An Evaluation of Parent Training Methods to Increase Treatment Integrity

Jennifer Bellotti

A Dissertation Submitted to the Faculty of

The Chicago School of Professional Psychology

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in Psychology

August 24, 2018

An Evaluation of Parent Training Methods to Increase Treatment Integrity

---

A Dissertation Submitted to the Faculty of

The Chicago School of Professional Psychology

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in Psychology

Jennifer Bellotti

2018

Approved By:

Julie A. Ackerlund Brandt, PhD, BCBA-D, Chairperson
Assistant Professor, Applied Behavior Analysis Online Program

---

Annette Griffith, PhD, BCBA-D, Member
Assistant Professor, Applied Behavior Analysis Online Program

---

Ariana Boutain, PhD, BCBA-D, Member
Director of Therapy and Clinical Services, KGH Consultation and Treatment, Inc.

Acknowledgements

I would like to acknowledge my advisor, Dr. Julie A. Ackerlund Brandt, for her contribution to the development and refinement of this research project. I would also like to thank my committee members, Dr. Annette Griffith and Dr. Ariana Boutain, for their support and contributions to this project. I would also like to acknowledge my fellow graduate students for their contributions to the conceptualization of this project.

Abstract

Treatment integrity is the accurate implementation of an intervention as written and is an area of upmost importance for applied research and professional practice. Treatment integrity allows a researcher or clinician to determine that the intervention was correctly implemented throughout the study or treatment. However, despite its importance, a limited amount of research has been conducted in this area, particularly for parents as implementation agents. As behavior analytic services provided in the home become more common, it is important to identify strategies to increase treatment integrity for this population. It is important to obtain high levels of treatment integrity for parents and to identify factors that may impede accurate implementation of interventions. Behavioral skills training has been found to be effective as a package to increase treatment integrity for staff; however, research has not been conducted for parent implementation. The purposes of the current study are to (a) examine the effects of behavioral skills training as a method for increasing treatment integrity of parents' implementation of behavioral interventions for their children, and to (b) assess any correlation between parent treatment integrity and child skill acquisition.

Table of Contents

Chapter 1: Nature of the Study

Background

Treatment integrity is often defined as the accurate implementation of an intervention and is frequently also called implementation integrity, procedural fidelity, procedural integrity, procedural reliability, treatment adherence, treatment implementation, treatment delivery, and intervention adherence (Barnett et al., 2014; DiGennaro-Reed & Codding, 2013). Interest in treatment integrity is growing rapidly in various fields, such as behavior analysis, education, psychology and public health (DiGennaro-Reed & Codding, 2013). High levels of treatment integrity have been repeatedly associated with more substantial behavioral changes and more effective and efficient interventions (Carroll, Kodak, & Fisher, 2013; Craig, 2010; DiGennaro, Martens, & Kleinmann, 2007; Gross, Duhon, Doerksen-Kloop, 2014; Mouzakitis, Codding, & Tyron, 2015; Noell, Gresham, & Gansle, 2002; Plavnick, Ferreri, & Maupin, 2010; St. Peter Pipkin, Vollmer, & Sloman, 2010).

Human-service research is carefully conducted to assess the impact of services and to create effective programs; however, programs can only be effective if implemented with high integrity. Dissemination is critical in translating well-researched procedures and programs into common, effective practices. Once effective treatment procedures have been identified and tested, the task of delivering those treatment procedures to clients consistently across settings is paramount (Fixsen & Blase, 1993). It is important for teachers and clinicians, as well as parents and other family members, to implement procedures consistently with integrity. If the procedures are only implemented effectively in one setting, generalization of behavior change will not occur or be as robust as possible. Additionally, it can be difficult to determine why behavior changes are not being made if interventions are not followed accurately.

Based on their review of school-based experimental studies published in the *Journal of Applied Behavior Analysis* (*JABA*), McIntyre et al. (2007) made several recommendations to ensure accurate treatment: (a) including treatment integrity as an assessment of independent variable implementation on a more frequent basis, and (b) remaining alert to the fact that various treatments may require different levels of treatment integrity for desired behavior change. Although additional research is needed to determine ideal levels, studies often report a "high level" to be approximately 90%. McIntyre et al. also recommended that researchers conduct regular documentation of treatment integrity for individual intervention components. By examining individual components, the specific component of an intervention and results can be assessed so that the researcher can determine what specifically led to behavior change, or the lack of, behavior change. This is also important for replication for both researchers and practitioners.

McIntyre et al. (2007) recommended regular documentation of treatment integrity within clinical settings to ensure that problems can be caught and addressed early. Consequences should then be programmed based on treatment integrity levels. For example, providing corrective feedback and additional training for staff who demonstrate low levels of treatment integrity and providing praise or rewards with positive feedback for high levels of treatment integrity. Training requirements can then be reduced when high levels of treatment integrity are then achieved.

Additionally, initial training for staff (or any other implementation agent) should be provided as soon as the intervention begins for the client, and this training should be conducted until the participant meets a preset criterion (McIntyre et al., 2007). Basing training on a set criterion, rather than ending training after a specific period of time (i.e., two sessions, 5 days, or

1 week) has passed, is beneficial because high levels of integrity can be ensured (McIntyre et al., 2007). It is important to note that the amount of training required is likely to vary based on individual factors (McIntyre et al., 2007). Intervention protocols such as procedural checklists should be provided to individuals required to implement interventions and individuals should be trained how to implement interventions and how to collect treatment-integrity data (McIntyre et al., 2007). Last, a sample of treatment-integrity assessment should be collected on all interventions considered to be high-risk for inaccurate implementation, such as complicated treatment programs for high-intensity problem behavior (McIntyre et al., 2007).

Hagermoser-Sanetti, Dobey, and Gritter (2012) also conducted a review of literature on treatment-integrity research published in the *Journal of Positive Behavior Interventions* from 1999 to 2009 and summarized trends and characteristics similar to McIntyre et al. (2007). They culled 72 articles that met the criteria to be included for their review. Results were that 86% of articles included an operational definition and technological description of the independent variable, or a reference to a more detailed source, but less than half the studies reported quantitative treatment-integrity data on the implementation of their independent variables. These results are worrisome because the true effects of an independent variable cannot be determined without knowing the integrity with which it was implemented. There was also a decreasing trend over time of providing an operational definition of the independent variable and an increasing trend of providing a reference to a more detailed source for the operational definition of the independent variable. Without an operational definition of the independent variable and quantitative treatment-integrity data, it is difficult to assess the validity of a study and replicate findings moving forward. Additionally, there was a variable trend of reporting quantitative treatment-integrity data across the years studied with the data increasing from 2003 to 2006,

decreasing in 2007 and 2008, and increasing again in 2009. These results further support the need for a greater amount of treatment integrity research because without knowing the degree to which an independent variable was implemented as intended, readers cannot make an informed decision regarding whether a functional relation exists between variables and therefore cannot with sound judgment replicate a study in clinical practice.

As demonstrated in the reviews discussed, there continues to be limited research on treatment integrity and a lack of quantitative treatment-integrity data. This is detrimental due to the importance of treatment-integrity measurement for quality, internal validity, external validity, replication, and assessment of effectiveness and efficiency (Hagermoser-Sanetti et al., 2012). However, some federal regulations have recently been introduced to increase measurement and reporting of treatment integrity for research and grants (DiGennaro-Reed et al., 2013; McIntyre et al., 2007). For example, the American Psychological Association has called for assessment and monitoring of treatment integrity, and the United States Department of Education's Institute of Education Sciences has called for researchers who submit single-case experimental design grant applications to describe how treatment integrity will be measured, the frequency of assessments, and how much variation in integrity will be accepted over the course of a study (McIntyre et al., 2007). Other programs that have also introduced treatment-integrity standards more recently include the Center for the Study of Prevention of Violence's Blueprints for Healthy Youth Development Program, the Task Force on Evidence-Based Interventions on School Psychology, and the Substance Abuse and Mental Health Services Administration's National Registry of Evidence-Based Programs and Practices (DiGennaro-Reed et al., 2013). The goal is that editorial and grant boards will increase support for inclusion of treatment-

integrity measurement in research publications and grant applications (DiGennaro-Reed et al., 2013).

Problem Statement

Implementing behavioral programming with high levels of treatment integrity lead to more substantial behavioral changes and more effective and efficient interventions. Therefore, it is important for implementation agents (including teachers, therapists, and parents) to deliver treatment procedures accurately and as intended and for supervisors to regularly document treatment integrity. By doing so, behavior analysts can determine if the intervention or components of the intervention led to behavior change or the lack of behavior change, or if behavior changes are due to some external variable. Additionally, problems can be caught and addressed early. High levels of treatment integrity and reporting of these data are also important in research so that the reader can ascertain if a functional relation was obtained and for replication purposes. Despite its importance, there continues to be a lack of research on treatment integrity.

