12-2004

Evaluating Progress in Behavioral Programs for Children with Pervasive Developmental Disorders: Continuous Versus Intermittent Data Collection

Anne Rena Cummings
Western Michigan University

EVALUATING PROGRESS IN BEHAVIORAL PROGRAMS FOR CHILDREN WITH PERVASIVE DEVELOPMENTAL DISORDERS: CONTINUOUS VERSUS INTERMITTENT DATA COLLECTION

by

Anne Rena Cummings

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

ADVISOR: DR. JAMES E. CARR

Western Michigan University
Kalamazoo, Michigan

December 2004

UMI Number: 3154489

ACKNOWLEDGMENTS

I would like to begin by thanking my advisor, Dr. James Carr, for believing in me and teaching me to believe in myself, as well as for proving to me that even at the age of forty I could enter graduate school and make my mind work again this way. I would also like to thank Jim for his constant encouragement and support, and for taking the time to shape my behavior as a clinician, researcher and professor. It has truly been an honor to study with him, and his enthusiasm for, and support of research and academic work in the field of Applied Behavior Analysis has inspired me to pursue the field, and ultimately led to the work contained in this dissertation. Jim's love for research and teaching will continue to inspire me throughout my career.

Secondly, I would like to thank the people that took the time to discuss with me their perceptions of the topic contained herein, specifically Tina and Dave Sidener and Eric Davis, who have given their time to discuss or read earlier drafts and help to coalesce my ideas into something tangible. I would also like to thank my committee members and mentors Dr. Jack Michael, Dr. Linda LeBlanc, Dr. Scott Gaynor, Dr. Nickola Nelson and my advisor and chair, Dr. James Carr for taking the time to review and shape my work, and for the excellent teaching and supervision I have received from them throughout my graduate career.

Lastly, I would like to thank my research assistants Kendra Priest and Stacy Osredkar for their dedicated and arduous work in helping to collect and analyze the data. Without them this comprehensive task would have been overwhelming.

Moreover, I would like to thank my family for their never-ending patience, support and belief in me.

Anne Rena Cummings

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Purpose

Applied behavior analysis is dedicated to producing reliable and predictable improvements in socially important behavior (Baer, Wolf, & Risley, 1968). One aspect of this process is the quantitative analysis of behavior, which is typically accomplished by operationalizing target behaviors and objectively recording them *throughout* the treatment process. Given this tradition, most clinicians presumably use *continuous data collection* systems in which data are recorded for every trial or learning opportunity. However, some clinicians have recently advocated that *intermittent data collection* (e.g., data recorded only for the first trial in a session) is a valid and time-saving alternative to continuous data methods (Dollins & Carbone, 2003; Sundberg & Hale 2003). Although, these alternative approaches have generated a substantial amount of recent discussion and disagreement among clinicians, there is little empirical evidence to inform the issue. The current study compared the two most extreme values of these data collection systems: trial-by-trial continuous data collection versus first-trial-only intermittent data collection.

Autism

Individuals with autism have historically provided unique challenges to the professionals who work with them, primarily due to the particular set of behaviors characterizing the syndrome, their pervasiveness, and elusive etiology. Leo Kanner first identified autism in 1943 as he was sub-classifying a unique group of children

who demonstrated relatively common characteristics and who differed from children with "childhood psychosis." Kanner described the syndrome of "early infantile autism" in an article he published entitled "Autistic Disturbances of Affective Contact." The 11 children he described showed common signs of dramatic social withdrawal, communication disorders, rigidity for sameness in their environment, and a predominance of stereotypic behaviors.

When autism is diagnosed in North America, the diagnosis is typically based on the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (*DSM-IV;* American Psychiatric Association, 2000). The *DSM-IV* includes 12 diagnostic criteria for "autistic disorder," which are grouped into three areas: (1) deficits in reciprocal social interaction, (2) deficits in vocal and nonvocal communication, and (3) a restricted repertoire of activities and interests. Within each area are four specific criteria, each representing a different symptom. The first criterion in each area is generally the one that can be detected at the earliest age, with the latter criteria becoming more apparent later in development.

Autism belongs to a diagnostic family of disorders known as the Pervasive Developmental Disorders (PDD; also known as "autism spectrum disorders"). According to the *DSM-IV*, the disorders within the PDD category are: (1) Autistic Disorder, (2) PDD, Not Otherwise Specified, (3) Asperger's Disorder, (4) Rett's Disorder, and (5) Childhood Disintegrative Disorder.

The current prevalence of autism is estimated to be as high as 10 out of every 10,000 children (Volkmar, Lord, Bailey, Schultz, & Klin, 2004). When both autism

and non-autistic PDDs are considered together, autism spectrum disorders occur in approximately 10 to 15 out of every 10,000 children (Siegel, 1996). Although these children share the same broad diagnosis, their behavioral symptoms vary greatly. In fact, variability may be one of the most commonly used descriptors of the characteristics of individuals with autism. Although all individuals with PDDs have deficits in reciprocal social interaction, deficits in vocal and nonvocal communication, and a restricted repertoire of activities and interests, the expression of those difficulties may differ immensely in both type and severity. Furthermore, the impact of those characteristics can change across the developmental period (Waterhouse, Fein, Nath, & Snyder, 1987).

A History of the Behavioral Treatment of Autism

Early attempts to treat autism were products of the "Zeitgeist" of the post-war era in psychology and medicine that were dominated by the psychodynamic approach. However, treating autism from this approach did not prove effective (DeMyer, Hingtgen, & Jackson, 1981; Rutter, 1974; Wing, 1966). Autism was among the first developmental disorders to be treated utilizing a behavioral approach (Van Houten, 1990). The behavioral intervention literature of the 1960s was pivotal in developing methods for teaching the self-care, basic communication, and pre-academic repertoires that would be the first steps toward autonomous functioning for children with autism. Much of the foundation in the development of behavioral services for people with developmental disabilities was provided by early behavior analysts from within the

educational and residential systems (e.g., Allyon & Michael, 1959; Bijou & Baer, 1961).

Ferster (1961) made the first significant connection between learning theory and autism. In his seminal theoretical paper, Ferster suggested that the aberrant behavior of children with autism was maintained by operant variables (e.g., reinforcement) and, thus, could be modified using operant techniques. Later, Ferster and DeMyer (1961) actually demonstrated that the same behavioral procedures used in the laboratory were relevant to the behavior of children diagnosed with autism. The authors reported a series of experiments in which children with autism were taught simple behaviors (e.g., pulling levers) by providing edible consequences contingent upon these behaviors. Thus, these early experiments demonstrated that the behavior of children with autism could be reliably brought under environmental control (Lovaas, 1979). These studies stimulated a substantial amount of research in which investigators began using the methodology of the experimental analysis of behavior to extend the behavioral model to other, more socially relevant behaviors such as tantrums, aggression, and speech (Schreibman, 1988).

Subsequently, Lovaas and his students at the University of California, Los Angeles conducted a series of experiments that employed various behavioral procedures to change the behavior of children diagnosed with autism in socially important ways (Lovaas, 1966). It was through the work of Lovaas that many of the now well-established methods for teaching language and social behavior to children with autism were developed. These early studies clearly established that autistic

symptoms could be drastically improved by consistent behavioral teaching interventions, and they provided the foundation for Lovaas' later groundbreaking work.

In 1987, Lovaas published a seminal study demonstrating that intensive behavioral treatment of early childhood autism could result in the development of elaborate repertoires such as language, and help reduce undesirable behaviors such as aggression and stereotypy. In this study, Lovaas reported on a behavioral intervention project that began in 1970. The goals of this project were to maximize treatment gains by implementing behavioral treatment in an unusually intensive manner. He assigned 38 children to one of two groups that both received treatment for more than 2 years during the study. The experimental group received an average of 40 hours per week of intensive, one-to-one behavioral treatment from trained staff, as well as extensive parent training which allowed treatment to be implemented during most of the participants' waking hours. Treatment included reinforcement and shaping of appropriate behaviors and punishment of aberrant behaviors, all of which occurred within a hierarchically organized developmental curriculum. The control group received less than 10 hours per week of the same behavioral treatment, no parent training, and no punishment interventions. Results indicated that, although there were no significant differences between the two groups at the onset of the study, at follow-up the experimental group scores were significantly higher than control-group scores on both educational placement and IQ. By the end of the study, of the 19 children in the experimental group 47% (9) completed grade 1 in general education, 42% (8)

completed grade 1 in classes for children with aphasia, and 10% (2) were placed in classes for children with autism. Overall, experimental-group participants gained an average of 30 IQ points, with an increase from 2 participants being classified with a normal IQ at onset to 12 participants being classified with a normal IQ at follow-up. Mean IQ of the control-group participants remained essentially unchanged.

The children who participated in the Lovaas (1987) investigation were evaluated approximately 6 years later by McEachin, Smith, and Lovaas (1993). This effort assessed some of the original participants who were then at an average age of 11.5 years. The authors documented that the original experimental group maintained its gains over the control group. In particular, the authors concentrated on the 9 participants from the experimental group who had attained the best outcomes during the original study, and found that 8 of them were "indistinguishable from average children on tests of intelligence and adaptive behavior" (p. 359).

Characteristics of Intensive Behavioral Treatment of Autism

It has been shown that children with autism do not readily learn from typical environments, although research shows that they can learn a great deal when provided with appropriate treatment (Harris & Handleman, 1994; Koegel & Koegel, 1995; Lovaas & Smith, 1989; Schreibman, 1988). Dunlap, Koegel, and O'Neill, (1985) outlined 3 steps for designing a behavioral treatment plan for children with autism. First, one carefully and precisely defines the behaviors targeted for change. Second, one decides how to accurately measure the behavior. Third, based on the first two

steps, a treatment intervention can be designed and implemented. These steps, of course, are applicable to all areas of behavioral treatment.

With his early students, Lovaas developed one-on-one, trial-by-trial teaching procedures for establishing attention, academic skills, and language (Lovaas, 1977). Many specific skills must be directly and systematically taught in order for generalization to occur. In addition, specific instruction must be provided to produce response "spontaneity" (i.e., prompt-independent responding). Appropriate behavioral instruction involves systematically teaching small, measurable units of behavior. Every skill must be broken down into smaller steps that can be taught in a discrete format because many of the curricula are hierarchically sequenced. This discrete-trial training (DTT) format involves presenting specific instructions, which are typically repeated numerous times in rapid succession until the child performs the response easily and independently (Smith, 2001). Sometimes a prompt needs to be added (such as gentle physical guidance) to occasion the appropriate behavior. A correct response is typically followed by immediate praise and brief access to a preferred stimulus. An incorrect response, and the absence of a response, will usually be followed by an error correction procedure consisting of feedback ("try again") and repetition of the trial with a prompt. The learner's responses are recorded and evaluated according to specific, objective definitions and criteria. Thus, the experimenter needs to simultaneously balance this rather intensive acquisition technology with management of problem behaviors and recording all of the child's responses as they are performed each trial. These data must then be evaluated and graphed to provide a visual display of the child's progress to

enable ongoing adjustments. Such adjustments are particularly important given that children can easily emit thousands of responses per week and be exposed to hundreds of behavioral programs over the course of treatment. It is, therefore, essential to have appropriate curricula written and implemented by competent, well-trained behavior analysts designing and supervising behavioral intervention programs for children with autism (Maurice, Green, & Luce, 1996).

Data Collection and Visual Inspection

One of the missions of applied behavior analysis is the production of reliable and predictable improvements in socially important behavior (Baer et al., 1968). A key aspect of this mission is the objective, quantitative analysis of behavior. Data are often collected in a repeated measures fashion, making possible the ongoing analysis of environmental effects on behavior. The repeated, objective assessment of behavior provides clinicians with feedback as to when a learner is progressing and, even more importantly, when a learner is making inadequate progress.

If therapists do not collect direct and frequent measurements of performance, two common errors are possible. First, ineffective intervention programs could be continued and, thus, a program may be carried on for an extended period without any measurable learning occurring. This error denies the learner valuable time in which new skills could be acquired. Second, many effective programs could be discontinued prematurely because subjective impressions are often insensitive in detecting actual behavior change (Cooper, Heron, & Heward, 1987). In addition, data collected during intervention can be useful in evaluating, to some extent, the integrity with which an

intervention is implemented. Unfortunately, despite the importance of data collection, the task of choosing how, where, and what to measure can be difficult.

Because behavior is not static, the behavior analyst must maintain direct and continuous contact with the behavior being examined. These data form the basis on which each decision is made. Making decisions from the actual numbers (raw data) can be cumbersome. However, graphical display of the same data often makes important features of the data more evident. Thus, line graphs are the major tool for organizing, storing, interpreting, and communicating the results of behavioral treatment evaluations (Cooper et al., 1987). As Barlow and Hersen (1984) indicated, data from individual-design approaches can be evaluated in different ways to address experimental and therapeutic change. In behavior analysis, visual inspection is the most commonly used method for evaluating graphical data.

Visual inspection is the process by which visual discriminations are made between data paths within and between phases in an attempt to detect differences. The combination of graphic data presentation and visual inspection has several major benefits (Cooper et al., 1987). First, ongoing access to a complete record of the individual's behavior can facilitate appropriate and timely decisions. Second, direct and continuous contact with the data enables the researcher or clinician to explore interesting variations in behavior as they occur. Third, the graphic display of data allows and encourages independent judgments and interpretations of the meaning and significance of behavior change. Finally, and perhaps most importantly, visual inspection is a conservative method of determining the significance of behavior

change, because only those variables that produce reliable and substantial effects are considered notable.

