Upper Saddle River, NJ: Prentice-Hall, Inc.

Barnett, D. W., Daly, E. J., Jones, K. M., & Lentz, F. E. (2003). Measuring intervention effort and needed support? A review of single case design tactics as aids for empirically derived service delivery. Manuscript submitted for publication.

Berliner, D. C. (1984). The half-full glass: A review of research on teaching. In P.L. Hosford (Ed.), *Using What We Know About Teaching* (pp. 51-85). Alexandria: Association for Supervision and Curriculum Development.

Billingsley, F. F. (1977). The effects of self- and externally-imposed schedules of reinforcement on oral reading performance. *Journal of Learning Disabilities, 10,* 21-31.

Bijou, S. W., Peterson, R. F., & Ault, M. H., (1968). A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. *Journal of Applied Behavior Analysis, 1,* 175-191.

Bonfiglio, C. M., Daly, E. J., Martens, B. K., Lin, R. H., & Corsaut, S. (in press). An experimental analysis of reading interventions: Generalization across instructional strategies, passages, and time. *Journal of Applied Behavior Analysis.*

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32,* 513-531.

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J.R. Levin (Eds.), *Single-case research design and analysis: Applications in psychology and education* (pp. 187-212). Hillsdale, NJ: Lawrence Erlbaum Associates.

Busse, R. T., Kratochwill, T. R., & Elliott, S. N. (1995). Meta-analysis for single-case consultation outcomes: Applications to research and practice. *Journal of School Psychology, 33,* 269-285.

Carlson, C. I., Scott, M., & Eklund, S. J. (1980). Ecological theory and method for behavioral assessment. *School Psychology Review, 9,* 75-82.

Carnine, D. W. (1976). Effects of two teacher presentation rates on off-task behavior, answering correctly, and participation. *Journal of Applied Behavior Analysis, 9,* 199-206.

Center, B. A., Skiba, R. J., & Casey, A. (1985-1986). A methodology for the quantitative synthesis of intra-subject design research. *Journal of Special Education, 19,* 187-400.

Clarke, S., Dunlap, G., & Peck-Stichter, J. (2002). A descriptive analysis of intervention research in emotional and behavioral disorders from 1980 through 1999. *Behavior Modification, 26,* 659-683.

Cohen, J. (1965). Some statistical issues in psychological research. In B.B. Wolman (Ed.), *Handbook of clinical psychology,* (pp. 95-121). New York: McGraw-Hill.

Cooper, L. J., Wacker, D. P., Thursby, D., Plagmann, L. A., Harding, J., Millard, T., & Derby, M. (1992). Analysis of the effects of task preferences, task demands, and adult attention on child behavior in outpatient and classroom settings. *Journal of Applied Behavior Analysis, 25,* 823-840.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). Planning, replicating, and evaluating research in applied behavior analysis. *Applied Behavior Analysis.* Upper Saddle River, NJ: Apprentice-Hall, Inc.

Cossairt, A., Hall, R. V., & Hopkins, B. L. (1973). The effects of experimenter's instruction, feedback, and praise on teacher praise and student attending behavior. *Journal of Applied Behavior Analysis, 6,* 89-100.

Daly, E. J., Fountaine, K., Bonfiglio, C. M., Lin, L.H., & Corsaut, S. (In Submission). An investigation of the combined effects of reinforcement, altering text difficulty, and instruction on generalized oral reading performance.

Daly, E. J., Lentz, F. E., & Boyer, J. (1996). The instructional hierarchy: A conceptual model for understanding the effective components of reading interventions. *School Psychology Quarterly, 11,* 369-386.

Daly, E. J., Martens, B. K., Dool, E. J., & Hintze, J. M. (1998). Using brief functional analysis to select interventions for oral reading. *Journal of Behavioral Education, 8,* 208-218.

Daly, E. J., Martens, B. K., Hamler, K. R., Dool, E. J., & Eckert, T. L. (1999). A brief experimental analysis for identifying instructional components needed to improve oral reading fluency. *Journal of Applied Behavior Analysis, 32,* 83-94.