Purpose of the Study

The purposes of this study are to (a) examine the effect of behavioral skills training (BST) as a method for increasing treatment integrity of parent's implementation of behavioral interventions for their children, and (b) to assess any correlation between parent treatment integrity and child skill acquisition. BST has been found to be effective as a package to increase treatment integrity for staff; however, research has not been conducted for parent implementation. As behavior analytic services continue to be implemented in the home setting, it will be necessary to train parents effectively how to implement interventions with integrity for their children.

Research Questions and Hypotheses

Research Question1: Is BST effective as a method for increasing treatment integrity of parent's implementation of behavioral interventions for their children?

H_1: BST increases parents' treatment integrity of behavioral interventions to mastery criterion set at 90% or above.

Research Question 2: Is there a correlation between parent treatment integrity and child skill acquisition?

H_2: High levels of treatment integrity will lead to high rates of skill acquisition for children.

Scope of the Study

The scope of this study includes parents of children aged 5–18 receiving services from an applied behavior analysis (ABA) company based out of Florida. BST was chosen as the training intervention for this study based on evidence of its effectiveness for increasing treatment integrity in previous literature (Barnett et al., 2014; DiGennaro-Reed & Codding, 2013; Jenkins and DiGennaro-Reed 2016; Pantermuehl, & Lechago, 2015). Parents included have not received BST for a skill acquisition program.

Summary

Treatment integrity is the accurate implementation of an intervention and is applicable to many fields such as behavior analysis, psychology, and education (DiGennaro-Reed & Codding, 2013). High levels of treatment integrity lead to more substantial and positive behavioral outcomes while helping researchers and practitioners determine that the intervention caused behavioral changes instead of an extraneous variable. Treatment integrity also helps researchers and practitioners catch and address problems early (McIntyre, Gresham, DiGennaro, & Reed,

2007). Despite its importance, limited research exists on treatment integrity, particularly on ways to increase treatment integrity, and examining treatment integrity for parent implementation. Therefore, there is a large need for research in this area. This dissertation will examine the use of BST to increase treatment integrity for parent implementation for their children's behavioral program(s).

Chapter 2 will present literature related to treatment integrity for staff and parent training, methods used to measure treatment integrity, common sources of treatment integrity errors, methods to increase treatment integrity including BST, and the social validity of treatment integrity. Chapter 3 will present the methods that will be used for this dissertation study.

Chapter 2: Literature Review

Introduction

Treatment integrity is the accurate implementation of an intervention as it was intended or written (Barnett et al., 2014; DiGennaro-Reed & Codding, 2013). It is important for all implementation agents from behavior analysts to staff members and teachers to implement interventions consistently and with integrity to ensure that behavior change occurs and can be attributed to the intervention and no other factors related to incorrect implementation (McIntyre et al., 2007). Treatment integrity is also important for determining functional relations, eligibility decisions, and high-risk or dangerous behaviors (Barnett et al., 2014). Several recommendations have been made regarding treatment integrity including assessing treatment integrity frequently, setting treatment-integrity mastery criteria at 90%, examining individual components of interventions, and programming consequences based on treatment integrity levels (McIntyre et al., 2007). Training should also be conducted to ensure high levels of treatment integrity among implementation agents (McIntyre et al. 2007).

Several methods have been used to measure treatment integrity, such as direct observation, self-report, interviews, and permanent product review (Barnett et al., 2014; Fiske, 2008; Hagermoser-Sanetti & Collier-Meek, 2014; Wilkinson, 2007). Direct observation, the most frequently used, is often referred to as the gold standard of treatment-integrity measurement (Hagermoser-Sanetti & Collier-Meek, 2014). In addition to choosing which general measurement to use for the collection of treatment-integrity data, it is also important to take into consideration global versus component integrity measures, scheduling of observation sessions, and whether to create a protocol (Hagermoser-Sanetti & Kratochwill, 2008; Hagermoser-Sanetti et al., 2014; Plumb & Vilardaga, 2010). These factors are important to ensure that a valid

measurement is obtained as poor treatment integrity has been found to lead to a lack of skill acquisition or behavior suppression. Additionally, based on the type of treatment integrity that occurs, such as an error of omission or commission, different effects may be obtained.

Several methods have been evaluated to determine their effectiveness in increasing treatment integrity, including using a lottery based incentive system, self-monitoring and video self-monitoring, token economies, differential reinforcement of other behavior (DRO), permanent product data collection, performance feedback, rehearsal, didactic instruction, modeling, coaching, goal-setting, and having the implementation agent engage in data collection (Codding, Livanis, Pace, & Vaca, 2008; Craig, 2010; DiGennaro-Reed, Codding, Catania, & Maguire, 2010; DiGennaro, Martens, & Kleinmann, 2007; Howard, Burke, & Allen, 2013; Jenkins & DiGennaro-Reed, 2016; Miller, Carlson, & Sigurdsson, 2014; Mouzakitis, Codding, & Tryon, 2015; Pantermuehl, & Lechago, 2015; Pelletier, McNamara, Braga-Kenyon, & Ahearn, 2010; Plavnick, Ferreri, & Maupin, 2010). Researchers have found advantages and limitations to each; however, BST and its components have been found effective consistently across studies. Lastly, in addition to methods to increase treatment integrity and measurement, social validity is another important factor to take into consideration when attempting to increase treatment integrity levels. For example, it has been found that the most acceptable and socially valid training procedure may also be the most effective (Strohmeier, Mulé, & Luiselli, 2014).

For this paper, a review of the literature was conducted on June 3, 2017, using ProQuest database accessible through The Chicago School of Professional Psychology. Search terms included treatment integrity, procedural fidelity, procedural integrity, and implementation integrity located in the document title within the date range of January 1, 2002 through June 3, 2017 in order to get an accurate assessment of the literature that has been conducted in this area

over the past fifteen years. Results were also limited to the following peer-reviewed journals: *Journal of Applied Behavior Analysis*, *Journal of Behavioral Education*, *Behavior Analysis in Practice*, *Behavior Modification*, *International Journal of Behavioral Consultation and Therapy*, *Journal of Organizational Behavior Management*, *Behavioral Interventions*, *Journal of Positive Behavior Interventions*, *Behavior Therapy*, *Behavioral Development Bulletin*, *The Behavior Analyst Today*, and *The Behavior Therapist*.

First, the importance of treatment integrity will be discussed. Second, there will be a discussion of methods used to measure treatment integrity. Third, studies showing lack of acquisition or behavior suppression based on poor treatment integrity will be discussed. Fourth, common types of errors regarding treatment integrity and errors of omission and commission will be presented. Fifth, methods used to increase treatment integrity will be reviewed. Sixth, there will be an examination of social validity measures as related to treatment integrity. Last, articles of treatment integrity research populations will be presented.

Importance of Treatment Integrity

Applied behavior analysts have developed countless effective interventions and have repeatedly demonstrated that, even in individuals' natural environments, behavior often responds quickly to interventions that are implemented accurately. Despite behavior analysis' demonstrations of important behavior changes, simply providing repeated demonstrations of effectiveness of behavior change procedures may hurt the field if others do not implement these procedures with integrity (Kunkel, 1987). Behavior analysts cannot be everywhere and need to rely on others to implement their procedures. The success of an intervention is dependent not only upon its effectiveness, but also upon its precise delivery by a clinician and the consistency with which parents and staff implement that treatment with all of its essential features.

Practical uses of treatment-integrity data are to guide clinical decision making such as changes in interventions and to provide feedback to intervention agents (McIntyre, Gresham, DiGennaro, & Reed, 2007; Vollmer, Sloman, & St. Peter Pipkin, 2008). Measures of treatment integrity help the researcher and practitioner make conclusions and decisions regarding intervention effects in all settings (McIntyre et al., 2007). Treatment integrity is also important for determining functional relations, eligibility decisions, and for high-risk or dangerous behaviors (Barnett et al., 2014). Additionally, without measuring treatment integrity, it is impossible to know if assessments and interventions were implemented as intended. Treatment procedures are developed, and changes are made based on the client and their needs, interactions between the client and therapist, details of treatment exposure, participant responsiveness, and quality of treatment delivery (DiGennaro-Reed & Codding, 2013); therefore, treatment integrity is important for several clinical areas.

Despite the importance of treatment integrity, a limited amount of research exists in this area for both staff and parent implementation. There are several barriers that may contribute to the lack of research in this area. Individuals often have a limited knowledge of treatment integrity and its theoretical basis including the importance of including measures of treatment integrity in research and practice (DiGennaro-Reed & Codding, 2013; McIntyre et al., 2007). If individuals do choose to include measurements of treatment integrity, they may have difficulties related to a lack of guidance, and constraints on time, cost, and labor required to collect and evaluate this additional data (DiGennaro-Reed & Codding, 2013; McIntyre et al., 2007). In terms of research, there has historically been a lack of accountability on behalf of editorial boards, specifically relating to space limitations and publication biases (DiGennaro-Reed & Codding, 2013; McIntyre et al., 2007). If interventions produce desired effects, researchers and editors

may not view treatment-integrity data as important; however, researchers and practitioners need to overcome these barriers, because measurements of treatment integrity help to make conclusions and decisions regarding intervention effects in all settings (McIntyre et al., 2007).