Continuous versus Intermittent Data Collection

Given the tradition of research and practice in behavior analysis, most clinicians presumably use *continuous data collection* systems in which data are recorded for every trial. Continuous data collection allows for a comprehensive, ongoing account of the learner's performance across all programmed learning opportunities. Despite the relative ease of managing data once collected, as mentioned earlier, considerable effort is still required during data collection. Thus, some clinicians have recently proposed that *intermittent data collection* (e.g., data recorded only for the first trial in a session) is a valid and time-saving alternative to continuous data methods (Dollins & Carbone, 2003; Sundberg & Hale 2003). Although, these alternative approaches have generated a substantial amount of recent discussion and disagreement among clinicians, there is little empirical evidence to inform the issue. Recently, Dollins and Carbone discussed the conflicting opinions of how much data are required to appropriately guide instruction.

Although Dollins & Carbone, (2003) specified that continuous data might represent a better overall sample, they maintained that continuous data collection has several possible limitations. These limitations include: (i) an increase in session length due to frequent data collection, (ii) a delay of reinforcer delivery due to data collection, (iii) a tendency to employ massed rather than interspersed trials, and (iv) possible false mastery of skills. This false mastery could be demonstrated at a later time period with

incorrect responding during maintenance probes. Although these concerns illustrate real problems, the concerns can be easily countered as follows: (i) data collection time should be minimal with a well-designed data sheet, (ii) sufficient training can produce a skilled therapist who can collect data without delaying reinforcer deliveries, (iii) interspersed (rather than massed) trials can be implemented under either mode of data collection, and (iv) it may actually take fewer trials to master a target skill under continuous data collection because continuous data represent a better sample overall.

The proponents of intermittent data collection (Dollins & Carbone, 2003; Sundberg & Hale, 2003) have proposed that the first trial, which they term the "cold trial," should be the only trial recorded. They propose the following benefits of this approach: (i) increased time spent on teaching rather than on data collection, (ii) increased trials conducted on a target skill, (iii) an increase in learner motivation due to an increased pace, (iv) facilitation of interspersed trials rather than mass-trial presentation, and (v) retention from one day to the next rather than a percentage of learning carried over. Dollins and Carbone claim that this cold-probe approach appears to reliably predict mastery of a skill. However, there are no data to support this claim as of yet.

Sundberg (personal communication, May 18, 2003) and Sundberg and Hale (2003) state that collecting cold-probe data seems to convey most of what we need to know and often is more accurate than continuous data. Sundberg continues to declare that most of his research shows that continuous data may show different effects than first trial data; however, he states that "first trial data tell you a lot more than the

continuous within-session data." He clarifies that the first-trial data is an unpracticed, unprompted trial and, therefore, better conveys the learner's true mastery of the skill. He also stresses that cold-probe data show behavior change more quickly than continuous data. It is interesting to note, however, that Dollins and Carbone (2003) and Sundberg and Halle (2003) revert to continuous trial-by-trial data collection to demonstrate experimental control in their research presentations.

A literature search revealed only two studies that are related to the issue of continuous versus intermittent data collection; unfortunately, they are only marginally related in that they did not focus on trial-by-trial data, but rather on overall performance data collected within an educational setting during a daily observation period. In addition, the focus of these studies was to evaluate the impact of data-collection frequency on evaluative judgments of graphical displays of data. Thus, continuous data in these studies refers to data collected every day, and intermittent data in these studies refers to data collected on alternate days. These studies primarily focused on the decisions made by teachers via visual inspection of graphs based on different amounts of data in educational settings.

Bijou, Peterson, Harris, Allen, and Johnston (1969) examined the effects of varying the frequency of observations on data collection and analysis. They displayed data for a 4-year-old boy on his frequency of verbalizations to other children during a play period each day in a "laboratory" nursery setting. These data were originally collected in an earlier study by Bijou, Peterson, and Ault (1968). The authors then displayed this frequency of verbalizations in three separate graphs. The first graph

illustrated the frequency of the data collected continuously (i.e., every day). The second graph included only data from every other day beginning with the first day. The third graph included only data from every other day beginning with the second day. The authors found that teachers interpreted the three graphs similarly, regardless of the continuous or intermittent nature of the graphed data. Therefore, this study seems to support intermittent data collection. However, it should be noted that this study employed intermittent data *analysis* after the fact and did not evaluate the impact of intermittent data *collection* on treatment decisions (e.g., whether a skill meets a mastery criterion).

Munger, Snell, and Loyd (1989) presented four graphs of actual student acquisition data from intervention programs across 60 days of continuous data collection (5 days per week). These four graphs were selected to represent four different data trends (ascending, descending, flat, and variable). Then for each graph, the authors created 3 additional versions of it to represent how the data would appear if they had been collected according to different intermittent frequencies. Thus, one version was created to represent how the data would have appeared if they had been collected 3 times per week (Mondays, Wednesdays, and Fridays). A second version was created by the authors to represent how the data would have appeared if they had been collected 2 times per week (Tuesdays and Thursdays). Finally, a third version was created by the authors to represent how the data would have appeared had they been collected only once per week (on Wednesdays).

The 16 graphs in total were then randomly sequenced and given to 59 randomly selected teachers of students with moderate to profound handicaps. For each graph, each teacher was asked to evaluate the progress of the student by selecting one of five possible statements to describe their performance (i.e., definitely making progress, probably making progress, staying about the same, probably decreasing in performance, and definitely decreasing in performance), and to make program recommendations by selecting one of four possible statements (i.e., definitely continue the program, probably continue the program, probably change the program, and definitely change the program).

The authors found that when graphs illustrated an ascending trend, teachers' assessments were similar regardless of the frequency of the data collection. However, when graphs illustrated descending, flat, or variable trends, teachers' judgements as well as program decisions differed between the continuous and intermittent data collection conditions. The results of this study were mixed in that when rating student progress, frequency of data collection showed statistically significant differences, whereas, when making program recommendations, the frequency of data collection was not found to yield significantly different recommendations. More specifically, these authors found that when graphs illustrated an ascending trend, the teacher's judgments were similar amongst all probe data, thus suggesting that if a student is making progress, these authors would support the use of probe/intermittent data collection as minimally as one time per week. However, these authors also found that when graphs illustrated descending, flat, or variable trends, teacher's judgments and

program decisions differed by recording frequency. Thus this study suggests that when graphs illustrate descending, flat, or variable trends, data need to be collected more often than one time per week and at best should be collected continuously.

In both the Bijou et al. (1969) and Munger et al. (1989) studies the main focus was primarily on the decisions made by teachers via visual inspection of graphs based on different amounts of daily data from educational settings. Both studies examined student performance data collected during continuous and intermittent conditions; however, these data conditions would be different from those that typically exist in early intervention because they do not represent trial-by-trial or session-by-session data, but rather overall daily performance. In other words both of these studies examined data collected over an extended rather than briefer (e.g., 10 trial) period of time. Furthermore, neither of the studies evaluated whether actual treatment decisions were affected by type of data collection.

Given the importance of data collection to autism treatment programs, the effort it requires of therapists, and the current debate over data-collection frequency, further research in this area is clearly warranted. Thus, the purpose of the current study was to experimentally compare continuous versus intermittent data collection systems across a number of curriculum areas in DTT programs for children with autism and other pervasive developmental disabilities. It was essential to compare these data collection systems across different curriculum areas to ensure that the similarities and/or differences were consistent across different curricular domains rather than specific to a certain type of learning program. The current study examined the two

most extreme values of these data collection systems: trial-by-trial continuous data collection versus first-trial-only intermittent data collection. Although some clinicians who support the use of intermittent data collections systems often employ the first 5 of 10 trials, the first 3 of 10 trials, or even the first and last of 10 trials (Dollins & Carbone, 2003; Sundberg & Hale, 2003), the current study evaluated the first-trial-only data collection method as the most extreme variant of the approach.

Three dependent measures were chosen for the present study because they are the measures on which intermittent data collection would likely have an impact. The first measure was the number of 10-trial sessions required to master each skill. The second was the cumulative minutes required to master each skill (across sessions). These measures were selected to evaluate the claims that continuous data collection requires an increased session length of time and can result in delayed reinforcement. The third measure was how well the skill maintained (i.e., percentage correct) at a follow-up assessment. This measure was selected to evaluate the claim that intermittent data collection could possibly lead to false mastery of skills.

Method

Participants and Setting

The participants in this study were six children diagnosed with autism or pervasive developmental disorder (not otherwise specified) between the ages of 4 and 8 years. All participants had at least one month of prior exposure to the DTT format; they exhibited learning deficits in at least 2-3 of 7 common curricular programs (see

Tables 1 and 2); and they were not involved with any other behavioral intervention programs during the course of the study.

Erin was 8-years old and was diagnosed with autism. She had limited vocal skills and was primarily echolalic with single-sound utterances. Erin's scores for each of the 12 items on the Behavioral Language Assessment[1] (BLA; Sundberg & Partington, 1998) are displayed in Table 3. Erin's scores (M = 1.9; SD = .9) are indicative of a rather limited verbal repertoire. Erin attended a special classroom designed for children diagnosed with autism and received speech and language training twice per week, occupational therapy once per week, and a specialized gym class once per week. Throughout the course of the study, Erin consistently received the following medications/supplements every day: Nexium (40 mg.), Vantac (10 ml.), Zyrtec (10 mg.), Singular (5 mg.), and the vitamin supplement Efalex (one pill). Erin's sessions were conducted in her home, in a large room approximately 10 meters by 6.6 meters. Erin's acquisition programs for the study were receptive discriminations and receptive instructions. The interspersed task for Erin was an echoic program.

Jeff was 5-years old and was diagnosed with PDD (not otherwise specified). He had good vocal skills and was able to comment, answer questions, and request using full sentences. Jeff's scores on the BLA (M = 4.3; SD = 1.05), which can be found in Table 3, indicate a rather sophisticated verbal repertoire. Jeff attended an

---

[1] The Behavioral Language Assessment is an informant assessment that contains 12 sections that assess a variety of basic language-related skills (e.g., cooperation, motor imitation, labeling, conversation). Each section is divided into 5 levels. Informants are asked to select a level that best represents the individual's repertoire in that area. Levels 1 and 5 are indicative of minimal and well-developed verbal repertoires,

autism classroom and received no additional services. Jeff did not receive any medication during the course of the study. Jeff's sessions were conducted in the kitchen of his home, which was approximately 3.3 meters by 4 meters. Jeff's acquisition programs for the study were tacts, receptive instructions, intraverbals, and receptive discriminations by function, feature, and class. The interspersed task for Jeff was drawing.

Patrick was 7-years old and was diagnosed with autism. He had good vocal skills and was able to comment, answer questions, and request using full sentences. Patrick's scores on the BLA (*M* = 4.4; *SD* = 1.08), which can be found in Table 3, indicate a rather sophisticated verbal repertoire. Patrick attended an autism classroom with daily integration in a general-education classroom. In addition, Patrick received speech and language training 1-2 times per week (for 8 weeks each year) and occupational therapy once every two weeks. Throughout the course of the study, Patrick received clonidine (.1 mg.) on a daily basis. Patrick's sessions were conducted in his school setting, in a separate therapy room attached to his classroom that was approximately 3.3 meters by 2.6 meters. Patrick's acquisition programs for the study were receptive discriminations, tacts, intraverbals, and receptive discriminations by function, feature, and class. The interspersed task for Patrick was drawing.

Peter was 5-years old and was diagnosed with PDD (not otherwise specified). He had good vocal skills and was able to comment, answer questions, and request using 1-2-word combinations. Although he was able to use full sentences when

---

respectively.

prompted, he did not do so independently. Peter's scores on the BLA (*M* = 4.1; *SD* = 1.24), which can be found in Table 3, indicate a rather sophisticated verbal repertoire. Peter attended a general-education kindergarten classroom and had a full-time educational assistant. In addition, Peter received speech and language training twice per week. Peter did not take any medications during the study. Peter's sessions were conducted in a university research/therapy room which was 3.3 meters by 4 meters and in the kitchen of his home which was approximately 3.3 meters by 4 meters. Peter's acquisition programs for the study were tacts, receptive instructions, intraverbals, and receptive discriminations by function, feature, and class. The interspersed task for Peter was answering social questions.

Mary was 4-years old and was diagnosed with autism. She had limited vocal skills and was primarily echolalic, making single-sound utterances and some single words and approximations. Mary's scores on the BLA (*M* = 2.5; *SD* = 1), which can be found in Table 3, indicate a rather limited verbal repertoire. Mary attended a special classroom for children with developmental delay and received speech and language training three times per week at school and occupational therapy and physical therapy once per week. Throughout the course of the study Mary consistently received the following medications on a daily basis: Stratera (40 mg.) and Risperidol (1 mg.). Mary's sessions were conducted in a university research/therapy room which was 3.3 meters by 4 meters and in the dining room of her home which was approximately 3.3 meters by 4 meters. Mary's acquisition programs for the study were motor imitation, requesting, and tacts. The interspersed task for Mary was receptive discriminations.

Allison was 6-years old and was diagnosed with autism. She had limited vocal skills and was primarily echolalic, making single-sound utterances and some single words and approximations. However, Allison used phrases and small sentences via a voice output machine and a PECS book. Allison's scores on the BLA (*M* = 3.4; *SD* = 1.51), which can be found in Table 3, indicate a moderate verbal repertoire. Allison attended an autism classroom and received speech and language training once per week and a specialized gym class once per week. Throughout the course of the study Allison consistently received the following medications/supplements on a daily basis: zinc, Melatonin, and Chloralhydrate (2.5 mg.). Allison's sessions were conducted in the kitchen of her home, which was approximately 3.6 meters by 3 meters. Allison's acquisition programs for the study were receptive discriminations, requesting, and echoics. The interspersed task for Allison was motor imitation.

Each study visit consisted of 3-5 programs for each participant and lasted approximately 15-20 min in duration. Visits were conducted 2-4 times per day, 3-7 days per week.

Experimenter

It should be noted that the experimenter was a well-trained behavioral consultant with approximately 20 years of experience in delivering DTT to children with pervasive developmental disorders. As a behavioral consultant, the experimenter wrote treatment programs, evaluated client progress, trained staff to implement programs, evaluated staff skills, and directly implemented DTT programs with children.