Daly, E. J., Martens, B. K., Kilmer, A., & Massie, D. (1996). The effects of instructional match and content overlap on generalized reading performance. *Journal of Applied Behavior Analysis, 29,* 507-518.

Daly, E. J., & Murdoch, A. (2000). Direct observation in the assessment of academic skills problems. In E.S. Shapiro & T.R. Kratochwill (Eds.), *Behavioral assessment in schools: Theory, research, and clinical foundations (2nd ed.).* New York: The Guilford Press.

Dooley, P., Wilczenski, F. L., & Torem, C. (2001). Using an activity schedule to smooth school transitions. *Journal of Positive Behavioral Interventions, 3,* 57-61.

Dunlap, G., & Kern, L. (1996). Modifying instructional activities to promote desirable behavior: A conceptual and practical framework. *School Psychology Quarterly, 11,* 297-312.

Dunlap, G., Kern-Dunlap, L., Clarke, S., & Robbins, F. R. (1991). Functional assessment, curricular revision, and severe behavior problems. *Journal of Applied Behavior Analysis, 24,* 387-397.

Eckert, T. L., Ardoin, S. P., Daly, E. J., & Martens, B. K. (2002). Improving oral reading fluency: An examination of the efficacy of combining skill-based and performance-based interventions. *Journal of Applied Behavior Analysis, 35,* 271-281.

Elliott, S. N. (1988). Acceptability of behavioral treatments: Review of variables that influence treatment selection. *Professional Psychology: Research and Practice, 19,* 68-80.

Ervin, R. A., Radford, P. M., Bertsch, K., Piper, A. L., Ehrhardt, K. E., & Poling, A. (2001). A descriptive analysis and critique of the empirical literature on school-based functional assessment. *School Psychology Review, 30,* 193-210.

Espin, C. A., & Deno, S. L. (1989). The effects of modeling and prompting feedback strategies on sight word reading of students labeled learning disabled. *Education and Treatment of Children, 12,* 219-231.

Faith, M. S., Allison, D. B., & Gorman, B. S. (1997). Meta-analysis of single case research. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 245-277). Mahway, NJ: Erlbaum.

Fuchs, L. S., & Deno, S.L. (1982). *Developing goals and objectives for educational programs.* Minneapolis: Institute for Research in Learning Disabilities, University of Minnesota.

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

Gettinger, M. (1995). Best practices for increasing academic learning time. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology-III* (pp. 943-954). Washington, DC: National Association of School Psychologists.

Gingerich, W. J. (1984). Meta-analysis of applied time-series data. *Journal of Applied Behavioral Science, 20,* 71-79.

Gorsuch, R.L. (1983). Three methods for analyzing limited time-series (N of 1) data. *Behavioral Assessment, 5,* 141-154.

Greenwood, C. R. (1996). The case for performance-based instructional models. *School Psychology Quarterly, 11,* 283-296.

Greenwood, C. R., Carta, J .J., Kamps, D., Terry, B., & Delquadri, J. (1994). Development and validation of standard classroom observation systems for school practitioners: Ecobehavioral assessment systems software (EBASS). *Exceptional Children, 6,* 197-210.

Greenwood, C. R., Delquadri, J., & Hall, R. V. (1984). Opportunity to respond and student academic performance. In W. Heward, T. Heron, D. Hill, & J. Trapporter (Eds.), *Behavior analysis in education* (pp. 58-88). Columbus, OH: Charles E. Merrill.

Greenwood, C. R., Hart, B., Walker, D., & Risley, T. (1994). The opportunity to respond and academic performance revisited: A behavioral theory of developmental retardation and its prevention. In R. Gardner III, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. W. Eshleman, & T. A. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 213-224). Pacific Grove, CA: Brooks/Cole Publishing Co.

Greenwood, C. R., Terry, B., Marquis, J., & Walker, D. (1994). Confirming a performance-based instructional model. *School Psychology Review, 23,* 652-668.

Gresham, F. M. (1981a). Assessment of children's social skills. *Journal of Psychoeducational Assessment, 1,* 299-307.