Hagermoser-Sanetti and Kratcochwill (2008) proposed that there was a double standard in research in that operational definitions and assessments of integrity, in the form of interobserver agreement or reliability, are often provided for the dependent variable in studies; however, there are rarely such descriptions for the independent variable. Researchers are more likely to define and describe the independent variables without measuring the accuracy of the implementation of them. The researchers also noted that researchers may leave out low treatment integrity values to increase chances of publication. This is a dangerous practice because if treatment integrity values are low for an intervention, a researcher cannot conclude that there was a functional relationship between the independent variable and the dependent variable even if behavior change was achieved. Therefore, the researcher cannot make the statement that their intervention was effective. Without the reader having access to this information, they may implement these interventions with their client, expecting similar results, and instead be implementing a non-evidence-based, and possibly ineffective, intervention. It is important that treatment-integrity data to be a requirement for publication for these reasons and for insurance that a functional relation was truly obtained.

McIntyre et al. (2007) conducted a review of the literature on school-based interventions with children, published in the *JABA* between 1991 and 2005 and, in doing so, demonstrated the lack of treatment integrity reporting in this research. The criteria for this review was that all studies had to be experimental, had to describe the intervention, use participants younger than 19 years old, evaluate results using a “true” experimental design (including a clear baseline and

manipulation of an independent variable), be completed in the school setting, and finished articles had to be longer than three pages. A total of 152 studies met these inclusionary criteria. The results were that most studies provided operational definitions of the independent variable (i.e., treatment procedures), approximately one third of the studies provided treatment-integrity data on the implementation of the independent variable, and studies that reported these data had a high percentage (M = 93%; SD = 9.93) of integrity. Additionally, reporting treatment-integrity data did not appear to differ consistently by publication year; however, results were variable based on year. For example, the years 1996, 1998, 1999, and 2005, included a larger percentage of studies with treatment-integrity data, and 1985, 1993, and 1995 did not include any studies with treatment-integrity data. Lastly, the researchers noted that reporting rates of treatment-integrity data have been remarkably stable, and low, over the past fifteen years (30% of studies included treatment-integrity data).

Methods Used to Measure Treatment Integrity

Barnett et al. (2014) conducted a review of the methods for collecting treatment-integrity data. They reviewed articles from 2005 through 2012 in *JABA*, *Journal of Behavioral Education*, Journal of School Psychology, *Psychology in the Schools*, *School Psychology Quarterl*y, and *School Psychology Revie*w. Inclusion criteria included studies that were experimental, studies that examined an academic or behavioral intervention in a school setting, participants who were under 19 years old and in grades prekindergarten through 12, and the target variable was student behavior. These criteria resulted in 266 articles. Results were that 70% of studies included treatment-integrity data, with most of these studies coming from *JABA*. In terms of method, the most frequently used was direct observation with a checklist or rating scale (66%), second frequent was direct observation alone (16%), and third was self-report (5%). Less than half of the

studies (45%) collected procedural fidelity data for more than 40% of sessions; however, more than half of studies (62%) included a plan for performance feedback checks across phases. The most frequent intervention agent was the researcher in 50% of studies; however, when the researcher implemented the intervention, they were less likely to report treatment-integrity data. For example, only 61% of studies in which the researcher was the implementation agent reported treatment-integrity data. In contrast, if a teacher or other school staff implemented the intervention (or a combination of both), treatment-integrity data were reported for over 70% of the studies. According to the results, there remain several opportunities for improvement regarding treatment integrity in present and future research, as well as the collection of these data in practice. Regarding improvements, Barnett et al. made several suggestions for future research and practice including increasing treatment-integrity data collection and measurement across settings to prevent intervention failure. Specifically, Barnett et al. recommended preventing low levels of treatment integrity from becoming a problem by using multi-element design to include conditions for the acquisition and fluency of interventions under natural and prompted conditions. They also recommended using BST and functional analysis (FA) procedures to address performance problems, assessing situational risk, using multiple treatment-integrity methods, and examining scheduling of integrity checks.

Barnett et al. (2014) reported that one of the most frequently used methods of treatment integrity collection is direct observation. Fiske (2008) defined direct observation as a method in which a rater observes the implementation of the intervention *in vivo* and records the occurrence or nonoccurrence of components of the intervention. Direct observation may result in accurate measurement of many components, including those that cannot easily be measured through permanent products, such as quick, repetitive behaviors, or vocal behaviors (Fiske, 2008). Fiske

(2008) noted that one difficulty with direct observation is the possibility of reactivity, in which the presence of the rater inflates the typical levels of integrity; however, data collected during inconspicuous observations may capture low levels of treatment integrity. Reactivity may be reduced using unscheduled and unobtrusive observations or the researcher's frequent presence in the environment for various reasons such that the individual being observed is unaware of the reason for the researcher's presence.

Direct observation, although the most frequently used, is only one method of collecting treatment-integrity data. Other methods include self-report, interviews, and permanent product review. Each method has unique strengths and weaknesses. Self-report, such as self-collected data or narrative accounts, reduces response effort required by the data collector; however, self-repot may overestimate the level of treatment integrity and produce inaccurate reports. If required to collect treatment-integrity data on their own implementation, individuals may inflate scores to demonstrate a greater skill level to their supervisor or other staff, to avoid or escape remedial-training sessions, or to receive an available reinforcer, including financial reimbursement, praise, or promotions. Additionally, individuals may unknowingly inflate their treatment-integrity scores for several reasons such as limited skills and resources including adequate time to collect data. Including the use of permanent products such as recording forms, notes, and worksheets may be beneficial in ameliorating these issues by increasing accuracy of data collection; additionally, by incorporating such permanent products, reactivity and time requirements for data collection may be reduced (Wilkinson, 2007). The implementation agent may not realize the data being collected are for treatment integrity purposes and to evaluate their implementation, and the presence of an observer is likely to be limited, reducing reactivity of their presence and time-requirements for frequent data collection for the observer or for the

implementation agent. Including multiple forms of data collection can also lead to heightened efficiency depending on the intervention of interest.

Using interviews to collect treatment-integrity data also reduces response effort required by the data collector. Additionally, interviews that incorporate performance feedback may increase the level of treatment integrity for staff's implementation of programs, enhancing the effectiveness of procedures (Wilkinson, 2007). The treatment-monitoring interview (TMI) is a semistructured interview that includes performance feedback and several other components. The TMI guides the consultant and consultee in the following areas: reviewing behavioral progress, identifying barriers or obstacles to intervention implementation, determining the need for modification, analyzing data, providing positive praise and corrective feedback, addressing questions and concerns, determining the need for further training and support, and reaffirming the consultee's commitment to implementation of the intervention.

Although the TMI provides a thorough assessment of treatment integrity, it also has several limitations (Wilkinson, 2007). First, the TMI is labor-intensive in terms of response effort and time. Labor-intensive assessments may be unlikely to be carried through by supervisors and instead should only be used when the benefits of doing so outweigh the costs, and supervisors are thoroughly trained in these procedures. Second, the TMI includes prompts for more frequent feedback than may be necessary and sufficient for high levels of integrity. The amount of feedback required for an individual to be effective and implement interventions with integrity should be individualized, as some individuals need more support than others to reach the same level. Preset criteria of feedback are unlikely to be useful for all implementation agents. Further, the use of consultee self-report may inflate estimated levels of integrity when compared to direct observation and permanent product data. Therefore, although interviews may be a

method used to collect treatment-integrity data, it should only be used when the benefits of doing so outweigh the limitations.

Collecting treatment-integrity data through the use of permanent products, termed permanent product review (PPR), also has strengths and weaknesses. With PPR the observer can collect data on all instances of implementation across time and settings; however, PPR may not be useful for all components of an intervention (Hagermoser-Sanetti & Collier-Meek, 2014). Hagermoser-Sanetti and Collier-Meek (2014) compared the relative merits and limitations of PPR with direct observation or specific direct observation (SDO). Participants of this study were three student–teacher dyads in a public-school setting. Teachers reported that these students engaged in behavior problems in their classrooms including off-task verbal and motor behavior, difficulty with redirection, excessive requesting, and noncompliance. Due to these behavior problems, each student had an individualized behavior support plan (BSP). Teachers' implementation of BSPs was measured through SDO and PPR to assess if different procedural integrity estimates were calculated based on the method used and the relationship of these differences to behavioral outcomes. SDO was used to measure teacher adherence to their assigned behavior support plan as a primary measure and quality of implementation as a secondary measure. Adherence was rated for each BSP step as implemented as planned, implemented with deviation, not implemented, or not observed. Quality for each BSP step was rated for each step as excellent, good, fair, or poor. PPR was used to measure teacher adherence to the plan but could not be used to measure quality. SDO was conducted for 30 min two to three times a week, and PPR was conducted one to three times a week. Student outcomes were measured through direct observation with observations conducted for 15 min (broken into sixty15 s intervals) two to three times a week. Student outcomes included academic engagement

measured by momentary time sampling, and disruptive behavior measured by partial interval recording when the student was most likely to engage in problem behaviors. For momentary time sampling, the observer measures if a behavior occurs at the specific time of observation, at a single preset moment. For partial interval recording, the observer measures if the behavior occurred at all during a preset interval. Several analyses were conducted including: descriptive statistics, Spearman's rho correlation, and visual analysis of graphed data. Results were that SDO estimates had a higher correlation with student outcomes than PPR estimates and represented more components of the behavior support plans than PPR. For example, if a behavioral event does not result in a permanent product, such as time off task, omission of a reinforcer, or a verbal-vocal response, PPR is not likely to detect this component. This suggests that there may be considerable differences in the utility of treatment-integrity data based on the assessment method used and that multiple treatment-integrity assessment methods provide richer data to base decisions on. These results support the authors' reference to SDO as the "gold standard" as it appears to be the most valid and reliable measure of treatment integrity. However, it is still important to consider the intervention before selecting an assessment method.