Specific Acquisition Programs

Eight different behavioral acquisition programs were implemented during the study, and five different programs were implemented during the study for interpersal between trials of the acquisition programs. The eight acquisition programs were as follows: tact/label training (receptive); receptive instruction following; tact/label training (expressive); modified intraverbal training; receptive by function, feature and class; motor (nonvocal) imitation; requesting (modified mand training); and echoic behavior (vocal imitation). The five interspersal programs were as follows: echoic behavior (vocal imitation); drawing; answering social questions; tact/label training (receptive); and motor (nonvocal) imitation. Descriptions of each program can be found in Appendix A, and specific exemplars taught to each participant during each program can be found in Table 4.

Data Collection - Dependent Variables

Each session consisted of 20 trials. The dependent measure for each session was the percentage of correct participant responses. A correct response was defined as the participant responding correctly and independently (i.e., without prompts) to the discriminative stimulus within 5 s. Data were collected continuously or intermittently depending on the experimental condition in effect. The mastery criterion for each skill was two consecutive sessions at 100%. The percentage-correct measure was calculated for each session. These data were then used to compute the first 2 of 3 primary dependent variables, which are described below.

Three primary dependent variables were assessed for every skill taught under each data-collection condition: number of sessions until mastery, cumulative time until mastery, and maintenance. *Number of sessions* was defined as the number of 20-trial sessions required to master each skill (i.e., two consecutive sessions at 100%). *Cumulative time* was defined as the cumulative number of minutes required to master each skill from the first until the last session. *Maintenance* was defined as the percentage of correct responses produced during each maintenance probe (described later).

Interobserver Agreement

Interobserver agreement (IOA) was assessed for every session for each participant. An agreement was defined as two independent observers agreeing on whether a participant's behavior during a trial was correct or incorrect. IOA was calculated using the overall agreement formula: # of agreements / (# of agreements + disagreements) x 100%. Peter's mean IOA score was 98% (range, 90-100%) Erin, Jeff, Patrick, Mary, and Allison's IOA scores were 100%.

Procedures

Curriculum assessment. Before the study was conducted, each participant was administered a brief DTT assessment, which was composed of five trials from each of the common acquisition training programs selected for the study (see Table 1). These acquisition programs were selected from four popular clinical manuals for intensive early intervention (Leaf & McEachin, 1999; Lovaas, 2003; Maurice, Green, & Luce, 1996; Sundberg & Partington, 1998). Each 35-trial assessment session was

administered to each participant at least three times before the study began. Based on the results of this assessment, each participant's programs were selected around their identified deficits. Two of these skill areas were chosen as the acquisition areas for the study and the third skill area was used to intersperse the acquisition training trials with non-acquisition trials. In total, eight different programs were implemented during the study for acquisition programs, and five different programs were implemented during the study for interspersal between trials of the acquisition programs.

As can be seen in Table 2, three of the participants (Erin, Mary, and Allison) had deficits in 4-5 curricular areas. Thus, these participants had several areas from which to choose acquisition programs and at least one area they passed from which to choose an interspersal program. Jeff, Patrick, and Peter had only 2 identified areas with deficits from the original seven curricular areas assessed. Therefore, up to three new more complex skill areas were added to the assessment (intraverbals; receptive by function, form, and class; and social questions). When these new areas were added into the assessment, the participants also had 4 curricular areas from which to select acquisition programs and at least one area they passed from which to select an interspersal program. All acquisition programs, interspersal programs, and specific stimuli chosen can be seen in Tables 2 and 4.

Stimulus preference assessments. The Reinforcer Assessment for Individuals with Severe Disabilities (*RAISD;* Fisher, Piazza, Bowman, & Amari, 1996) interview was administered to the caregivers of each participant by the experimenter. The RAISD is a structured interview during which caregivers nominate a number of

putative reinforcers for their child's behavior in the following categories: visual, audible, olfactory, edible, social, tactile, and toys. These are merely suggested categories and do not require nominations in each. In addition, caregivers were asked to specify the conditions under which these stimuli are preferred (e.g., blowing bubbles while sitting in a rocking chair). When the list had been compiled, caregivers were then asked to rank order the stimuli according to their predicted preference by their child. In addition, the experimenter directly observed each participant during his or her leisure time to determine what activities and foods were frequently contacted in the natural environment. Based on the results of the RAISD and direct observation, a pool of stimuli was selected for subsequent stimulus preference assessments.

Two brief multiple-stimulus preference assessments (MSWO; Carr, Nicolson, & Higbee, 2000) were conducted for each participant at the beginning of each visit. Eight toys and 8 edibles were presented in different assessments, each of which consisted of three stimulus arrays. Before the first MSWO assessment, the participant was given a sample of each edible to ingest and each toy to access (20 s). Items in the stimulus array were quasi-randomly placed in a horizontal row on the tabletop approximately four inches apart. The experimenter then instructed the participant to select one stimulus (e.g., "Pick one."). After the selection was made, the participant was given either the edible to ingest or 30-s access to the toy. The chosen item was then removed from the array (toy) or not replaced (edible). The positions of the remaining items were changed by taking the item at the left of the array and moving it to the right, then shifting the other items so that they were again equally spaced. This

process was repeated until all stimuli had been selected and was then repeated twice. A selection percentage was calculated by dividing the number of times a specific item was chosen by the number of times that item was available. The top three stimuli from the pre-session assessments were used during the subsequent sessions.

General training method. During the course of the study, each participant received DTT in 3-5 training programs (including interspersal programs). Each skill was taught in a 20-trial, interspersed training format until the participant reached the mastery criterion. The acquisition program trials comprised 10 of the trials and were scored in either the continuous or intermittent data collection condition. The other 10 trials consisted of non-acquisition program trials and were inserted between each acquisition program trial to constitute the interspersed format. The framework for each acquisition (scored) training trial was as follows. First, the experimenter delivered an S^D to the participant, who was given 5 s to respond. If the participant responded correctly, he or she received immediate praise from the experimenter and brief access to one of the preferred stimuli. If the participant did not respond or responded incorrectly, the experimenter immediately provided corrective feedback (e.g., “try again”) and issued the S^D again. If the participant responded correctly, he or she received immediate praise. If not, the experimenter used a verbal, gestural, or mild physical prompt (depending on the participant and skill) to occasion the correct response. An incorrect response, or no response, was followed by an error correction procedure consisting of feedback (“try again”) and repetition of the trial with a prompt. A correct prompted response was followed by social praise only.

The framework for non-acquisition (unscored) trials was as follows. First, the experimenter delivered an S^D to the participant, who was given 5 s to respond. Immediately following the participant's response (correct or incorrect) the experimenter delivered a neutral statement (e.g., "okay"). If the participant did not respond within 5 s, the experimenter immediately moved on to the next scored acquisition trial.

Training. The term "training set" will be used to refer to two separate skills being taught concurrently (one per data-collection condition) in the same skill area. Thus, two consecutive 20-trial sessions, one in each data-collection condition, were administered during each "training set." Session order varied across sets. No more than four training sets were implemented during each visit with an individual participant. Each participant continued in an acquisition program area until he/she met the mastery criterion for 4 to 8 skills. Participants were taught 16 skills in each of two acquisition training programs, 12 skills in each of three acquisition training programs, or 8 skills in each of four acquisition training programs. During training on one of the skills, the experimenter recorded data on every trial (continuous data collection condition) and during training on the other skill the experimenter only recorded data from the first of 20 trials (intermittent data collection condition). For example, if a participant was being taught receptive discriminations, he was taught to discriminate "car" under continuous data collection and "book" under intermittent data collection (both comprised training-set one). When the participant met the mastery criteria for both "car" and "book," training on those items ended and discrimination training began for

"shoe" under continuous data collection and "brush" under intermittent data collection (these comprised training-set two).

Jeff, Patrick, and Peter mastered the first few skills within their initial acquisition training areas very rapidly. Therefore, their protocol was changed from two acquisition programs with 16 exemplars to four programs with 8 exemplars. This rapid learning also occurred for both Mary and Allison within one of their two acquisition programs. Thus, their protocol was changed from two acquisition programs with 16 exemplars to three programs with 12 exemplars.

Condition I – Continuous data collection. During this condition the experimenter recorded each participant's responses during every acquisition trial. Thus, each session score represented the percentage of correct responses for the 10 acquisition trials. Performance on the 10 interspersed trials was not included in this score.

Condition II – Intermittent data collection. During this condition, the experimenter only recorded the participant's response for the first of 20 trials. Thus, each session score represented 0% or 100% correct. Since no data were recorded after the first trial, it was possible for the experimenter to lose track of the number of trials left to be administered. In order for the experimenter to administer exactly 10 acquisition trials in a session, she came to the session with 20 different stimuli: 10 stimuli for the acquisition program and 10 stimuli for the interspersal program. This helped to ensure that the experimenter conducted 10 acquisition trials, which is a function served by data collection in the continuous data-collection condition.

Maintenance Probes

To assess the study's third primary dependent measure, maintenance probes were conducted for each skill 3 weeks after it had been mastered. Each probe was implemented using the general-teaching format described above except that it was administered in a 5-trial, massed-trial training format with no programmed consequences (i.e., no preferred stimuli or error correction were provided). Data were recorded for every trial during maintenance probe sessions.

Experimental Design and Data Analysis

A variation of the alternating treatments design was used to evaluate, for each participant, the "number of sessions" and "maintenance" dependent measures across both data-collection conditions. Thus, each participant's data resulted in 4-8 bar graphs, each of which included two data series to represent the two data-collection conditions. Each bar graph was visually inspected for reliable differences between conditions.

Treatment Integrity

Because the primary focus of the study involved a comparison between two conditions, treatment integrity was assessed in two ways. First, The experimenter's behavior was scored during each trial to determine whether (a) stimuli were presented correctly, (b) the instruction was presented correctly, (c) the correct programmed consequence (e.g., a preferred stimulus, a verbal prompt) was immediately delivered after the participant's response, and (d) the experimenter collected data on the trial (these data were examined for each session to verify the presence of continuous or

intermittent data collection). For each session, the experimenter's score for each of these behaviors was computed as the number of correct behaviors divided by the total number of behaviors. This treatment-integrity measure was calculated for 100% of the sessions and averaged 100% for all 6 participants. IOA was assessed for 100% of these sessions using the overall agreement formula. An agreement was defined as two independent observers agreeing on whether the experimenter's behavior was correct for the entire trial. Mean IOA for this treatment-integrity measure was 100%.

The second measure of treatment integrity involved experienced clinicians rating the quality of DTT of the experimenter from samples of each condition. A video was edited to display 5 trials of 4 programs being conducted with 3 different participants, producing a total of 60 trials. Half of the trials were from the continuous data collection condition and half were from the intermittent data collection condition. The experimenter was videotaped from the neck up, so as not to demonstrate any data collection behavior, or lack thereof, during trials. The experimenter's behavior was rated after each block of 5 trials to determine whether (a) the instruction and S^D were given clearly with proper emphasis, (b) different vocal tones were used for the instruction versus praise, (c) program pacing was appropriate for each participant, (d) consequences were delivered immediately, (e) differential reinforcement was provided for prompted trials, and (f) disruptive behavior was controlled appropriately and effectively. All six behaviors were rated on a likert scale that included three anchors: 1 ("did not happen at all"), 2 ("inconsistent performance"), and 3 ("fully consistent performance each trial") (see Appendix D). Each of the 12 video samples was rated by

nine senior therapists from a DTT program in Canada who were blind to the purposes of the study. These senior therapists write DTT programs, evaluate DTT programs, evaluate DTT therapist skills, and implement DTT programs on a daily basis for their profession. These nine senior therapists did not have a relationship with the experimenter at the time of the study. However, it was known that at some point the experimenter would supervise the therapists. Therefore, to ensure the integrity of the ratings, the therapists scored the videotape segments anonymously and sent the data in a signed, sealed envelope directly to a third party. Furthermore, the therapists were not informed about the purpose of the study. The data from this assessment indicate that both conditions were implemented equally and consistently across all of the dimensions, with a score of 3 ("fully consistent performance each trial") for each segment by each of the nine raters (see Table 5).

Results

Evaluation of Acquisition Rates

As seen in Figure 1, for Erin's receptive discrimination program, the average sessions to mastery were 8 for the continuous data collection system and 7 for the intermittent data collection system. For Erin's receptive instruction following program, the average sessions to mastery were 6.5 for the continuous data collection system and 5.5 for the intermittent data collection system. As seen in Table 7, exemplars in 11 of Erin's 16 training sets were mastered more quickly during intermittent data collection, exemplars in 4 of the 16 training sets were mastered more quickly during continuous data collection, and there was no difference with 1 training set. Likewise, as seen in

Table 8, exemplars in 10 of the 16 training sets were better maintained at the 3-week follow-up in the continuous data collection condition, an exemplar in 1 training set was better maintained under intermittent data collection, and exemplars in 5 of the 16 training sets were equally maintained.

As seen in Figure 2, for Jeff's tact program, the average sessions to mastery were 4.25 for the continuous data collection system and 4 for the intermittent data collection system. For Jeff's intraverbal program, the average sessions to mastery were 3.25 for both the continuous and intermittent data collection systems. For Jeff's receptive instruction following program, the average sessions to mastery were 5 for the continuous data collection system and 4.75 for the intermittent data collection system. For Jeff's receptive by function, feature, and classification program, the average sessions to mastery were 3.25 for the continuous data collection system and 2.75 for the intermittent data collection system. As seen in Table 7, exemplars in 5 of Jeff's 16 training sets were mastered more quickly during intermittent data collection, 0 of the 16 training sets were mastered more quickly during continuous data collection, and there was no difference with 11 training sets. Likewise, as seen in Table 8, exemplars in 2 of the 16 training sets were better maintained at the 3-week follow-up in the continuous data collection condition, exemplars in 0 training sets were better maintained under intermittent data collection, and exemplars in 14 of the 16 training sets were equally maintained.