Gresham, F. M. (1981b). Social skills training with handicapped: A review. *Review of Educational Research, 51,* 139-176.

Gresham, F. M., Watson, T. S., & Skinner, C. H. (2001). Functional behavioral assessment: Principles, procedures, and future directions. *School Psychology Review, 30,* 156-172.

Hall, R. V., Lund, D. E., & Jackson, D. (1968). Effects of teacher attention on study behavior. *Journal of Applied Behavior Analysis, 1,* 1-12.

Harding, J., Wacker, D. P., Cooper, L. J., Millard, T., & Jensen-Kovalan, P. (1994). Brief hierarchical assessment of potential treatment components with children in an outpatient clinic. *Journal of Applied Behavior Analysis, 27,* 291-300.

Haring, N. G., & Eaton, M. D. (1978). Systematic instructional procedures: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton, & C. L. Hansen (Eds.), *The Fourth R: Research in the Classroom* (pp. 23-40). Columbus, OH: Charles Merrill.

Hendrickson, J. M., & Gable, R. A. (1981). The use of modeling tactics to promote academic skills development of exceptional learners. *Journal of Special Education Technology, 4,* 20-29.

Heward, W. (1994). Three "low-tech" strategies for increasing the frequency of active student response during group instruction. In R. Gardner III, E. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. W. Eshleman, & T. A. Grossi (Eds.), *Behavior Analysis in Education: Focus on Measurably Superior Instruction* (pp. 283-320). Pacific Grove, CA: Brooks/Cole Publishing Co.

Hoge, R. D., & Andrews, D. A. (1987). Enhancing academic performance: Issues in target selection. *School Psychology Review, 16,* 228-238.

Howell, K. W., & Nolet, V. (2000). Curriculum-based evaluation: Teaching and decision making (3rd ed.). Belmont, CA: Wadsworth

Iwata, B. A., Vollmer, T. R., & Zarcone, J. R. (1990). The experimental (functional) analysis of behavior disorders: Methodology, applications, and limitations. In A. C. Repp & N.N. Singh (Eds.), *Perspectives on the use of nonaversive and aversive interventions for persons with developmental disabilities* (pp. 301-330). Sycamore, IL: Sycamore Publishing.

Jayaratne, S., Tripodi, T., & Talsman, E. (1988). The comparative analysis and aggregation of single-case data. *Journal of Applied Behavioral Science, 24,* 119-128.

Jenkins, J. R., Barksdale, A., & Clinton, L. (1978). Improving reading comprehension and oral reading: Generalization across behaviors, settings, and time. *Journal of Learning Disabilities, 11,* 5-15.

Kazdin, A. E. (1980a). Acceptability of alternative treatment for deviant child behavior. *Journal of Applied Behavior Analysis, 13,* 259-273.

Kazdin, A. E. (1980a). Acceptability of timeout from reinforcement procedures for disruptive child behavior. *Behavior Therapy, 11,* 329-344.

Kazdin, A. E. (1981). Acceptability of child treatment techniques: The influence of treatment efficacy and adverse side effects. *Behavior Therapy, 12,* 493-506.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings.* New York: Oxford University Press.

Lalli, J. S., Browder, D. M., Mace, F. C., & Brown, D. K. (1993). Teacher use of descriptive analysis data to implement interventions to decrease students' problem behaviors. *Journal of Applied Behavior Analysis, 26,* 227-238.

Lalli, E. P., & Shapiro, E. S. (1990). The effects of self-monitoring and contingent reward on sight word acquisition. *Education and Treatment of Children, 13,* 129-141.

Lentz, F. E. (1988). On-task behavior, academic performance and classroom disruptions: Untangling the target selection problem in classroom interventions. *School Psychology Review, 17,* 243-257.

Lentz, F. E. & Shapiro, E. S. (1986). Functional assessment of the academic environment. *School Psychology Review, 15,* 346-357.