If choosing to use direct observation to collect treatment-integrity data, it is important to take into consideration scheduling of observation sessions. Previous researchers indicated that when observation sessions are broken into three segments, the second or middle segment may be most indicative of whole session data (Weck, Grikscheit, Höfling, & Stangier, 2014). In contrast, the first or third segment show lower interrater reliability, weaker relationships with treatment outcomes, and lower correlations with treatment integrity ratings for entire sessions (Weck et al., 2014).

Weck et al. (2014) compared adherence and competence ratings based on session segments to ratings based on entire sessions for reliability and relationship to treatment outcome for cognitive behavior therapy (CBT) for three disorders; mood disorder, anxiety disorder, and somatoform disorder. CBT conditions for three randomized control trials were considered for adherence and competence of therapists for 84 clients and 50 therapists. Therapists were trained through workshops for 20 hrs and received supervision once a month. Two judges evaluated the treatment integrity of entire sessions and two different judges evaluated the treatment integrity of the middle segments. Rating scales were created and used to assess adherence and competence using statistical analyses to evaluate the results. Results were that ratings of therapist adherence and competence based on segments were equivalent and strongly correlated to ratings based on entire sessions. Reliability values for ratings of therapist adherence and competence based on segments did not differ significantly from entire sessions. This suggests that segment-based ratings may be suitable for comparison to session-based ratings and highlights the importance of examining components of interventions when evaluating effectiveness. However, limitations of this study are that this study exclusively examined CBT and therefore the results may not be generalizable to different therapeutic approaches. Additionally, only one treatment session was included for the assessment of treatment integrity and assessments conducted were in highly standardized contexts, further reducing the generalizability of the results.

Component integrity data can be collected by creating a task analysis that includes all components of the intervention and rating whether each component was implemented as written when the opportunity to observe the component was available and noting times when the opportunity to observe certain components was not available or inappropriate. A percentage may then be calculated by dividing the number of components implemented as written by the total

number of components observed. Fiske (2008) discussed the benefits of this approach, such as the availability of detailed information that can lead to specific performance feedback for the implementation agent and modifications to programming and training as needed. Although direct and systematic observation is the most common method of assessing treatment integrity, it also is vulnerable to reactivity effects and can be labor intensive (Fiske, 2008; Wilkinson, 2007). To ameliorate this limitation, Wilkinson (2007) suggested using consultee self-report, permanent products, behavioral interviews, and feedback from consultants as other methods to assess degree of integrity in the implementation of behavioral interventions.

Hagermoser-Sanetti and Kratochwill (2008) also discussed the importance of component measurements of treatment integrity. Treatment integrity is usually measured as a percent of components implemented as intended, even though it may be a multidimensional construct. Dimensions that may be important in the measurement of treatment integrity are adherence, exposure, quality of delivery, program differentiation, and participant responsiveness. Not all components of a treatment or intervention may be equally important, hence the importance of component integrity measures. Another option is to use global measures, meaning an overall score of components, to assess treatment integrity, although this does not provide as specific of a measurement and it is difficult to determine which components need additional training and are essential for behavior change. Therefore, component integrity leads to more effective, data-based decisions.

Cook, Subramaniam, Brunson, Larson, Poe, and St. Peter (2015) conducted a specific comparison of global and component treatment-integrity measurements. They compared global and component treatment-integrity scores by systematically assessing how training and feedback differentially affected treatment integrity of discrete trial teaching (DTT). Participants in this

study were four behavior therapists with no experience implementing DTT. All participants worked with an 8-year-old child diagnosed with mild intellectual disability, attention deficit hyperactivity disorder (ADHD), posttraumatic stress disorder, and phonological disorder. The dependent variables were global and component treatment-integrity scores for nine components of DTT. Observers used a modified version of the Discrete-Trial Teaching Evaluation Form (DTTEF; Fazzio et al. 2010) to measure treatment integrity from video recorded sessions of a 12-trial DTT program. Global integrity was calculated by dividing the number of correctly performed steps by the total number of opportunities to implement each step in a session and converting to a percent. Component integrity was calculated by dividing the number of correctly implemented responses within a single component by the total number of opportunities to implement each component response within a session and converting to a percent. The number of sessions conducted each day by therapists varied; however, mastery criteria was set at three consecutive sessions with 80% or greater treatment integrity. All participants taught the child expressive letter sounds from the same set of targets. Additional interventions for participants were put in place to increase treatment integrity if mastery criteria were not met. Interventions for participants began with the least amount of trainer time required and faded into interventions requiring the most amount of trainer time as needed. Baseline occurred the day before conducting sessions in which participants were provided a 30-page written instruction manual describing how to implement the components of DTT, written at an eighth-grade reading level and containing pictures to further demonstrate correct techniques. During video modeling, participants were provided a 2-min video focusing on reinforcer identification and delivery with rationale. Videos included written text and still photos to highlight important details. Video models were only used for the reinforcer component of the DTT programs to assess its effects on

this component as well as generalization to other DTT skills. Written feedback consisted of two to three pages of praise for components implemented correctly and constructive feedback for components implemented incorrectly, delivered to the participant in between subsequent sessions. Written feedback included pictures of correct implementation, brief descriptions, and rationale. Results were that global integrity scores generally increased across conditions for all participants; however, increases in global integrity did not represent an increase in performance across all individual components. Therefore, global scores (correct implementation across all components) provided different measurements of treatment integrity as compared to component scores (correct implementation of component responses). Overall, participants showed improvement in accurate implementation of DTT steps with training, but accuracy of individual components remained variable throughout the study, with error-correction remaining one of the least accurate steps implemented across most participants. Cook et al. (2015) highlighted the importance of including a component integrity measure as compared to relying solely on global measures, as more specific information and results may be demonstrated. Additionally, by relying solely on a global integrity measure, researchers and practitioners are in danger of misinterpreting whether additional training is needed and for which components training is required. By knowing if and what to train, behavior analysts can be more effective and efficient in their practices. Global scores may also result in over-estimation of correct implementation and hide important effects of training on staff behavior and implementation on client behavior.

In addition to taking into consideration global versus component integrity measures, scheduling of observation sessions, and form of measurement used when collecting treatment-integrity data, it can also be beneficial to create a protocol to use when collecting such data. Plumb and Vilardaga (2010) provided some guidelines to be considered when developing a

treatment integrity protocol: (a) providing standardized procedures for treatment integrity raters to identify and operationalize elements of treatment, (b) identifying (in writing) items to be assessed and a rubric for assessing them, (c) relying on observable behaviors only, and (d) choosing a coding unit. Standardized procedures are important because treatment integrity should be compared within and across interventions and so that training can be adjusted as needed. By laying out in writing the components that need to be assessed and creating a rubric and coding system by which to do so, treatment integrity can become an objective part of research and practice. Furthermore, behavior analysts rely on clearly and objectively defined observable behaviors in measurements so that their effects can be replicated and for consistency across sessions, settings, and practitioners or researchers.

It is also necessary when developing a treatment integrity protocol to collect data on all components of an intervention and collect an adequate data sample that is representative across time and settings (Hagermoser-Sanetti et al., 2014). This way, the researcher or practitioner can ensure that generalization of behavior change has occurred. Treatment-integrity measurements should also include competence or consistency and adherence (Plumb & Vilardaga, 2010). It is important to know if individuals are consistently implementing specific components incorrectly or are implementing components with variable levels of accuracy. Additionally, it may be beneficial to train reliability before completing data collection and it may also be helpful to look at processes and procedures to assess strengths and weakness of implementation agents (Plumb & Vilardaga, 2010). Based on these measures, training can be individualized based on the implementation agent.

Reed, Fienup, Luiselli, and Pace (2010) examined planned treatment integrity observations on schedule-induced responding for implementation agents and provided some

best-practice recommendations for promoting treatment integrity. These best practices include selecting an intervention that is backed by evidence, is likely to be successful, requires minimal response effort and resources, has buy in, and utilizes positive reinforcement procedures (Reed et al., 2010). They also discussed the importance of implementing antecedent and consequent strategies, remaining alert to any effects of reactivity, and conducting observations in a steady, intermittent distribution. Additionally, the temporal sequence of observations is important to remain alert to when aiming to reduce reactivity of the observer's presence.