Figure 3 shows that for Patrick's receptive discrimination program, the average sessions to mastery were 5 for the continuous data collection system and 3 for the

intermittent data collection system. For Patrick's intraverbal program, the average sessions to mastery were 3.25 for the continuous data collection system and 2.75 for the intermittent data collection system. For Patrick's tact program, the average sessions to mastery were 5.5 for the continuous data collection system and 5 for the intermittent data collection system. For Patrick's receptive by function, feature, and classification program, the average sessions to mastery were 3.25 for the continuous data collection system and 2.5 for the intermittent data collection system. As seen in Table 7, exemplars in 8 of Patrick's 16 training sets were mastered more quickly during intermittent data collection, exemplars in 1 of the 16 training sets was mastered more quickly during continuous data collection, and there was no difference with 7 training sets. Likewise, as seen in Table 8, exemplars in 0 of the 16 training sets were better maintained at the 3-week follow-up in the continuous data collection condition, 0 of the 16 training sets were better maintained under intermittent data collection, and exemplars in 16 of the 16 training sets were equally maintained.

Figure 4 shows that for Peter's tact program, the average sessions to mastery were 4 for the continuous data collection system and 3.5 for the intermittent data collection system. For Peter's intraverbal program, the average sessions to mastery were 3.25 for both the continuous and intermittent data collection systems. For Peter's receptive instruction following program, the average sessions to mastery were 4.25 for the continuous data collection system and 4.5 for the intermittent data collection system. For Peter's receptive by function, feature, and classification program, the average sessions to mastery were 3.5 for the continuous data collection system and

2.75 for the intermittent data collection system. As seen in Table 7, exemplars in 7 of Peter's 16 training sets were mastered more quickly during intermittent data collection, exemplars in 3 of the 16 training sets were mastered more quickly during continuous data collection, and there was no difference with 6 training sets. Likewise, as seen in Table 8, exemplars in 2 of the 16 training sets were better maintained at the 3-week follow-up in the continuous data collection condition, 0 training sets were better maintained under intermittent data collection, and exemplars in 14 of the 16 training sets were equally maintained.

As shown in Figure 5, for Mary's motor imitation program, the average sessions to mastery were 5.75 for the continuous data collection system and 5.5 for the intermittent data collection system. For Mary's requesting program, the average sessions to mastery were 4 for both the continuous and intermittent data collection systems. For Mary's tact program, the average sessions to mastery were 4.5 for the continuous data collection system and 3.25 for the intermittent data collection system. As seen in Table 7, exemplars in 9 of Mary's 18 training sets were mastered more quickly during intermittent data collection, exemplars in 4 of the 18 training sets were mastered more quickly during continuous data collection, and there was no difference with 5 training sets. Likewise, as seen in Table 8, exemplars in 10 of the 18 training sets were better maintained at the 3-week follow-up in the continuous data collection condition, 0 training sets were better maintained under intermittent data collection, and exemplars in 8 of the 18 training sets were equally maintained.

As depicted in figure 6, for Allison's receptive discrimination program, the average sessions to mastery were 5 for the continuous data collection system and 4.25 for the intermittent data collection system. For Allison's requesting program, the average sessions to mastery were 3.75 for both the continuous and intermittent data collection systems. For Allison's echoic program, the average sessions to mastery were 6 for the continuous data collection system and 4.5 for the intermittent data collection system. As seen in Table 7, exemplars in 9 of Allison's 18 training sets were mastered more quickly during intermittent data collection, exemplars in 4 of the 18 training sets were mastered more quickly during continuous data collection, and there was no difference with 5 training sets. Likewise, as seen in Table 8, exemplars in 9 of the 18 training sets were better maintained at the 3-week follow-up in the continuous data collection condition, 0 training sets were better maintained under intermittent data collection, and exemplars in 9 of the 18 training sets were equally maintained.

Interestingly, in 16 instances, 5 of the 6 participants (Erin, Patrick, Peter, Mary and Allison) mastered a single skill in slightly fewer sessions during the continuous data collection condition than in the intermittent data collection system. However, as depicted in Table 6, in all of these instances the same skill was mastered in cumulatively fewer minutes during the intermittent data collection condition compared to the continuous data collection condition. As shown in Figure 4 and Table 6, there existed only one incident in which a participant (Peter), mastered a training set (4, receptive discrimination program) more quickly in the continuous data collection

condition (5 sessions, 21.2 min) compared to the intermittent data condition (7 sessions, 22.2 min). However, the difference was negligible.

It should be noted that the minutes to mastery data were virtually identical to the data on number of sessions until mastery. Thus, these data are only depicted in a summary table (Table 6).

Maintenance Evaluation

As seen in Figure 1, for Erin's receptive discrimination program, the average percentage correct score was 90% for the continuous data collection system and 70% for the intermittent data collection system. For Erin's receptive instruction following program, the average percentage correct score was 95% for the continuous data collection system and 70% for the intermittent data collection system.

As shown in Figure 2, for Jeff's tact program, the average percentage correct score was 100% for the continuous data collection system and 95% for the intermittent data collection system. For Jeff's intraverbal program, the average percentage correct score was 100% for both the continuous and intermittent data collection systems. For Jeff's receptive instruction following program, the average percentage correct score was 100% for the continuous data collection system and 95% for the intermittent data collection system. For Jeff's receptive by function, feature, and classification program, the average percentage correct score was 100% for both the continuous and intermittent data collection systems.

Figure 3, shows that for all 4 of Patrick's programs, (receptive discriminations; intraverbals; tacts; receptive by function, feature, and classification), the average

percentage correct score was 100% for both the continuous and intermittent data collection systems.

Figure 4, shows that for Peter's tact program, the average percentage correct score was 100% for the continuous data collection system and 95% for the intermittent data collection system. For Peter's intraverbal program, the average percentage correct score was 100% for both the continuous and intermittent data collection systems. For Peter's receptive instruction following program, the average percentage correct score was 100% for the continuous data collection system and 95% for the intermittent data collection system. For Peter's receptive by function, feature, and classification program, the average percentage correct score was100% for both the continuous and intermittent data collection systems.

As shown in Figure 5, for Mary's motor imitation program, the average percentage correct score was 95% for the continuous data collection system and 87.5% for the intermittent data collection system. For Mary's requesting program, the average percentage correct score was 95% for the continuous data collection system and 90% for the intermittent data collection system. For Mary's tact program, the average percentage correct score was 100% for the continuous data collection system and 90% for the intermittent data collection system.

As depicted in Figure 6, for Allison's receptive discrimination program, the average percentage correct score was 100% for the continuous data collection system and 95% for the intermittent data collection system. For Allison's requesting program, the average percentage correct score was 100% for the continuous data collection

system and 95% for the intermittent data collection system. For Allison's echoic program, the average percentage correct score was 95% for the continuous data collection system and 75% for the intermittent data collection system.

As shown in Figures 1-6, the percentage correct scores were generally the same or better for all participants in the continuous data collection system. More specifically, for 3 (Erin, Mary, Allison) of the 6 participants, each skill mastered in the continuous data collection condition was maintained at a higher percentage correct score compared to the matched skill acquired during the intermittent data collection condition. For 2 (Jeff, Peter) of the 6 participants, 2 of their 4 skill areas mastered under continuous data collection were maintained at a higher percentage correct score than the matched skill acquired during the intermittent data collection condition. For Patrick, all of his skills were equally maintained across conditions (100% correct responding). However, it should be noted that none of the differences in maintenance scores were substantial.

Session Length

As seen in Table 9, the average number of minutes per session for the continuous data collection condition ranged from 3.36 minutes to 4.24 minutes with an average session length of 3.48 minutes. Whereas, the average number of minutes per session for the intermittent data collection condition ranged from 3.18 minutes to 4.18 minutes with an average session length of 3.36 minutes. Overall for all six participants, the continuous data collection condition took longer per session with a difference ranging from 0.06 seconds longer to 0.18 seconds longer per session. Therefore, there

were negligible differences in the amount of time it took to run a session in either condition.

Discussion

The current study was designed to compare continuous (i.e., trial by trial) versus intermittent (i.e., first-trial only) data collection systems in the context of behavioral treatment for children who have PDDs. Comparisons were made across 6 children and 8 different curricular programs. Specific exemplars ranged from simple motor imitation responses (e.g., clapping hands) to answering complex questions such as "Who would use a bandage?" In general, the findings indicate intermittent data collection resulted in slightly more efficient mastery in terms of number of sessions and minutes to mastery. Approximately half (49%) of the participants took longer to master training sets during the continuous data collection condition, whereas, 16% of the participants took longer to master training sets in the intermittent data collection condition and for 35% of the participants there was no difference in either condition. In addition, continuous data collection appeared to result in slightly better maintenance of skills at a 3-week follow-up assessment. Overall, 33% of the training sets maintained better scores at 3-week follow-up probes in the continuous data collection condition, whereas, only 1% of the training sets maintained better scores in the intermittent condition, and 66% of the training sets were equally maintained.

Thus, the data can probably best be interpreted as evidence that both continuous and intermittent data collection appear to similarly impact behavioral instruction. Although the experimental preparation employed in this study is not

representative of all teaching circumstances, the data suggest that collecting data on only the first trial of a session might be a reasonable approach. If the margin of difference between the two data collection systems is as minimal as the data suggest, it may be a reasonable assumption that the intermittent data collection system would be less cumbersome, and more efficient while still accurately portraying data.

The current study does offer some context in which to evaluate some of the claims made regarding continuous and intermittent data collection. One claim was continuous data collection would result in lengthier sessions. The current finding relevant to this claim is that there were no reliable or robust differences in cumulative minutes to mastery between conditions. Another claim is that continuous data collection results in delayed reinforcement (due to therapist data collection). Although the current study does not directly inform this issue, delays in reinforcer delivery presumably would have resulted in slower acquisition under continuous data collection. This did not happen in the current study. However, it is possible that without a well-designed data sheet (see Appendix C) the therapist time that is devoted to continuous data collection might result in lengthier sessions or delayed reinforcement. A third concern is that continuous data collection is more likely to result in false skill mastery. The present data clearly do not indicate this to be the case. Moreover, the data in the current study suggest that intermittent data collection was more likely to produce false skill mastery than continuous data collection. When examining the maintenance probes during the 3-week follow-up sessions, it appears that 3 of the 6 participants (Erin, Mary, Allison) reliably scored lower on the

intermittent data collection probes in *all* of their programs. Likewise, 2 of the 6 participants (Jeff and Peter) scored lower on the intermittent data collection probes in *half* of their programs. It appears that this false-mastery phenomenon would be more likely to occur in receptive programs than during expressive programs due to the higher probability of response selection in receptive tasks. This issue would lend itself well to future research.

An additional caveat for continuous data collection is that it might lead a therapist to teach using massed trials rather than using task interspersal, which has been shown in some cases to be a more effective strategy (Dunlap, 1984). However, the current study showed that continuous and intermittent data collection could both be accomplished using a task interspersal arrangement. It should be noted, that although some of the potential limitations of continuous data collection were not evident in the present study, Sundberg and Hale (2003) appear to have been correct in claiming that the "cold probe" (i.e., first-trial only) method of intermittent data collection is a useful assessment of learner performance.

Even though there was essentially no difference between the two data collection systems in the current study, one could argue that continuous data collection might still offer at least one benefit over intermittent data systems. If data were collected, not only on the learner's response (i.e., correct vs. incorrect), but also on the type of prompt (e.g., gestural vs. physical) required to occasion the response, a clinician could determine whether a learner was making progress by requiring progressively less intrusive prompts, even though percentage correct would still be

poor. Although these kinds of data could still be recorded during intermittent data collection, such patterns are more likely to be evident at the within-session level rather than the between-session level. Thus, continuous data collection (or a less intermittent system than the one evaluated in the current study) would better enable a clinician to evaluate progress across prompt levels. Ultimately, this is an empirical question that could be answered in future studies.

Four strengths of the current study are worth noting. First, because the primary focus of the study was a treatment comparison, treatment integrity was rigorously assessed. Two assessments - one based on direct observation and one on subjective clinical impressions - both showed acceptable and comparable treatment implementation. Second, the comparison was conducted across 8 curriculum programs to ensure that the findings were consistent across different curricular domains rather than specific to a certain type of learning program. The current data do not indicate any program-specific effects. Third, the study was conducted with multiple children, boys and girls, of different ages who were diagnosed with different pervasive developmental disorders. The fact that findings were consistent across such a diverse group of learners speaks to their robustness, at least within this experimental preparation. Finally, a stringent mastery criterion was employed. It is not uncommon for behavioral treatment programs to employ 80% or 90% correct mastery criteria. However, because mastery criteria had to be consistent across conditions, and the intermittent condition could only result in scores of 0% or 100% correct, the latter was chosen as the mastery criterion. Interestingly, this issue is directly related to the notion that continuous data

collection could produce false mastery. The concern is only relevant when massed trials (e.g., 10 consecutive trials of the same exemplar) are conducted with a mastery criterion less than 100%. In these cases, a learner could make an error on the first trial and then simply repeat/echo for all subsequent trials the correct response produced by the therapist during error correction. It should be noted that a stringent mastery criterion (e.g., 100% correct, 90% correct with the first trial of a session correct) adequately addresses this concern, even during massed trials.