Lehrman, D. C., & Iwata, B .A. (1993). Descriptive and experimental analyses of variables maintaining self-injurious behavior. *Journal of Applied Behavior Analysis, 26,* 293-319.

Leper, M. R. (1985). Microcomputers in education. *American Psychologist, 40,* 1-18.

Levy, B. A., Nicholls, A., & Kohen, D. (1993). Repeated readings: Process benefits for good and poor readers. *Journal of Experimental Child Psychology, 56,* 303-327.

Mace, F. C., & Lalli, J. S. (1991). Linking descriptive and experimental analyses in the treatment of bizarre speech. *Journal of Applied Behavior Analysis, 24,* 553-562.

MacMann, G. M., & Barnett, D. W. (1999). Diagnostic decision making in school psychology: Understanding and coping with uncertainty. In C. R. Reynolds & T. B. Gutkin (Eds.), *The Handbook of School Psychology* (pp. 519-548). New York, NY: John Wiley & Sons, Inc.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children, 52,* 266-276.

Martens, B. K., & Witt, J. C. (1988). On the ecological validity of behavioral modification. In J. C. Witt and S. C. Elliott (Eds.), *Handbook of Behavior Therapy in Education* (pp 325-341). New York, NY: Plenum Press.

Martens, B. K., Witt, J. C., Daly, E. J., & Vollmer, T. R. (1999). Behavior analysis: Theory and practice in educational settings. In C. R. Reynolds & T. B. Gutkin (Eds.), *The Handbook of School Psychology* (pp. 638-663). New York, NY: John Wiley & Sons, Inc.

Martens, B. K., Witt, J. C., Elliot, S. N. & Darveaux, D. X. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology: Research & Practice, 16,* 191-198.

McComas, J. J., Wacker, D. P., Cooper, L. J., Asmus, J. M., Richman, D., & Stoner, B. (1996). Brief experimental analysis of stimulus prompts for accurate responding on academic tasks in an outpatient clinic. *Journal of Applied Behavior Analysis, 29,* 397-401.

McGlinchey, M. T., & Hixson, M. D. (in press). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review.*

Miltenberger, R. G. (2001). Behavior modification: Principles and procedures. Belmont, CA: Wadsworth/Thomson Learning.

Moncrieff, J. (1998). Research synthesis: Systematic reviews and meta-analysis. *International Review of Psychiatry, 10,* 304-311.

Morgan, D. L. & Morgan, R. K. (2001). Single-participant research design: Bringing science to managed care. *American Psychologist, 26,* 119-127.

National Association of School Psychologists (2000). *Standards for training and field placement programs in school psychology.* **Retrieved November 1, 2001, from http://www.naspweb.org/whats_new/index.html.**

The National Center for Education Statistics (2001). The nation's report card: Fourth grade reading 2000. *The National Center for Education Statistics.* **Retrieved November 1, 2001, from http://nces.ed.gov/nationsreportcard/.**

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology.* San Francisco: Freeman.

Paine, S. C., Radicchi, J., Rosellini, L. C., Deutchman, L., & Darch, C. B. (1983). *Structuring your classroom for academic success.* Champaign, IL: Research Press Company.

Reschly, D. J., & Ysseldyke, J. E. (1995). School psychology paradigm shift. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (pp. 17-32). Washington, DC: The National Association of School Psychologists.

Risley, T. R. (1970). Behavior modification: An experimental-therapeutic endeavor. In L. A. Hamerlynck, P. O. Davidson, & L. E. Acker (Eds.), *Behavior modification and ideal mental health services.* Calgary, Alberta: University of Calgary Press.

Rose, T. L., & Beattie, J. R. (1986). Relative effects of teacher-directed and taped previewing on oral reading. *Learning Disability Quarterly, 9,* 193-199.

Rosenberg, M. S. (1986). Error-correction during oral reading: A comparison of three techniques. *Learning Disability Quarterly, 9,* 182-192.

Sainato, D. M., Strain, P. S., & Lyon, S. L. (1987). Increasing academic responding of handicapped preschool children during group instruction. *Journal of the Division of Early Childhood Special Education, 12,* 23-30.