Lack of Acquisition and Behavior Suppression

Several studies have investigated the impact of treatment integrity on intervention effectiveness, as well as numerous methods to increase treatment integrity in work with clients. For example, Fryling, Wallace, and Yassine (2012) conducted a review of the research on treatment integrity and made several conclusions. First, improved integrity led to intervention effectiveness. Second, the extent to which integrity failures affected intervention efficacy might be intervention-specific, especially when interventions involve multiple components. Third, compliance systematically varied according to the level of treatment integrity. Fourth, teachers and caregivers can implement an intervention with low integrity and still maintain treatment outcomes if they initially implement the intervention with high levels of integrity. Fifth, some errors in implementation during discrete-trial training may be just as detrimental as 100% errors. This further reinforces the importance of treatment integrity in that it has a significant effect on intervention effectiveness. Moreover, when looking at why an intervention may not have been implemented accurately, there are several factors to consider. Factors that may influence treatment integrity's effects on intervention effectiveness include whether the participant was exposed to a low level of integrity before a high level and the relationship between specific

components of the intervention to the level of integrity associated with that component (Noell, Gresham, & Gansle, 2002).

Two approaches to examining treatment integrity include *descriptive* and *experimental*. In the descriptive approach, the researcher can examine how variations in implementation are associated with variations in behavioral outcomes in the natural setting, it does not allow causal inferences because an experimental design and manipulation of variables are not used (Noell et al., 2002). For example, if a researcher implements a staff training intervention and then examines the effect of differing levels of treatment integrity on client outcomes, this would represent a descriptive approach to examining treatment integrity. In contrast, in an experimental approach, the researcher manipulates the levels of treatment integrity as an independent variable through two or more conditions within an experimental design and then examines the effects on a dependent variable such as client behavior (Noell et al., 2002). An example of an experimental approach would be if a computer program systematically manipulates the level of treatment integrity through multiple conditions and then compares the effects on student behavior. It is best to use the experimental approach; however, despite the approach used, it is important to assess treatment integrity's effects on behavior.

Several studies examined the effects of levels of treatment integrity on behavior. Noell et al. (2002) examined the impact of varying degrees of treatment integrity using computer-provided instruction to teach math skills, an experimental approach to examining treatment integrity. The researchers used a multiple-baseline, multiple-sequence design across students and curricular materials, meaning the computer program implemented the program at varying levels of integrity. For example, prompts, accuracy feedback, or praise were not presented in baseline. In implementation phases, depending on the level of integrity programmed, either 100%, 67%, or

33%, prompts were only provided for that percentage of presented problems. In implementation phases, accuracy feedback was provided for all problems and praise was provided on a VR-3 schedule. The order of conditions was counterbalanced. Results were that higher treatment integrity levels, 100% as compared to 67% or 33%, resulted in greater accuracy for 10 out of 12 participants. For one participant, mean accuracy in the 67% integrity condition was greater than accuracy in the 100% integrity condition; however, the 67% condition followed the 100% condition. For the second participant, mean accuracy in the 67% integrity condition was equal to accuracy in the 100% integrity condition. Results for the rest of the participants in the 67% accuracy condition were variable. A question remains to be addressed: what level of treatment integrity is necessary for an intervention to be effective? This answer may vary based on the intervention. For example, operant reinforcement programs may remain effective despite reduced levels of integrity, and skills teaching may require a greater level of treatment integrity than fluency training. Limitations of this study were that the use of a computer to deliver prompts may have reduced the external validity of the results as inconsistencies are more likely with an individual implementing the intervention as compared to a computer. Additionally, the preprogrammed length of phases resulted in shorter phases than may have been used in an applied setting.

Wilder, Atwell, and Wine (2006) conducted a parametric analysis of varying levels of treatment integrity for an antecedent intervention to examine the effects on child compliance for two 4-year-old children. The antecedent intervention involved three steps: presenting an instruction, providing brief praise for compliance, obtaining eye contact and modeling of the correct behavior for noncompliance, and re-presenting the instruction with guided compliance for continued noncompliance. The levels of treatment integrity presented to the participants were

100%, 50%, and 0%. Each level of integrity was associated with one of three instructions: “come here,” “give me …,” or “put the toy away.” The three levels of integrity were associated with the percentage of trials on which the procedure was used. For example, during the 100% integrity condition, procedures were implemented for all commands, during the 50% integrity condition, procedures were implemented for only half of the demands, and during the 0% integrity condition, the procedures were not implemented at all. Compliance to these demands was measured as the dependent variable, specifically if a participant complied within 10 s of the first prompt, for the three-step prompting intervention. Data were collected on the percentage of trials in which the participants complied. Three settings were used (tutoring room, classroom, and playground) and in each setting, a different instruction and level of treatment integrity was implemented. Results were evaluated using a multi-element design with baseline and were that the integrity with which the three-step prompting procedure was implemented had a large impact on its effectiveness in increasing compliance. Specifically, inconsistent implementation of interventions resulted in a reduction of the effectiveness of the intervention, meaning 100% integrity resulted in the most compliance, 0% integrity resulted in the least amount of compliance, and 50% integrity resulted in compliance levels between these two. Limitations of this study were that preferences may have varied among participants, which was not controlled for in the study and could have affected levels of compliance.

In addition to varying levels of treatment integrity, specific errors regarding the accuracy of implementation have also been examined to determine their effects on behavior. Carroll, Kodak, and Fisher (2013) assessed the effects of treatment integrity errors on skill acquisition for children with autism during DTT in a clinic setting across three experiments. In the first experiment, observations were conducted to assess treatment integrity errors that normally

occurred during instruction. The most common errors included the delivery of tangible reinforcers, providing controlling prompts incorrectly, and presenting instructions more than once. In the second experiment, Carroll et al. conducted a comparison of the effects of high- and low-treatment integrity conditions on skill acquisition through the use of an adapted, alternating treatment design. Low levels of treatment integrity decreased the effectiveness and efficiency of teaching procedures. Specifically, all but one of the participants mastered targets in the high-integrity condition; however, only one participant mastered targets in the low-integrity condition. In the third experiment, Carroll et al. evaluated the effects of the three most common treatment-integrity errors against high treatment integrity using an adapted alternating treatments design. The four conditions were high integrity, errors in instruction, errors in controlling prompt, and errors in reinforcement. The results were that all three participants showed slower acquisition of targets in one of the low-integrity conditions; however, the conditions varied across participants, stressing the importance of implementing interventions with a high degree of integrity. Additionally, the effects that errors in implementation of intervention components will have on an individual may depend on individual factors such as difficulty of a specific skill and targeted items. Limitations of this study are that previous exposure to low-integrity conditions may have affected responding in Experiment 3 and an early-termination criterion was used, which may have affected time to mastery of target stimuli.

Pence and St. Peter (2015) conducted a study to evaluate treatment integrity errors on mand acquisition across two experiments. In the first experiment, incorrect reinforcer delivery was assessed in relation to mand acquisition. The dependent variable was percent of independent target mands, and the independent variable was the intervention, which was evaluated using different treatment integrity levels. Specifically, the intervention consisted of

gaining the student's attention, using a least to most prompting hierarchy, and the delivery of preferred, nonpreferred, or neutral items. A multi-element arrangement was used within a multiple-baseline design across participants. Phases of the multiple-baseline design were baseline and training across students to examine the effects of training on mand acquisition. The multiple elements contained within this design were the levels of treatment integrity (0%, 40%, 70%, and 100%) during training to evaluate the effects of differing integrity levels on acquisition. Treatment-integrity levels were randomly alternated. In baseline, no prompts were provided for manding such as saying the name of the reinforcer; however, if the student manded for the reinforcer correctly, the reinforcer was provided to them. Training consisted of using a least to most prompting hierarchy. A nonspecific prompt such as stating, "What do you want?" was given if the student did not emit the correct mand within 10 s of the start of the trial. A partial prompt, the first syllable of the target word, was then provided if another 10 s passed and the student did not emit the correct mand. If the student did not emit the correct mand at this point, the therapist provided a full prompt by stating, "What do you want?" and the name of the reinforcer and continued this prompt every 10 s until the participant emitted the correct mand or the trial ended. If a non-targeted-mand or other vocalization was emitted by the student, the trial was marked as an error. For error trials, such as in conditions with reduced levels of treatment integrity, the therapist provided an item other than the reinforcer, either neutral or nonpreferred, for correct manding. Ten randomized sequences of trials were created for 40% and 70% treatment integrity so that the order of correct implementation and implementation with errors was not repeated. For example, for the 70% integrity conditions, there were three error trials and seven full integrity trials within a session. Results of this experiment were that mands were acquired more quickly in the 100% integrity condition for both participants; however, there were

variable effects in mand acquisition during the reduced integrity conditions. For the 0% integrity condition, correct mands were acquired for one out of two participants, which may reflect that acquisition of a generalized mand, such as manding for toys in general, rather than a specific toy. For one participant, 40% and 70% integrity conditions resulted in reduced acquisition; however, the second participant still met mastery criteria in both conditions. Yet, for both participants, 40% integrity was more detrimental to acquisition than 70% integrity; however, this may reflect a limitation such that the toy assigned to the 70% condition may have been a more potent reinforcer.