The findings of the current study should be viewed in the context of at least two limitations. First, no *long-term* maintenance data were collected. The third dependent measure in the study was how well skills maintained at a 3-week follow-up assessment. It is unknown whether a 3-week follow-up is sufficient to adequately assess maintenance, especially in light of the previously mentioned concerns regarding "false mastery." However, it should be noted that, in all likelihood, false mastery could probably be detected as early as several days after mastery. The second, and perhaps more troublesome, limitation is the possibility of multiple-treatment interference in the data. An alternating treatments design was employed to evaluate the effects of data collection. With this design, experimental control (i.e., internal validity) is demonstrated with a reliable difference between conditions. However, the data in the current study were comparable across conditions. A lack of data separation is often a function of one of the conditions impacting data in the other condition - multiple-treatment interference. On the surface, it might appear that this could have occurred in the current study. However, because the independent variable was the data

collection behavior of a therapist, the basis for multiple-treatment interference would have to be some aspect of one data collection system appearing in sessions for the second system. This is unlikely given the rigorous treatment integrity data that were collected, which indicated that both conditions were implemented as intended. Thus, it is likely that the effects evident in the alternating treatments design are a function of the actual conditions and not multiple-treatment interference.

An interesting phenomenon exists in the data that is worth noting because of its relevance to methodology in this area. It appears from the data that perhaps the participants could be separated into two different groups according to their learning pace. The data indicate that 3 out of the 6 participants demonstrated a quicker learning pace in that they acquired each new exemplar within 2-3 sessions (with a mastery criterion of 2 consecutive session at 100%). The other three participants demonstrated a slower learning pace in that they took up to 16 sessions to acquire each new operant. Interestingly, the participants with a faster learning pace were all male, whereas the participants with a slower learning pace were all female. It should also be noted that girls with the disorder are more likely to be diagnosed at the more severe end of the autistic spectrum (American Psychiatric Association, 2000). Regardless of its relation to participant sex, learning pace can potentially confound an experimental evaluation. By using a research approach in which exemplars within a curricular area are randomly assigned to different experimental conditions, it is necessary for exemplars to be learned at a moderate-to-slow pace such that differences have a time to be produced and detected. The fact that 3 of the participants in the current study acquired

exemplars so quickly might also be considered a limitation of the study. In that case, Erin, Mary, and Allison, who all acquired exemplars more slowly (thus providing time for experimental effects to be detected) might best be considered the "purest" assessment of the study's research question. Fortunately, the current study produced comparable findings across participants.

The current study examined the two most extreme values of data collection: trial-by-trial continuous data collection versus first-trial-only intermittent data collection. Proponents of intermittent data collection systems sometimes employ the first 5 of 10 trials, the first 3 of 10 trials, or even the first and last of 10 trials. The current findings should be reasonably extrapolated to these other "less intermittent" forms of data collection. However, the findings might not apply to other training arrangements. The cold-probe assessment has primarily been recommended in the context of language programs based on Skinner's analysis of verbal behavior (see Sundberg & Partington, 1998). It is common in these programs for different curricular programs (e.g., mands, tacts, intraverbals) to be taught in an interspersal/mixed format such that a learner might be presented with, for example, 8 different programs across 8 trials. In the current study, sessions included only two programs (10 trials of each). It is possible that intermittent and continuous data collection might differentially impact learning with such an incredibly varied interspersal format. For example, if every trial constituted a different program, data collection might indeed interfere with learning. On the other hand, under such an arrangement, collecting data on only one trial might result in a skewed performance evaluation.

This current study provides some initial answers regarding the required frequency of data collection in behavioral treatment programs for children with pervasive developmental disorders. However, further research in this area is clearly warranted. The current study has been conducted with a well-trained 20-year veteran as the experimenter. Given the knowledge and practice of such a senior clinician, it would be important to examine the same research question with less experienced therapists. Therefore, the current study should only be viewed as the beginning of a series of studies. The most valuable next step in this line of research would be to evaluate continuous and intermittent data systems with a more pervasive interspersal system, possibly in a Skinnerian language program and replicate the current study with less-experienced therapists.

Table 1

Seven Common Curriculum Areas

Curriculum Areas	# of Clinical Manuals that Recommend the Area	Stimulus Control of the Response
1. Mand Training	4	Vocal S^D – Vocal or Motor R - possible S^D-R correspondence - specific S^R
2. Motor (Nonvocal) Imitation	4	Vocal & Motor S^D – Motor R - possible S^D-R correspondence - non-specific S^R
3. Echoic Behavior (Vocal Imitation)	4	Vocal S^D – Vocal R - S^D-R correspondence - non-specific S^R
4. Receptive Instruction Following	4	Vocal S^D – Motor R - no S^D-R correspondence - non-specific S^R
5. Tact/Label Training (expressive) - objects, actions, & people	4	Vocal S^D – Vocal R - no S^D-R correspondence - non-specific S^R
6. Tact/Label Training (receptive discriminations) - objects, actions, & people	4	Vocal S^D – Motor R - possible S^D-R correspondence - non-specific S^R
7. Drawing	3	Vocal & Motor S^D – Motor R - possible S^D-R correspondence - non-specific S^R

Table 2

Summary Data for the Curricular Assessment

Participant	Curricular Areas								
	Mand Training	Motor (Nonvocal) Imitation	Echoic Behavior (Vocal Imitation)	Receptive Instruction Following	Tact/Label Training (expressive)	Tact/Label Training (receptive discriminations)	Drawing	Intraverbals	RFFC
Erin	PASS	PASS	PASS	DEFICIT	DEFICIT	DEFICIT	DEFICIT	N/A	N/A
Jeff	PASS	PASS	PASS	DEFICIT	DEFICIT	PASS	PASS	DEFICIT	DEFICIT
Patrick	PASS	PASS	PASS	PASS	DEFICIT	DEFICIT	PASS	DEFICIT	DEFICIT
Peter	PASS	PASS	PASS	DEFICIT	DEFICIT	PASS	PASS	DEFICIT	DEFICIT
Mary	DEFICIT	DEFICIT	PASS	DEFICIT	DEFICIT	PASS	DEFICIT	N/A	N/A
Allison	DEFICIT	PASS	DEFICIT	DEFICIT	DEFICIT	DEFICIT	PASS	N/A	N/A

Table 3

Summary Data for the Behavioral Language Assessment

Participant	Cooperation	Request (Mand)	Motor Imitation	Vocal Play	Vocal Imitation	Match to Sample	Receptive	Label (Tact)	Receptive by FFC	Conversation (Intraverbal)	Letters & Numbers	Social Interaction	Avg.
Erin	4	1	3	2	2	2	2	2	1	1	1	2	1.9
Jeff	4	5	4	5	5	5	5	5	2	3	5	3	4.3
Patrick	5	5	5	5	5	5	5	5	2	3	5	3	4.4
Peter	2	5	5	5	5	5	5	5	2	3	4	3	4.1
Mary	2	3	2	4	3	3	4	3	1	2	1	2	2.5
Allison	5	5	4	3	2	5	5	4	1	1	3	3	3.4

Note: Levels 1 and 5 are indicative of minimal and well-developed verbal repertoires, respectively.

Table 4

Exemplars Taught to Each Participant in Each Program

Program	Participant	Data System	Stimuli 1	2	3	4	5	6	7	8
Tact/Label Training (receptive)	Erin	Continuous	apple	ball	cat	balloon	dog	cookies	bubbles	candy
		Intermittent	baby	banana	horse	puzzle	bird	cake	blocks	chips
	Patrick	Continuous	Michigan	Wisconsin	Illinois	Montana				
		Intermittent	Minnesota	Washington	Mississippi	Louisiana				
	Allison	Continuous	Zebra	Penguin	Octopus	Flamingo	Rhino	Swan		
		Intermittent	Elephant	Dolphin	Buffalo	Owl	Lizard	Turkey		
Receptive Instruction Following	Erin	Continuous	stand up	clap hands	arms up	come here	put on table	tap table	open mouth	nod head yes
		Intermittent	jump	blow kiss	stomp feet	give me five	turn around	wave	cover eyes	shake head no
	Jeff	Continuous	look right	look both ways	stop with sign	green light				
		Intermittent	look left	hold my hand	walk with sign	red light				
	Peter	Continuous	look left	hold my hand	walk with sign	red light				
		Intermittent	look right	look both ways	stop with sign	green light				
Tact/Label Training (expressive)	Jeff	Continuous	Policeman	Barber	Teacher	Dentist				
		Intermittent	Mailman	Librarian	Garbage Man	Fireman				
	Patrick	Continuous	Pharmacist	Mechanic	Policeman	Janitor				
		Intermittent	Paramedic	Firefighter	Soldier	Lifeguard				
	Peter	Continuous	cymbals	maracas	trombone	trumpet				
		Intermittent	violin	xylophone	french horn	clarinet				
	Mary	Continuous	dog	cake	cookies	fries	balloon	juice		
		Intermittent	cat	apple	blocks	van	banana	milk		
Intraverbals	Jeff	Continuous	Doctor	Dentist	Teacher	Bus Driver				
		Intermittent	Mechanic	Policeman	Janitor	Lifeguard				
	Patrick	Continuous	Mailman	Librarian	Garbage Man	Fireman				
		Intermittent	Doctor	Dentist	Teacher	Bus Driver				
	Peter	Continuous	Doctor	Garbage Man	Dentist	Farmer				
		Intermittent	Mailman	Teacher	Librarian	Policeman				
Receptive by Function, Feature and Class	Jeff	Continuous	guitar	flamingo	dolphin	candy				
		Intermittent	banana	cake	apple	dog				
	Patrick	Continuous	balloons	apple	guitar	dog				
		Intermittent	flamingo	dolphin	cake	candy				
	Peter	Continuous	apple	balloons	guitar	puzzle				
		Intermittent	dolphin	flamingo	cake	cookies				
Motor (nonvocal) Imitation	Mary	Continuous	wave	nod head yes	stomp feet	clap hands	touch nose	wiggle fingers		
		Intermittent	rub hands together	shake head no	arms up	touch head	open/close hands	blow kiss		
Requesting (modified Manding)	Mary	Continuous	candy	duck	doll	cookie	marshmallow	music		
		Intermittent	light	car/truck	drink	squishy	book	wand		
	Allison	Continuous	bubbles	music	marshmallow	puzzle	scissors	doll		
		Intermittent	candy	wand	Barney	book	light	ball		
Echoic Behavior (Vocal Imitation)	Allison	Continuous	pop	happy	hi	help	mama	bus		
		Intermittent	hot	baby	bye bye	sit	yes	no		

Table 5

Clinician Ratings of Treatment Integrity

	Question A		B		C		D		E		F	
Raters	Cont.	Int.	Cont.	Int.	Cont.	Int.	Cont.	Int.	Cont.	Int.	Cont.	Int.
	Mean (SD)	X (SD)	X (SD)	X (SD)	X (SD)	X (SD)	X (SD)	X (SD)	X (SD)	X (SD)	X (SD)	X (SD)
1	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)
2	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)
3	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)
4	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)
5	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)
6	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)
7	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)
8	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)
9	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)	3 (0)
Overall	**3 (0)**	**3 (0)**	**3 (0)**	**3 (0)**	**3 (0)**	**3 (0)**	**3 (0)**	**3 (0)**	**3 (0)**	**3 (0)**	**3 (0)**	**3 (0)**

Table 6

Mean Minutes to Mastery for Each Participant

	Erin		Jeff		Pat		Peter		Mary		Allison	
	Cont.	Int.	Cont.	Int.	Cont.	Int.	Cont.	Int.	Cont.	Int.	Cont.	Int.
Program 1	33.17	<u>25.23</u>	19.35	<u>17.27</u>	21.33	<u>13.03</u>	19.37	<u>16.25</u>	23.27	<u>19.11</u>	19.55	<u>15.06</u>
Program 2	23.13	<u>18.44</u>	27.12	<u>24.21</u>	23.51	<u>20.2</u>	<u>21.23</u>	22.18	17.09	<u>16.38</u>	13.4	<u>12.23</u>
Program 3			11.18	<u>7.38</u>	12.57	<u>11.23</u>	13.59	<u>13.58</u>	16.39	<u>11.56</u>	21.18	<u>16.01</u>
Program 4			11.27	<u>9.28</u>	13.19	<u>10.28</u>	14.56	<u>11.11</u>				

Note: Each number that is underlined represents the lower mean of each training set.