Sasso, G. M., Reimers, T. M., Cooper, L. J., Wacker, D., Berg, W., Steege, M., Kelly, L., & Allaire, A. (1992). Use of descriptive and experimental analyses to identify the functional properties of aberrant behavior in school settings. *Journal of Applied Behavior Analysis, 25,* 809-821.

Shapiro, E. S. (1996a). *Academic Skills Problems: Direct Assessment and Intervention* (2nd ed.). New York, NY: The Guilford Press.

Shapiro, E. S. (1996b). *Academic Skills Problems Workbook*. New York, NY: The Guilford Press.

Shinn, M. R. (1989). *Curriculum-Based Measurement: Assessing Special Children*. New York, NY: The Guilford Press.

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. Boston, MA: Authors Cooperative, Inc.

Skinner, C. H., Ford, J. M., & Yunker, B. D. (1991). An analysis of instructional response requirements on the multiplication performance of behavior disordered students. *Behavior Disorders, 17,* 56-65.

Spache, G. (1953). A new readability formula for primary grade materials. *Elementary English, 53,* 410-413.

Stanley, S. O. & Greenwood, C. R. (1983). Assessing opportunity to respond in classroom environments through direct observation: How much opportunity to respond does the minority, disadvantaged student receive in school? *Exceptional Children, 49,* 370-373.

Stanley, S. O., & Greenwood, C. R. (1981). Code for instructional structure and student academic response (CISSAR): Observer's manual. Kansas City: Juniper Gardens Children's Project, Bureau of Child Research, University of Kansas.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349-367.

Tan, A., Moore, D. W., Dixon, R. S., & Nicholson, T. (1994). Effects of training in rapid decoding on the reading comprehension of adult ESL learners. *Journal of Behavioral Education, 4,* 177-189.

Tarnowski, K. J., & Simonian, S. J. (1992). Assessing treatment acceptance: The abbreviated acceptability rating profile. *Journal of Behavior Therapy and Experimental Psychiatry, 23,* 101-106.

Taylor, J. C., & Romanczyk, R. G. (1994). Generating hypotheses about the function of student problem behavior by observing teacher behavior. *Journal of Applied Behavior Analysis, 27,* 251-265.

University of Oregon. Dynamic Indicators of Basic Early Literacy Skills (DIBELS). *University of Oregon.* **Retrieved March 12, 2003 from** http://dibels.uoregon.edu.

White, D. M., Rusch, R. R., Kazdin, A. E., & Hartmann, D. P. (1989). Applications of meta-analysis in individual subject research. *Behavioral Assessment, 11,* 281-296.

Witt, J. C., Daly, E. J., & Noell, G. (2000). *Functional assessments: A step-by-step guide to solving academic and behavior problems.* Longmont, CO: Sopris West.

Witt, J. C., & Elliott, S. N. (1985). Acceptability of classroom management strategies. In T. R. Kratochwill (Ed.), *Advances in school psychology.* (Vol.4, pp. 251-288). Hillsdale, NJ: Erlbaum.

Witt, J. C., & Martens, B. K. (1983). Assessing the acceptability of behavioral interventions used in classrooms. *Psychology in the Schools, 20,* 510-517.

Witt, J. C., Martens, B. K., & Elliott, S. N. (1984). Factors affecting teachers' judgments of the acceptability of behavioral interventions: Time involvement, behavior problem severity, and type of intervention. *Behavior Therapy, 15,* 204-209.

Wolery, M., Bailey, D. B., Jr., & Sugai, G. M. (1988). *Effective teaching: Principles and procedures of applied behavior analysis with exceptional children.* Boston: Allyn & Bacon, Inc.

Wolery, M. & Harris, S. R. (1982). Interpreting results of single-subject research designs. *Physical Therapy, 62,* 445-452.

Ysseldyke, J. E., Dawson, P., Lehr, C., Reschly, D., Reynolds, M., & Telzrow, C. (1997). *School psychology: A blueprint for training and practice II.* Bethesda, MD: National Association of School Psychologists.