In Experiment 2, Pence and St. Peter (2015) assessed the effect of response-independent delivery of the target item on mand acquisition. Sessions, data collection, baseline, and treatment were conducted similarly to Experiment 1; however, treatment-integrity errors consisted of delivering the preferred item noncontingently once 5 s had elapsed in the trial. No prompts were delivered for correct mands. Results of this experiment were that, like Experiment 1, the quickest acquisition of independent mands was in the 100% integrity condition. Across participants, differentially higher levels of correct mands were emitted in the 70% integrity condition as compared to the 40% integrity condition, and participants never emitted correct independent mands in the 0% integrity condition. Overall, the results of this study demonstrate that lower levels of treatment integrity are detrimental to skill acquisition. Limitations of this study are that standardized assessments were not conducted to evaluate current repertoires for students and maintenance and generalization were not assessed in relation to treatment-integrity errors.

Carroll, Kodak, and Adolf (2016) evaluated treatment integrity errors on the implementation of DTT. In DTT, reinforcement is immediately provided for correct responding;

however, many practitioners make errors in integrity by inserting a delay between when the correct response is emitted and delivery of reinforcement to the client. Therefore, this study evaluated the effects of these treatment-integrity errors. Specifically, the effects of delays to reinforcement on skill acquisition during DTT for two children with autism were examined. Three conditions were included for this study. First, a condition took place in which combined reinforcement, consisting of praise and a preferred item, was delivered immediately following a correct response. In the second condition, only praise was delivered immediately, with delivery of a preferred item delayed. In the third condition, delivery of both praise and a preferred item were delayed. The dependent variable was correct child responses, defined as the participant saying the correct action or feature on a provided card in the absence of prompts, and the independent variables were the different types of reinforcement conditions. An adapted alternating treatment design was used. Results indicated that the immediate reinforcement condition was more effective and efficient; however, the other two conditions had variable results based on the participant. For example, the first participant mastered targets during all conditions, but required more training sessions to acquire targets in the condition where both praise and a preferred item were delayed. In contrast, the second participant only mastered targets in the immediate reinforcement condition. This study speaks to the importance of implementing DTT programs with perfect integrity to achieve the most effective and efficient programming. A limitation of this study is that reinforcer assessments were not conducted.

Treatment integrity has a significant effect on intervention effectiveness. Factors that may influence this relationship are sequencing effects and levels of integrity for components of the intervention (Carroll et al., 2013; Noell et al., 2002). Interventions may require different levels of integrity to be effective; however, across interventions, it has been consistently reported that

higher levels of integrity result in more effective and efficient procedures (Carroll et al., 2013; Pence & St. Peter, 2015; Noell et al., 2002; Wilder et al., 2006). Low levels of treatment integrity can be detrimental to skill acquisition and behavior reduction including DTT procedures (Carroll et al., 2016; Pence & St. Peter, 2015). Limitations of these studies include: (a) low external validity due to computerized programing, (b) low social validity in regards to preference differences of participants or conducting reinforcer assessments, (c) lack of control for the current behavioral repertoires, arbitrary nature of skills, early-termination criteria, lack of component analysis, and not assessing generalization and maintenance (Carroll et al., 2013, 2016; Grow et al., 2009; Noell et al., 2002; Pence & St. Peter, 2015; Wilder et al., 2006)

Common Types of Errors

Some common sources of treatment integrity errors include: inadequate training, complexity of the training or intervention protocol, lack of generalization, performance drift, and competing contingencies such as environmental contingencies or schedules of reinforcement (Vollmer et al., 2008). A potential solution in addressing errors in treatment integrity is to emphasize accuracy, provide consistent feedback about integrity levels, and provide reinforcement for high levels of integrity (Vollmer et al., 2008). This begins with professional development pertaining to treatment protocols and procedures such as BST. BST includes five components: instruction, modeling (in person and video), rehearsal or role-play, demonstration with the client in the moment, and feedback (DiGennaro-Reed & Codding, 2013). Additional options to promote and maintain high levels of treatment integrity include coaching support and follow-up interventions, regularly scheduled follow-up meetings, and directed rehearsal of treatment components that are implemented correctly or incorrectly (DiGennaro-Reed & Codding, 2013).

Errors of Omission and Commission

There are common errors that contribute to low treatment-integrity scores. First, errors of omission occur when individuals implementing interventions do not provide the appropriate response when a specific event occurs (Vollmer et al., 2008). Errors of omission involve not providing the appropriate response when a specific event occurs, failing to document a response or event, or failing to deliver a reinforcer (Vollmer et al., 2008). For example, in a fixed-time schedule of reinforcement, an error of omission may include not proving reinforcement when scheduled, such as at the end of a preset 5-min interval. Second, errors of commission occur when individuals implementing interventions provide a response at an inappropriate time (Vollmer et al., 2008). Errors of commission involve providing a response at an inappropriate time, recording an event when it did not occur, recording an event in place of a different event, or delivering an antecedent or consequence at an inappropriate time (Vollmer et al., 2008). For example, in a fixed-time schedule of reinforcement, an error of commission may include proving reinforcement in the middle of a preset interval, in addition to when that interval times out. These types of errors have differing effects on behavior and can impact the effectiveness and efficiency of programming depending on the behavior change intervention.

St. Peter Pipkin, Vollmer, and Sloman (2010) conducted a study to examine the effects of errors of omission and commission. The intervention used in this study was differential reinforcement of alternative behavior (DRA). DRA is a commonly used intervention for problem behavior in which a problem behavior is placed on extinction and reinforcers are provided for an appropriate alternative behavior (St. Peter Pipkin et al., 2010). When implemented with integrity, DRA is very effective; however, reduced levels of treatment integrity have negative effects, which may differ depending on the level of integrity for each component. St. Peter Pipkin et al.

examined the effects of treatment integrity failures on DRA using three experiments. In Experiment 1, college students were instructed to select circles to earn points. A black circle represented contingencies of reinforcement for problem behavior and a red circle represented contingencies of reinforcement for appropriate behavior. In baseline, the black circle was on a FR1 schedule, and selecting the red circle did not earn points. During a full-integrity DRA condition, selecting the black circle did not earn points, and selecting the red circle was on a FR1 schedule. Treatment integrity failures were evaluated using a random-ratio (RR) schedule for error of commission, errors of omission, or a combination. Students were assigned to four subsets. Group 1 experienced only errors of omission. Group 2 experienced only errors of commission. Group 3 experienced both omission and commission errors. Group 4 experienced 50% integrity for reinforcement and extinction components. Results were that the efficacy of DRA treatments decreased based on different kinds of treatment integrity failures. Specifically, errors of commission had a greater impact on responding as compared to errors of omission, but only at relatively low levels of treatment integrity. Additionally, combined errors seemed to have the same effects on behavior as commission errors alone. Therefore, commission errors may be responsible for more detrimental effects when treatment integrity is assessed as a single factor, and components are not analyzed separately.

In Experiment 2, St. Peter Pipkin et al. (2010) evaluated the effects of combined errors on the occurrence of problem behavior, off-task behavior, and appropriate behaviors, on-task behavior and task completion, measured for 15 s durations during DRA for a fourth-grade girl diagnosed with autism in a school setting. The procedure consisted of baseline, DRA, and treatment integrity errors. In baseline, the therapist attended to the student in each 15 s interval that she was off-task and ignored her on-task behavior. In DRA, the therapist attended to the

student when she was on-task for 15 s and ignored off-task behavior. Treatment integrity was then evaluated for four levels of combined omission and commission errors. These levels were 80%, 60%, 40%, and 30% integrity, which were presented to the participant in a counterbalanced order using a reversal design. Results were that the participant's behavior was affected by the degree of treatment integrity failure, but not significantly by condition sequence. Specifically, the participant engaged in less off-task behavior in the 80% and 60% integrity conditions as compared to the 40% and 20% integrity conditions. Additionally, the participant's behavior in subsequent presentations of conditions matched her behavior during the initial presentation. This differed from Experiment 1, in which the sequence did influence responding. For example, in Experiment 1, initially implementing the DRA with high integrity led to weakened effects of integrity failures as participants engaged in a greater amount of appropriate behaviors during the 50% integrity condition that followed the full-integrity DRA condition.