Table 7

Number of Exemplars Mastered more Quickly during Continuous or Intermittent Data Collection Conditions

	Continuous	Intermittent	Tied
Erin	11	4	1
Jeff	5	0	11
Patrick	8	1	7
Peter	7	3	6
Mary	9	4	5
Allison	9	4	5
OVERALL	49	16	35

Table 8

Number of Exemplars Maintained more Effectively during Continuous or Intermittent Data Collection Conditions

	Continuous	Intermittent	Tied
Erin	10	1	5
Jeff	2	0	14
Patrick	0	0	16
Peter	2	0	14
Mary	10	0	8
Allison	9	0	9
OVERALL	33	1	66

Table 9

Average Number of Minutes per Session for Each Condition

	Continuous	Intermittent	Difference
Erin	3.48	3.36	0.12
Jeff	4.06	3.54	0.12
Patrick	4	3.48	0.12
Peter	4.24	4.18	0.06
Mary	3.54	3.42	0.12
Allison	3.36	3.18	0.18

Figure 1: Number of sessions to reach the mastery criterion (top panels) and percentage-correct scores during maintenance probes (bottom panel) for Erin across 2 program areas

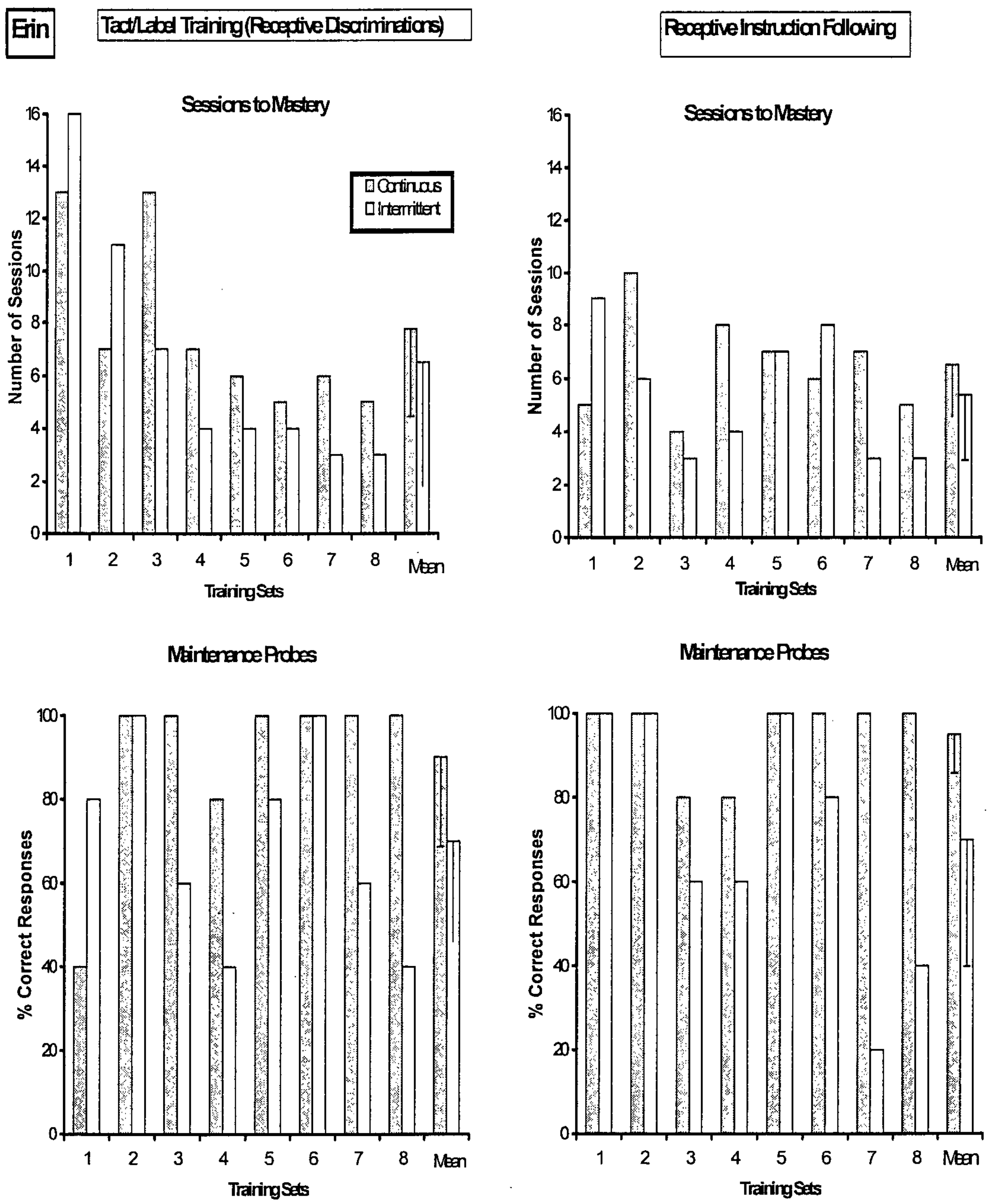


Figure 2: Number of sessions to reach the mastery criterion (top panels) and percentage-correct scores during maintenance probes (bottom panel) for Jeff across 4 program areas

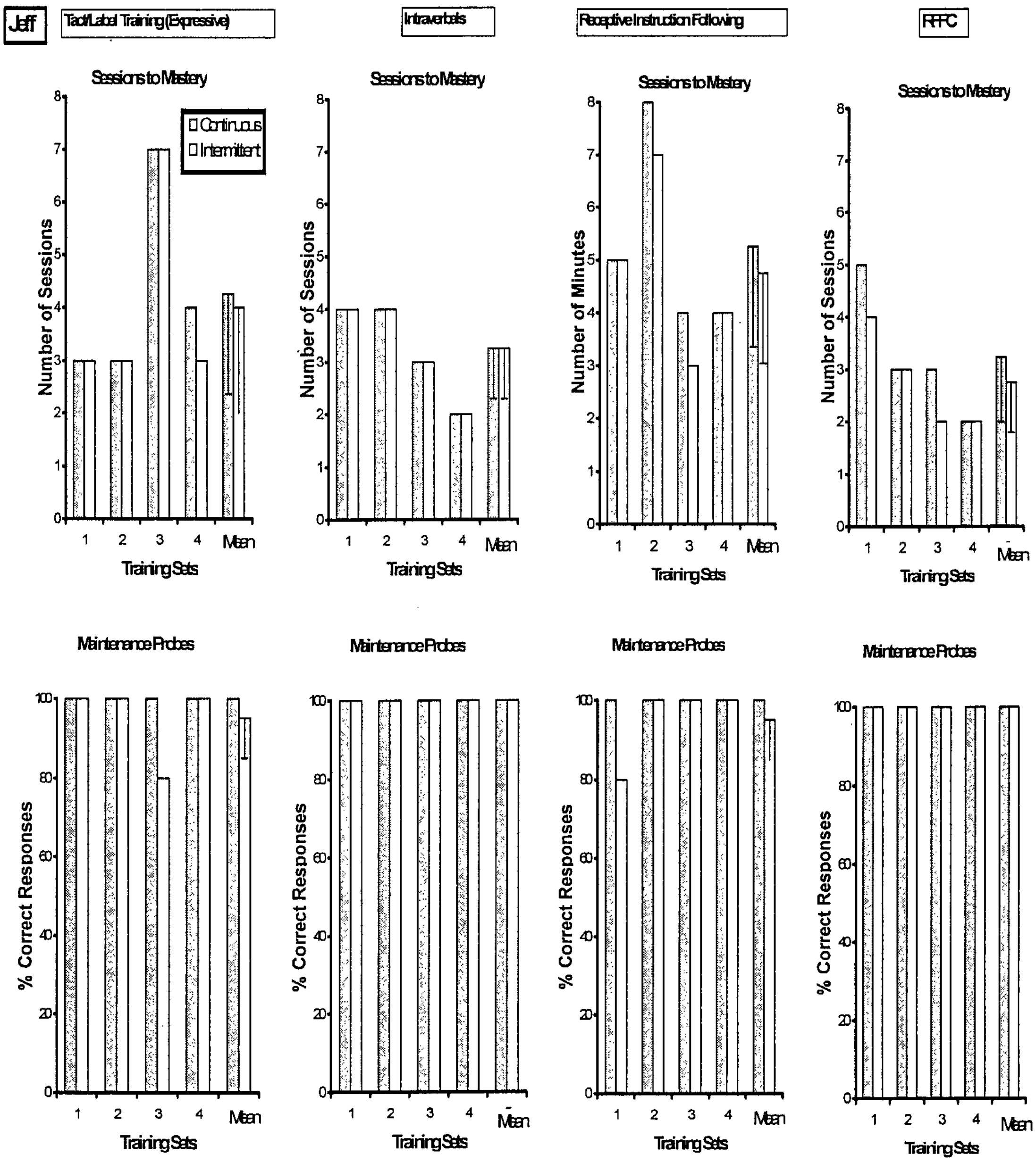


Figure 3: Number of sessions to reach the mastery criterion (top panels) and percentage-correct scores during maintenance probes (bottom panel) for Patrick across 4 program areas

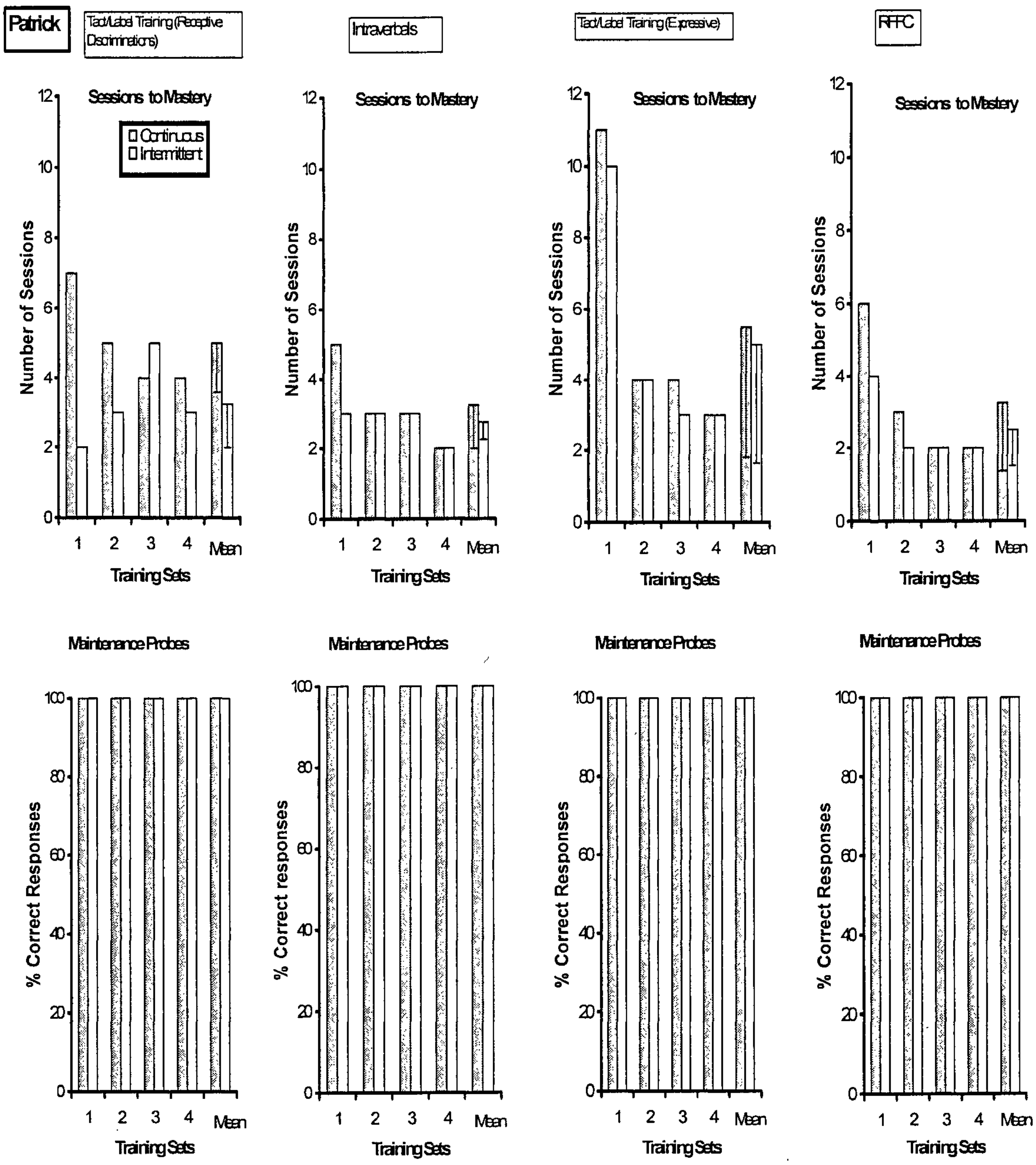


Figure 4: Number of sessions to reach the mastery criterion (top panels) and percentage-correct scores during maintenance probes (bottom panel) for Peter across 4 program areas

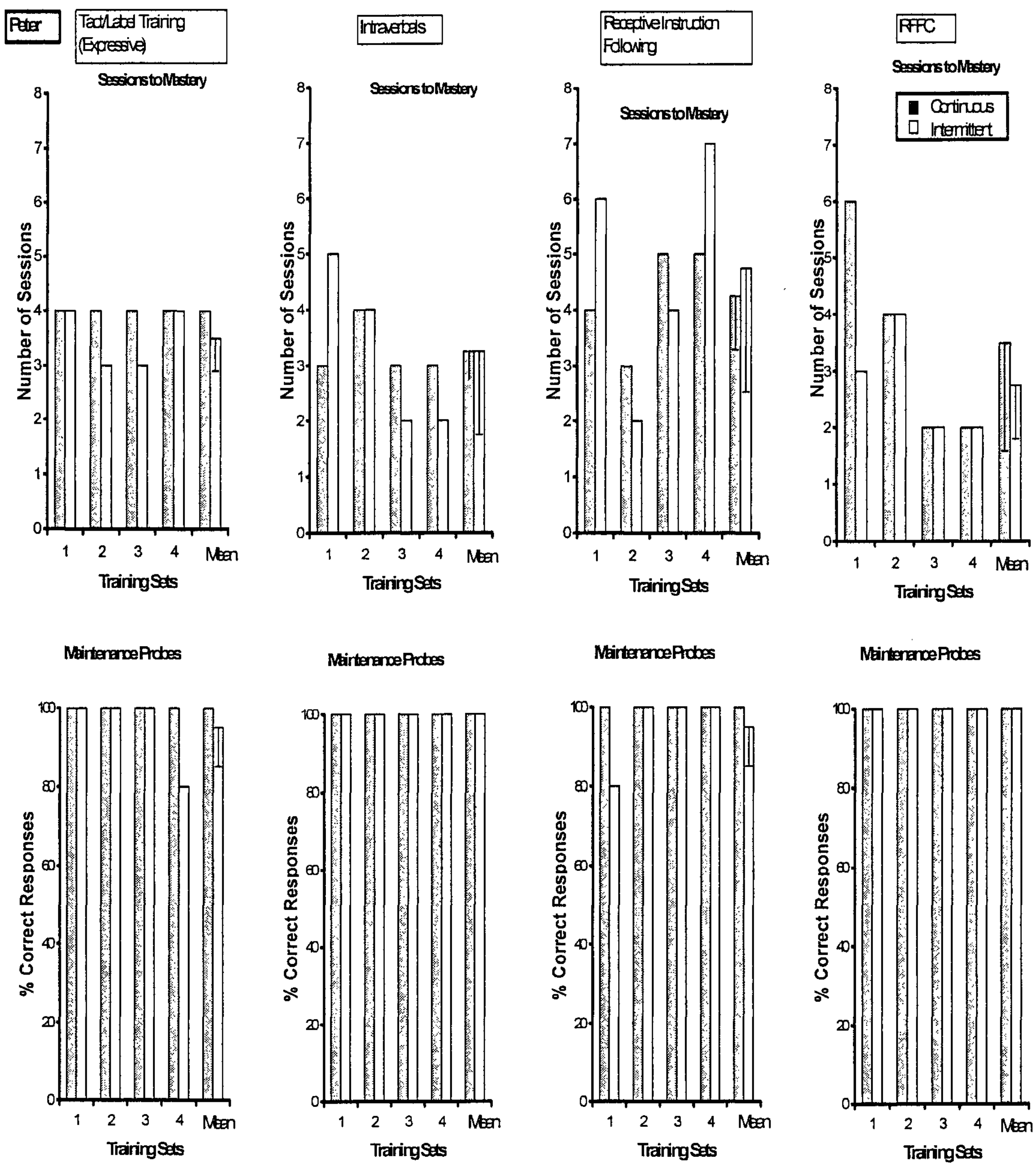


Figure 5: Number of sessions to reach the mastery criterion (top panels) and percentage-correct scores during maintenance probes (bottom panel) for Mary across 3 program areas