For Experiment 3, St. Peter Pipkin et al. (2010) examined sequence effects on treatment integrity failures by replicating the results of group 4 from Experiment 1 with an adolescent diagnosed with a developmental disability in a school setting. The procedure involved decreasing aggressive behaviors and increasing appropriate verbal communication through use of a DRA. During baseline, the student was ignored until aggression occurred, and then attention in the form of a neutral comment was provided. During DRA, prompting of appropriate greetings was immediately followed by a neutral comment as a form of attention as a reinforcer, and extinction was implemented for aggression. During the treatment integrity failure condition, a 50% level treatment integrity was implemented according to a RR2 schedule for aggression and appropriate verbal communication or greetings. Results were that integrity failures were more detrimental to the treatment when they followed baseline than when they followed treatment with perfect

integrity, and treatment effects during 100% integrity were more difficult to regain following a 50% integrity phase than following baseline. The results of this experiment suggest that if an intervention is initially implemented with reduced treatment integrity, effects of an intervention may be negatively affected, despite the intervention ending with high levels of integrity, particularly for a DRA intervention. Both errors of commission and omission can negatively impact treatment effects; however, reduced levels of treatment integrity may be less harmful if the intervention is initially implemented with a high level of integrity. Limitations of this study include that the use of college students in an analogue procedure may limit the generality of the results and that phases were changed at set points in time rather than based on behavior; therefore, behavior did not have a chance to stabilize.

Leon, Wilder, Majdalany, Myers, and Saini (2014) also examined the effects of different levels of treatment integrity and sequencing on child compliance across two experiments, for errors of omission and errors of commission, with two children in a preschool setting. Both children engaged in noncompliance when instructed to give up a preferred toy. During Experiment 1, the experimenters examined the effects of errors of omission during a DRA for compliance. The procedure consisted of four conditions: baseline, 100% integrity, 60% integrity, and 20% integrity, all of which were evaluated using a reversal design. In baseline, compliance with instruction resulted in praise and the experimenter kept the toy for 30 s before giving it back to the child. Noncompliance resulted in continued access to the toy. In the 100% integrity condition, the child was given a verbal direction of the contingency, compliance resulted in a preferred piece of candy and the experimenter kept the toy for 30 s before giving it back to the child. In 60% and 20% integrity conditions, compliance resulted in the delivery of candy on 60% and 20% of trials respectively. Noncompliance resulted in the child keeping the toy. Each

condition was associated with a different experimenter. Results of this experiment were that, for both participants, compliance varied per the integrity of the intervention. Compliance was lowest in baseline and in the 20% integrity condition and highest in the 100% integrity condition. Another important result is that when following baseline, the 60% integrity condition produced clinically significant increases in compliance. However, compliance was much lower in the 60% integrity condition when this condition followed 100% integrity.

During Experiment 2, Leon et al. (2014) examined the effects of errors of commission during DRA for compliance for three preschool children. The procedure consisted of four conditions: baseline, 100% integrity, 50% integrity, and 0% integrity, all of which were evaluated using a reversal design. Variables and baseline were the same as in Experiment 1. However, instead of a verbal description of the contingency, the experimenter told the participant that they would be given a piece of candy for handing over their toy or they may still be given a piece of candy if they do not. Therefore, the verbal signal was the same for all conditions. In baseline, compliance resulted in praise and the experimenter kept the toy for 30 s before giving it back to the child. Noncompliance resulted in continued access to the toy. In the 100% integrity condition, compliance resulted in praise and the delivery of candy. Neither was delivered for noncompliance. In the 50% integrity condition, compliance and noncompliance resulted in candy and praise for 50% of the total number of trials. In the 0% integrity condition, praise and delivery of candy was provided on every trial regardless of compliance. Results of this experiment were that, for all three participants, compliance was low during 0% integrity, high during 100% integrity, and for 50% integrity, compliance was low when it followed baseline and when it followed 100% integrity. Therefore, compliance varied according to the level of treatment integrity put in place regardless of the sequencing of conditions. The results of Experiments 1

and 2 suggest that with DRA errors of commission may be more detrimental than errors of omission, and less affected by the preceding context. Limitations of this study include that the use of an analogue procedure may limit the generality of the results, and trials in which reinforcement was delivered may have affected responding, which was unaccounted for in the results.

St. Peter, Byrd, Pence, and Foreman (2016) evaluated errors of commission and omission for a response cost procedure for 19 college students. In their first experiment, the researchers randomly assigned participants to either a 20% or 50% integrity condition for both commission and omission errors at that value. Participants were required to click on a red or black circle on a computer screen. Clicking on the designated circle would result in the delivery of points dependent on the differing schedules of reinforcement based on the conditions of the study. For example, in baseline, the red circle was on a variable-interval, 5 s schedule (VI5s) and the black circle was on a fixed-ratio, one schedule (FR1). In the response-cost condition, the red circle remained on the VI5s schedule, but the black circle resulted in the implementation of a response-cost procedure of FR1, with every click losing one point. After this, reduced integrity conditions, 20% integrity for omission errors, 20% integrity for commission errors, 50% integrity for omission errors, and 50% integrity for commission errors were put in place. For example, in the 20% omission condition, the red circle was on a variable-interval, 5 s schedule (VI5s) and the black circle was programmed with a 20% integrity response cost procedure, so for every five clicks, four clicks resulted in a point and one click resulted in loss of a point in a randomized order. The only difference in the 50% omission condition was that 50% of clicks on the black circle resulted in response cost. For reduced treatment integrity conditions for commission errors, the red circle remained on the VI5s schedule but also intermittently resulted in response cost for

a percent of clicks matching the treatment integrity percent. Results of Experiment 1 were that the greatest suppression of responding occurred in the 100% integrity of the response cost procedure. Across participants, as the integrity level increased for omission phases, there was an orderly decrease in black-circle responding and an orderly increase in red-circle responding even though there was no change in contingencies for clicking the red circle. For almost all participants, phases with commission errors reduced black circle responding similarly to response cost implemented with perfect integrity. Therefore, commission errors had less of an effect on responding during the response cost procedure. However, a limitation of this experiment was the use of a between-subjects design, so generalization of these results is limited as effects may be due to individual differences.

In response to this limitation, in Experiment 2, St. Peter et al. (2016) assessed whether these differences were attributable to individual differences for six undergraduate students as participants. All the procedures were identical, except for the design used. In Experiment 1, a between-subjects design was used; whereas, in Experiment 2, a within-subjects design was used. Results of Experiment 1 were replicated indicating that results of Experiment 1 were due to manipulations of the independent variable, levels of treatment integrity, not individual differences. St. Peter et al. highlighted the importance of implementing interventions with a high degree of treatment integrity, particularly when it comes to errors of omission for response cost procedures. A limitation of this study is that the fixed duration of phases sometimes resulted in phase changes before behavior stabilized.

These studies further demonstrate the importance of implementing interventions with high degrees of integrity and highlighted important aspects of integrity evaluation, such as evaluating component integrity rather than global scores for an entire intervention, sequence

effects, and analyzing the effects of errors based on the specific intervention (for example, errors of omission and commission; Leon et al., 2014; St. Peter et al., 2016; St. Peter Pipkin et al., 2010). They also highlighted the fact that, regardless of the intervention or procedures used, implementation at high levels of integrity is the most effective and efficient way to obtain behavior change (Leon et al., 2014; St. Peter et al., 2016; St. Peter Pipkin et al., 2010)

St. Peter and colleagues (2010) demonstrated that even if an intervention ends with high levels of treatment integrity, the level of integrity it is started with can affect responding to a great degree. The researchers demonstrated this effect for a DRA intervention. Additionally, errors of omission and commission can negatively impact an intervention. Effects of errors may depend on the intervention components and individual skill levels (Carroll et al., 2013). However, for certain interventions, such as reinforcement procedures, errors of commission may be more detrimental than errors of omission and less affected by sequencing effects (Leon et al., 2014). For other interventions, such as a response-cost or punishment procedure, errors of omission may be more detrimental then errors of commission (St. Peter et al., 2016). Limitations of these studies include (a) limited generality of findings due to analogue procedures, (b) limited chance for stable behavior patterns due to time-based phase changes rather than behavior-based changes, and (c) limited internal validity due to possible extraneous factors such as when reinforcement was delivered.

Increasing Treatment Integrity

Several methods have been evaluated to determine their effectiveness in increasing treatment integrity, including using a lottery-based incentive system, self-monitoring and video self-monitoring, a token economy, DRO, permanent product data collection, performance feedback, rehearsal, didactic instruction, modeling, coaching, goal-setting, and having the

implementation agent engage in data collection. Researchers have found advantages and limitations of each which will be further discussed.

Miller, Carlson, and Sigurdsson (2014) conducted a study to improve three special-education staff members (a 1:1 aide, teaching assistant, and teacher) implementation of DTT using a lottery-based incentive system. The dependent variables were treatment integrity of components of the DTT program, specifically, adherence to a five-step prompt hierarchy, accuracy of data collection, and writing in the activity for the following day. The prompt hierarchy included presentation of a discriminative stimulus, a verbal prompt, gestural prompt, partial physical prompt, and full physical prompt. Accuracy of data collection was defined as recording the most intrusive prompt used, even if applied incorrectly, and recording an incorrect response when appropriate. DTT procedures were video recorded once a day for three to five activities. A concurrent multiple-baseline design across participants was used to evaluate experimental control. In baseline, staff was not aware of the lottery system and intervention. During intervention, the staff was provided feedback about their past performance before each session, either baseline performance or the previous day, and was also provided a graph of their performance as well as verbal feedback. A weekly goal for performance was then set at 80% correct implementation. If staff met or exceeded this goal and were present for at least three sessions that week, they were entered in a weekly lottery for a $25 gift card to a retailer of their choice. Results were that most staff errors consisted of incorrect prompt implementation; however, the introduction of feedback and the lottery system consistently increased staff performance on DTT procedures above mastery criteria. Limitations of this study are that length and frequency of sessions for participants varied, data was not collected on student behavior, and it was not determined which component of the intervention had the greatest effects on

improvement of participant treatment integrity. Therefore, although a lottery-based system can be effective in increasing treatment integrity, it may not always be appropriate, such as in the case of limited funding or a small amount of implementation agents and may not be best for determining the effectiveness of individual components.