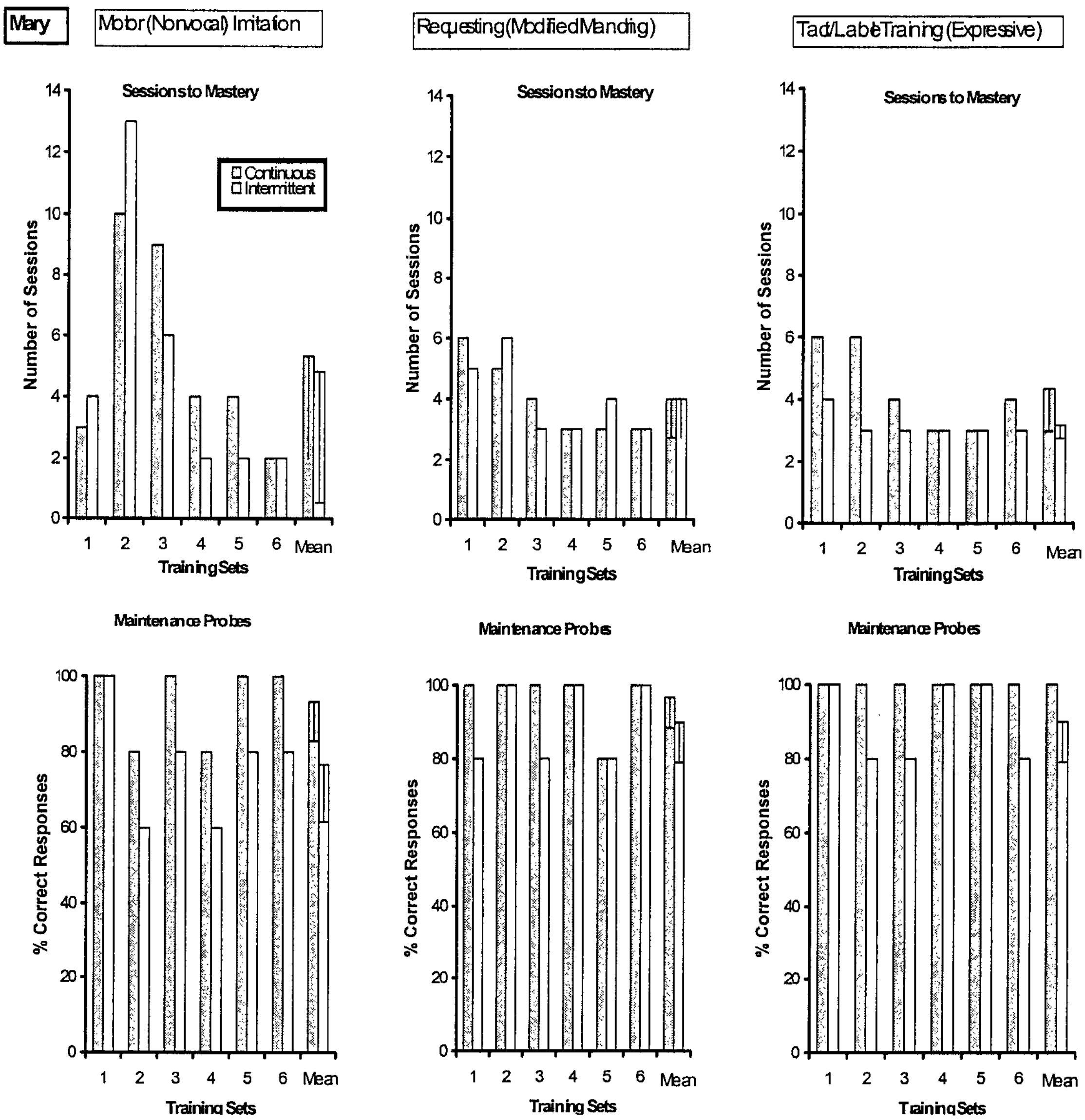
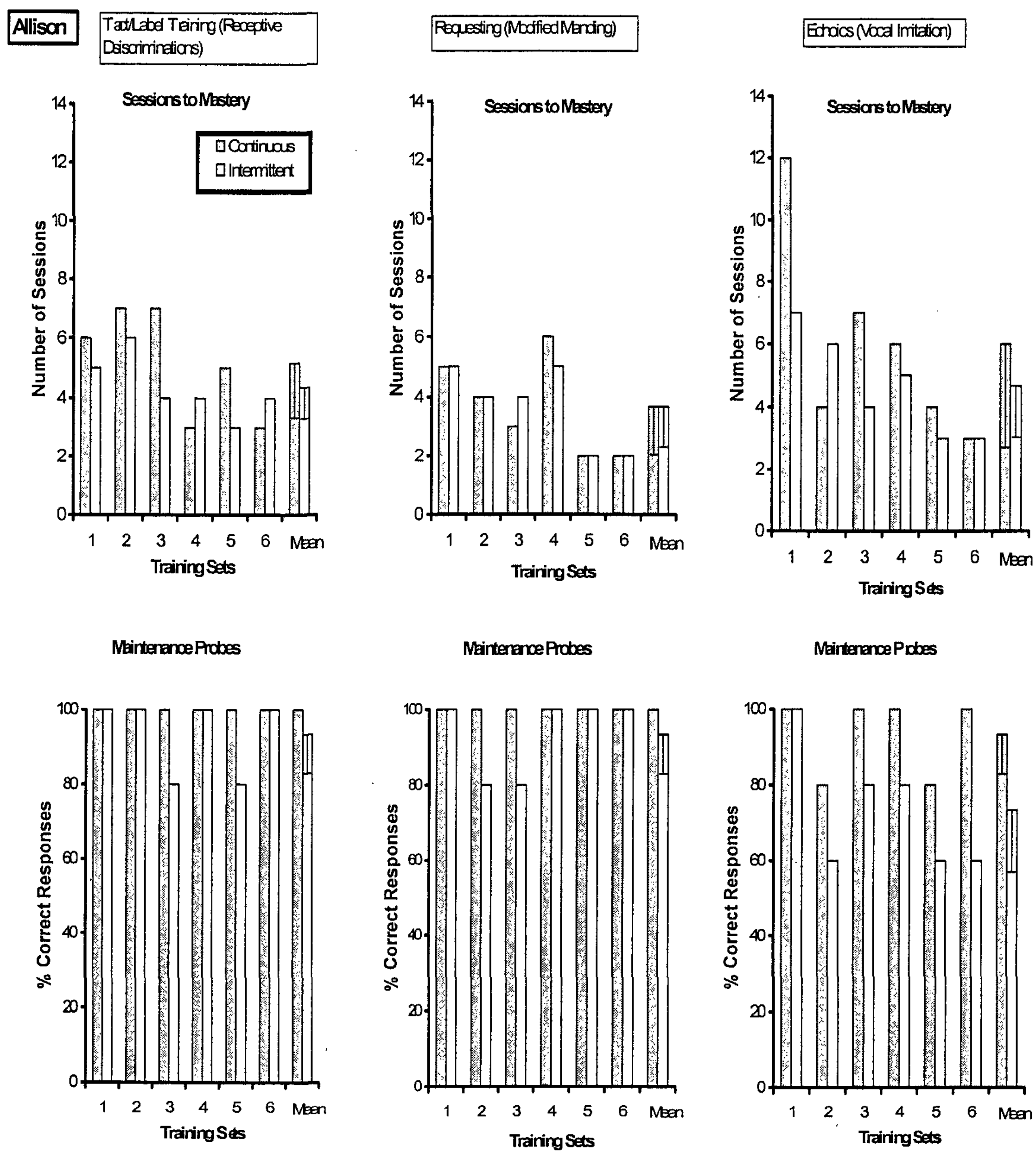


Figure 6: Number of sessions to reach the mastery criterion (top panels) and percentage-correct scores during maintenance probes (bottom panel) for Allison across 3 program areas



Appendix A

Acquisition Programs

The *tact/label training (receptive)* program involved laying out three pictures on a Velcro strip attached to the front of a binder. All pictures used in the study were actual photographs with the exception of the state maps used with Patrick, which were three-dimensional gray-scale drawings. All pictures were placed inside a plastic baseball cardholder. Velcro was attached to the backs of the cardholders for easy and secure placement to the top of an 8 x 11 inch, 3-ring binder that contained three strips of Velcro on the cover. The binder was placed closed on the table so that it lay at an angle for the participant to view easily. When the three pictures were secured to the front of the binder, the experimenter presented the binder to the participant and clearly delivered the vocal S^D "Show/give me _____." The participant was taught to select the picture by removing it from the binder and handing it to the experimenter. This program was implemented for Erin, Patrick, and Allison for their acquisition programs. The specific stimuli used for Erin were apple, ball, cat, balloon, dog, cookies, bubbles and candy in the continuous data collection condition; and baby, banana, horse, puzzle, bird, cake, blocks and chips in the intermittent data collection condition. The specific stimuli used for Patrick were Michigan, Wisconsin, Illinois, and Montana in the continuous data collection condition; and Minnesota, Washington, Mississippi, and Louisiana in the intermittent data collection condition. This program was also implemented in the same manner for Mary as an interspersal program.

The *receptive instruction following* program involved the experimenter clearly delivering the vocal S^D, which named the action for the participant to do (e.g., "stand up" or "look to the right"). The participant was taught to follow the action requested by the experimenter. This program was implemented for Erin, Jeff, and Peter for their acquisition programs. The specific stimuli used for Erin were stand up, clap your hands, arms up, come here, put on table, tap table, open mouth, and nod head "yes" in the continuous data collection condition; and jump, blow kiss, stomp feet, give me five, turn around, wave, cover eyes, and shake head "no" in the intermittent data collection condition. The specific stimuli used for Jeff were look right, look both ways, stop at the stop sign and go at the green light in the continuous data collection condition; and look left, hold my hand, walk with the sign, and stop at the red light in the intermittent data collection condition. The specific stimuli used for Peter were and look left, hold my hand, walk with the sign, and stop at the red light in the continuous data collection condition; and look right, look both ways, stop at the stop sign, and go at the green light in the intermittent data collection condition.

The *tact/label training (expressive)* program involved the experimenter holding up one picture and presenting the vocal S^D, "What/who is it?." All pictures used in the study were actual photographs and were placed inside a plastic baseball cardholder. The participant was taught to respond vocally by naming the person or object. This program was implemented for Jeff, Patrick, Peter, and Mary for their acquisition programs. The specific stimuli used for Jeff were policeman, barber, teacher, and dentist in the continuous data collection condition; and mailman, librarian, garbage

man, and fireman in the intermittent data collection condition. The specific stimuli used for Patrick were pharmacist, mechanic, policeman, and janitor in the continuous data collection condition; and paramedic, firefighter, soldier, and lifeguard in the intermittent data collection condition. The specific stimuli used for Peter were cymbals, maracas, trombone, and trumpet in the continuous data collection condition; and violin, xylophone, french horn, and clarinet in the intermittent data collection condition. The specific stimuli used for Mary were dog, cake, cookies, french fries, balloon, and juice in the continuous data collection condition; and cat, apple, blocks, van, banana, and milk in the intermittent data collection condition.

The *intraverbal* program involved the experimenter clearly delivering the vocal S^D "Who would use ___" and listing or showing a picture of a series of stimuli such as a stethoscope, a bandage, a needle, and a tongue depressor. The participant was taught to answer the vocation of the person, such as "a doctor." This program was implemented for Jeff, Patrick, and Peter for their acquisition programs. The specific stimuli used for Jeff were doctor, dentist, teacher and bus driver in the continuous data collection condition; and mechanic, policeman, janitor and lifeguard in the intermittent data collection condition. The specific stimuli used for Patrick were mailman, librarian, garbage man and firefighter in the continuous data collection condition; and doctor, dentist, teacher and bus driver in the intermittent data collection condition. The specific stimuli used for Peter were doctor, garbage man, dentist and farmer in the continuous data collection condition; and mailman, teacher, librarian and policeman in the intermittent data collection condition.

The *receptive by function, feature, and class* program involved the experimenter presenting three photographs using the materials described in the *receptive discrimination* program. When the three pictures were secured to the front of the binder, the experimenter presented the binder to the participant and clearly delivered the vocal S^D "Show/give me the one that ___," (specifying a function, feature, or classification of one of the stimuli). The participant was taught to select the picture by removing it from the binder and handing it to the experimenter. This program was implemented for Jeff, Patrick, and Peter for their acquisition programs. The specific stimuli used for Jeff were guitar, flamingo, dolphin and candy in the continuous data collection condition; and banana, cake, apple and dog in the intermittent data collection condition. The specific stimuli used for Patrick were balloons, apple, guitar and dog in the continuous data collection condition; and flamingo, dolphin, cake and candy in the intermittent data collection condition. The specific stimuli used for Peter were apple, balloons, guitar and puzzle in the continuous data collection condition; and dolphin, flamingo, cake and cookies in the intermittent data collection condition.

The *motor (nonvocal) imitation* program involved the experimenter clearly delivering the vocal S^D "Do this" as she simultaneously modeled a motor response for the participant. The participant was taught to perform the same motor response that was modeled by the experimenter. This program was implemented for Mary for her acquisition program. The specific stimuli used for Mary were wave, nod head "yes", stomp feet, clasp hands, touch nose and wiggle fingers in the continuous data

collection condition; and rub hands together, shake head "yes", arms up, touch head, open/close hands and blow kiss in the intermittent data collection condition. This program was also implemented in the same manner for Allison as an interspersal program.

The *requesting (modified manding)* program involved the experimenter holding up one picture or one object and asking the participant the vocal S^D "What do you want?" All pictures used in the study were actual photographs and were placed inside a plastic baseball cardholder. The participant was taught to respond by saying or signing, "I want ___" (either by naming the object vocally or with a sign). This program was implemented for Mary and Allison for their acquisition programs. The specific stimuli used for Mary were candy, duck, doll, cookie, marshmallow and music in the continuous data collection condition; and light, car/truck, drink, squishy, book and wand in the intermittent data collection condition. The specific stimuli used for Allison were bubbles, music, marshmallow, puzzle, scissors and doll in the continuous data collection condition; and candy, wand, Barney, book, light and ball in the intermittent data collection condition.

The *echoic behavior (vocal imitation)* program involved the experimenter clearly delivering the vocal S^D "Say ___" as she simultaneously modeled a vocal word for the participant. The participant was taught to perform the same vocal response that was modeled by the experimenter. This program was implemented for Allison for her acquisition program. The specific stimuli used for Allison were pop, happy, hi, help, Mama and bus in the continuous data collection condition; and hot, baby, bye bye, sit, yes and no in the intermittent data collection condition. This program was also implemented in the same manner for Erin as an interspersal program.

Interspersal Programs

In conjunction with the 3 programs mentioned above as interspersal programs (tact/label training [receptive], motor [nonvocal] imitation, and echoic behavior [vocal imitation] program) two additional programs were implemented as interspersal programs. The *drawing* program involved the experimenter clearly delivering the vocal S^D "draw a______/make a _____," which named the stimuli for the participant to draw/spell (e.g., "draw a D," "make an E," "draw "mommy" or "make a house"). The participant was then expected to draw/spell the letter, word or picture of the stimuli requested by the experimenter. This program was implemented for Jeff and Patrick for their interspersal program.

The *answering social questions* program involved the experimenter clearly delivering the vocal S^D which named the social question to be answered by the participant (e.g., "when is your birthday?," "who makes dinner in your house?," "where do you go to school?" or "what is your brother's name?"). The participant was then expected to answer the specific social question posed by the experimenter. This program was implemented for Peter for his interspersal program.

Appendix B

Curriculum Assessment Data Sheet

Participant: ____________ **Date:** ___________ **IOA: YES NO**
Trainer's Name: ____________ **Session #:** ________ **Primary/Secondary:** ________

Program:	Manding		
Trial:	Stimuli:	Score:	Prompt:
1			
2			
3			
4			
5			
TOTAL:			

Program:	Motor Imitation		
Trial:	Stimuli:	Score:	Prompt:
1			
2			
3			
4			
5			
TOTAL:			

Program:	Echoic		
Trial:	Stimuli:	Score:	Prompt:
1			
2			
3			
4			
5			
TOTAL:			

Program:	Rec Instruction Following		
Trial:	Stimuli:	Score:	Prompt:
1			
2			
3			
4			
5			
TOTAL:			

Program:	Tact - Expressive		
Trial:	Stimuli:	Score:	Prompt:
1			
2			
3			
4			
5			
TOTAL:			

Program:	Tact - Receptive		
Trial:	Stimuli:	Score:	Prompt:
1			
2			
3			
4			
5			
TOTAL:			

Program:	Drawing		
Trial:	Stimuli:	Score:	Prompt:
1			
2			
3			
4			
5			
TOTAL:			

NOTES:

Appendix C

Data Sheet

Participant: _______ Date: ______ IOA: YES NO
Trainer's Name: ___________ Session #: _____ Primary/Secondary: ______
Total Session Time: ________ Pref. Ass't Item (1) _____ (2) _____ (3) ______

	Instructional PRESENT STIMULUS	Sequence SPECIFIC INSTRUCTION	Participant's SCORE + − NR	Response PROMPT LEVEL V G/M HOH	IOA
1			+ − NR	V G/M HOH	A D
2			+ − NR	V G/M HOH	A D
3			+ − NR	V G/M HOH	A D
4			+ − NR	V G/M HOH	A D
5			+ − NR	V G/M HOH	A D
6			+ − NR	V G/M HOH	A D
7			+ − NR	V G/M HOH	A D
8			+ − NR	V G/M HOH	A D
9			+ − NR	V G/M HOH	A D
10			+ − NR	V G/M HOH	A D

IOA = A / (A+D) X 100 = ☐

TX Integrity

STIM CORR	INSTR CORR	CONSQ CORR	DATA CORR	ALL CORR	IOA
Y N	Y N	Y N	Y N	Y N	A D
Y N	Y N	Y N	Y N	Y N	A D
Y N	Y N	Y N	Y N	Y N	A D
Y N	Y N	Y N	Y N	Y N	A D
Y N	Y N	Y N	Y N	Y N	A D
Y N	Y N	Y N	Y N	Y N	A D
Y N	Y N	Y N	Y N	Y N	A D
Y N	Y N	Y N	Y N	Y N	A D
Y N	Y N	Y N	Y N	Y N	A D
Y N	Y N	Y N	Y N	Y N	A D
			%correctly implem trials		

IOA = A / (A+D) X 100 = ☐

Participant: _______ Date: ______ IOA: YES NO
Trainer's Name: ___________ Session #: _____ Primary/Secondary: ______
Total Session Time: ________ Pref. Ass't Item (1) _____ (2) _____ (3) ______

	Instructional PRESENT STIMULUS	Sequence SPECIFIC INSTRUCTION	Participant's SCORE + − NR	Response PROMPT LEVEL V G/M HOH	IOA
1			+ − NR	V G/M HOH	A D

TX Integrity

STIM CORR	INSTR CORR	CONSQ CORR	DATA CORR	ALL CORR	IOA
Y N	Y N	Y N	Y N	Y N	A D
			%correctly implem trials		

IOA = A / (A+D) X 100 = ☐

IOA = A / (A+D) X 100 = ☐

Appendix D

ABA Teacher Evaluation Form for Treatment Integrity

ABA Teacher	**Evaluator**
Date	**Time**

KEY
1 = Did not happen at all
2 = inconsistent performance
3 = fully consistent performance each trial

A. All instructions & S^Ds given clearly with proper emphasis

1 2 3

B. Uses different vocal tone for instruction vs. reinforcer

1 2 3

C. Program pacing is appropriate for participant

1 2 3

D. Gives consequences immediately

1 2 3

E. Provides differential reinforcement for prompted trials or demands

1 2 3

F. Controls disruptive behavior appropriately & effectively

1 2 3

Appendix E

HSIRB Approval Letter

Date: July 29, 2003

To: James Carr, Principal Investigator
Anne Cummings, Student Investigator for dissertation

From: Mary Lagerwey, Chair

Re: HSIRB Project Number 03-07-02

This letter will serve as confirmation that your research project entitled "Evaluating Progress in Intensive Intervention Programs for Children with Developmental Disabilities: Continuous versus Intermittent Data Collection Systems" has been **approved** under the **full** category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

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

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

Approval Termination: July 16, 2004

References

Allyon, T., & Michael, J. (1959). The psychiatric nurse a behavioral engineer. *Journal of the Experimental Analysis of Behavior, 2,* 323-334.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed.). Text Revision.* Washington, DC.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91-97.

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change (2nd ed).* Needham Heights, MA: Allyn & Bacon.

Bijou, S. W., & Baer, D. M. (1961). *Child development: A systematic and empirical theory, Vol. 1.* New York: Appleton-Century-Crofts.

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.

Bijou, S. W., Peterson, R. F., Harris, F. R., Allen, K. E., & Johnston, M. C. (1969). Methodology for experimental studies of young children in natural settings. *The Psychological Record,*

19, 177-210.

Carr, J.E., Nicolson, A. C., & Higbee, T. S. (2000). Evaluation of a brief multiple-stimulus assessment in a naturalistic context. *Journal of Applied Behavior Analysis, 33,* 353-357.

Cooper, J. O., Heron, T. E., & Heward, W. L. (1987) *Applied behavior analysis.* Columbus, OH: Merrill.

DeMeyer, M. K., Hingtgen, J. N., & Jackson, R. K. (1981). Infantile autism reviewed: A decade of research. *Schizophrenia Bulletin, 7*, 388-451.

Dollins, P., & Carbone, V. J. (2003, May). Using probe data recording methods to assess learner acquisition of skills. In V. J. Carbone (Chair), *Research related to Skinner's analysis of verbal behavior with children with autism.* Symposium conducted at the 29th Annual Convention of the Association for Behavior Analysis, San Francisco.

Dunlap, G. (1984). The influence of task variation and maintenance tasks on the learning and affect of autistic children. *Journal of Experimental Child Psychology, 37,* 41-64.

Dunlap, G., Koegel, R. L., & O'Neill, R. E. (1985). Pervasive developmental disorders. In P. H. Bornstein & A. E. Kazdin (Eds.), *Handbook of clinical behavior therapy with children* (pp. 499-540). Homewood, IL: Dorsey Press.

Ferster, C. B. (1961). Positive reinforcement and behavioral deficits of children with autism. *Child Development, 32*, 437-456.

Ferster, C. B., & DeMeyer, M. K. (1961). The development of performance in autistic children in an automatically controlled environment. *Journal of Chronic Diseases, 13*, 312-345.

Fisher, W. W., Piazza, C. C., Bowman, L. G., & Amari, A. (1996). Integrating caregiver report with a systematic choice assessment to enhance reinforcer identification. *American Journal on Mental Retardation*, *101*, 15-25.

Harris, S. L., & Handleman, J. S. (1994). *Preschool education programs for children with autism.* Austin. TX: PRO-ED.

Kanner, L. (1943). Autistic disturbances in affective contact. *Nervous Child, 2*, 217-250.

Koegel, R. L., & Koegel, L. K. (1995). *Teaching children with autism.* Baltimore, MD: Paul H. Brookes.

Leaf, R., & McEachin, J. (1999). *A work in progress: Behavior management strategies and a curriculum for intensive behavioral treatment of autism.* New York: DRL Books.

Lovaas, O. I. (1966). A program for the establishment of speech in psychotic children. In J. K. Wing (Ed.), *Early childhood autism* (pp. 74-87). Oxford: Pergamon.

Lovaas, O. I. (1977). *The autistic child.* New York: Irvington.

Lovaas, O. I. (1979). Contrasting illness and behavioral models for the treatment of autistic children: A historical perspective. *Journal of Autism and Developmental Disorders, 9*, 315-323.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*, 3-9.

Lovaas, O. I. (2003). *Teaching individuals with developmental delays: Basic intervention techniques.* Austin, TX: PRO-ED.

Lovaas, O. I., & Smith, T. (1989). A comprehensive behavioral theory of autistic children: Paradigm for research and treatment. *Journal of Behavior Therapy and Experimental Psychiatry,* 20, 17-29.

Maurice, C., Green, G., & Luce, S. C. (1996). *Behavioral intervention for young children with autism: A manual for parents and professionals.* Austin, TX: PRO-ED.

McEachin, J. J., Smith, T., & Lovaas, O. I. (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal on Mental Retardation, 97*, 359-372.

Munger, G. F., Snell, M. E., & Loyd, B. H. (1989). A study of the effects of frequency of probe data collection and graph characteristics on teachers' visual analysis. *Research in Developmental Disabilities, 10*, 109-127.

Rutter, M. (1974). The development of infantile autism. *Psychological Medicine, 4,* 147-163.

Schreibman, L. (1988). *Autism.* Newbury Park CA: Sage.

Siegel, B. (1996). *The world of the autistic child. Understanding and treating autistic spectrum disorders.* New York: Oxford University Press.

Smith, T. (2001). Discrete trial training in the treatment of autism. *Focus on Autism and Other Developmental Disabilities, 16*, 86-92.

Sundberg, M. L., & Hale, L. (2003, May). Using textual stimuli to teach vocal-intraverbal behaviors. In A. I. Petursdottir (Chair), *Methods for teaching intraverbal behavior to children.* Symposium conducted at the 29th Annual Convention of the Association for Behavior Analysis, San Francisco.

Sundberg, M. L., & Partington, J. W. (1998). *Teaching language to children with autism or other developmental disabilities.* Pleasant Hill, CA: Behavior Analysts, Inc.

Van Houten, R. (1990). Emotional problems II: Autism. In J. M. Matson (Ed.), *Handbook of behavior modification with the mentally retarded* (2nd ed.; pp. 421-441). New York: Plenum.

Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T., & Klin, A. (2004). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, 45,* 135-170.

Waterhouse, L., Fein, D., Nath, J., & Snyder, D. (1987). Pervasive developmental disorders and schizophrenia occurring in childhood: A review of critical commentary. In G. L. Tischler (Ed.), *Diagnosis and classification in psychiatry: A critical appraisal of DSM-III* (pp. 335-368). New York: Cambridge University Press.

Wing, J. K. (1966). *Early childhood autism.* Elmsford, NY: Pergamon.