Self-monitoring has been shown to increase treatment integrity but has not been isolated from other procedures, also limiting the ability to determine component effectiveness. Plavnick, Ferreri, and Maupin (2010) conducted a study to identify the effects of a self-monitoring checklist on treatment integrity of a token economy, implemented by school staff. A secondary purpose was to identify any related academic-readiness student behaviors, specifically appropriate sitting and appropriate vocalizing. Three staff members, para-professionals at an early childhood special education program, and two students, one diagnosed with autism and the other diagnosed with Williams syndrome and specific language impairment, participated in this study. The dependent variable for staff was the percentage of token economy components implemented accurately and in the correct order during 15-min observations. The dependent variable for the students was the percent of intervals in which they engaged in two academic-readiness behaviors simultaneously, appropriate sitting and vocalizations. Two multiple-baseline designs were used to evaluate the effects of self-monitoring on treatment integrity and the effects of staff behavior on student behavior. Staff baseline consisted of three phases: pretraining, training, and implementation of the self-monitoring checklist. In pretraining, token economy components that were implemented accurately were recorded by a consultant as they had throughout the school year. The token economy had been put in place before the start of this study. Experimenters conducted the training phase for two, 1-hr sessions. During these sessions, experimenters trained staff to implement the token economy checklist. The first session of

training consisted of didactic review, modeling, and a role-play with staff. The second session of training consisted of the experimenter modeling the token economy procedures with student participants and then staff participants were instructed to implement the procedures with the students. Coaching and feedback were provided to staff until mastery criteria or 80% for one session was met by staff participants. This phase continued until each staff participant's treatment integrity of the token economy was stable or decreasing. The intervention was then put in place, self-monitoring. Self-monitoring consisted of the experimenter meeting briefly with staff to explain the monitoring checklist, review procedures, and answering questions. Staff continued to implement the token economy along with a token economy checklist after two sessions of their choice each day. Experimenters observed and recorded the occurrence or nonoccurrence of academic readiness behavior during all experimental conditions. The results were that training and self-monitoring increased levels of treatment integrity for implementation of the token economy. In pretraining staff did not implement the token economy. This increased to an average of 70% during training and an average of 84% during self-monitoring. This suggests that self-monitoring can improve the implementation of a token economy system. Additionally, student participants demonstrated improvements in academic readiness behaviors following the introduction of the token economy, increasing from an average of 25% during pretraining to an average of 89% during self-monitoring. Limitations of this study are that staff and student participants demonstrated variability in responding after the introduction of self-monitoring and effects of individual components on student outcomes were not examined.

As previously discussed, one limitation to self-monitoring is that the individual may misreport or inaccurately report data; this can be addressed using a permanent product. Craig (2010) conducted a study self-monitoring with a token-based, DRO procedure to decrease his

own behavior of nail biting. Permanent product data were gathered using a digital camera to photograph the condition of the fingernails at the beginning of each DRO interval, after each token delivery, and after each occurrence of nail biting. The photographs were assessed by a second observer to assess treatment integrity of the self-monitoring system and a reversal design was used to study the effects of the intervention on nail biting. Baseline consisted of self-monitoring only. A 1-hr DRO interval was used and a token was provided for each interval in which no nail biting occurred. If nail biting occurred at any point during the interval the interval was reset, and a token was not delivered. Tokens were exchanged for 15 min of a leisure activity chosen by the participant. The participant took pictures of the condition of fingernails at the beginning of each DRO interval, after each token delivery, and after each time nails were bit. Each instance of nail biting was also tallied in a notebook by the participant. Results were that the package of DRO, self-monitoring, and permanent product data were effective for decreasing nail biting. Craig also concluded that the treatment integrity procedure with permanent product may have led to greater adherence to the self-monitoring program. Additionally, treatment integrity checks by means of social support such as a family member may enhance ABA techniques and reduce response effort required by an outside party. Frequently BCBAs have a limited amount of time to spend with their clients. By having family members collect or check data collection, a greater amount of data can be collected without requiring a substantial amount of time from the BCBA. However, it is important to ensure that the data being collected is accurate as to not portray a misleading picture of behavior. Including permanent product measurement when feasible is an excellent way to monitor and check accuracy of data collection.

Because self-monitoring has been frequently assessed as a method to increase treatment integrity but as part of an intervention package, Mouzakitis, Codding, and Tryon (2015)

conducted a study to assess whether self-monitoring (referred to as *SM*) alone would increase treatment integrity to an acceptable level of at least 80%. They also evaluated whether performance feedback (PFB) would be required to increase or maintain treatment integrity. The researchers examined the link between treatment integrity and student behavior change, and generalization of treatment integrity to behavior intervention plans (BIPs) with novel students. Participants were four special education teachers and eight students put into pairs. Each teacher was paired with a target and a generalization student. Each student engaged in escape-maintained, off-task behavior as determined by a functional behavior assessment (FBA). Independent variables were self-monitoring and self-monitoring with written PFB. The dependent variables for teachers were treatment integrity for implementation of the target student's BIP, and treatment integrity for implementation of the generalization student's BIP. The dependent variables for students were percentage of on-task behavior for the target students and percentage of on-task behavior for the generalization students. A multiple-baseline design across teachers was used to evaluate the effects of this study. Phases occurred in the following order: baseline, initial BIP training, treatment integrity baseline, self-monitoring training, self-monitoring, self-monitoring plus PFB, and return to self-monitoring. Each teacher was observed in 30-min increments, two to three times a week. Results of this study were that self-monitoring alone improved treatment integrity for behavior plan implementation for three out of four teachers, but not to criterion levels; to meet the criterion, PFB needed to be added. Additionally, treatment integrity was maintained when the intervention was faded to self-monitoring only for two teachers. All teachers successfully implemented the BIP with their target and generalization student, and higher levels of treatment integrity were associated with higher levels of on-task behavior. These results suggest that, although PFB was required to meet criterion levels, self-

monitoring may be a useful strategy for maintaining results while fading consultant support. Limitations of this study as reported by the researchers are that experience in teaching and implementing behavior plans was not held constant and results may not be generalizable for behavior problems not maintained by escape.

The use of technology such as videos has also been used to increase treatment integrity. Pelletier, McNamara, Braga-Kenyon, and Ahearn (2010) evaluated a video self-monitoring treatment package to increase treatment integrity for one student's BIP. Participants were three teachers working with a 15-year-old student with autism and behavior problems. Before the intervention, teachers demonstrated less than 80% levels of treatment integrity. During baseline, the teachers were provided written guidelines, observed a more-experienced teacher implement the BIP were observed by a senior staff member, and received verbal feedback and written performance feedback twice a year. Steps of this intervention package were a pretraining of a mock-implementation video scored with feedback, then scoring a video of themselves implementing the student's BIP, the experimenter scoring the same video and providing verbal feedback regarding their integrity, being videotaping to assess improvement or any changes, and a follow-up probe session for two of three participants. A nonconcurrent, multiple-baseline across-participants design was used to evaluate experimental control. Results were that the video self-monitoring treatment package was successful at increasing correct BIP implementation for three staff members (100%, 90-100%, and 86%), and maintaining results at a 1-month follow up for two staff members (86% and 100%). Therefore, using videos may be a viable option to increase the effectiveness of self-monitoring as a treatment integrity intervention.

Videos have also been used for modeling purposes. DiGennaro-Reed, Codding, Catania, and Maguire (2010) conducted a study on individualized video modeling and direct performance

feedback to increase treatment integrity for treatment of problem behavior of three teachers. The independent variable measured for this study was the intervention put in place to increase treatment integrity, consisting of video modeling and performance feedback. The dependent variable measured for this study was treatment integrity in carrying out a behavioral intervention, specifically, percentage of steps implemented correctly. A concurrent, multiple-baseline design across participants was used to evaluate the effects of individualized video modeling and individualized video modeling plus performance feedback on treatment integrity. Treatment integrity was assessed through 15 min direct observations of the participant implementing a student's behavioral intervention in a classroom or residence. During baseline, feedback was not provided. During the individualized video modeling condition, each participant viewed an individualized instructional video that depicted a model demonstrating accurate implementation of all the intervention steps and were then asked to implement the intervention with their students. During the individualized video modeling plus performance feedback condition, the experimenter provided verbal feedback about prior-session performance before playing the video and highlighted the relevant segments while playing the video. For example, the experimenter paused the tape at relevant points and asked the participant to pay attention to the next segment because the errors exhibited in the previous teaching session would be addressed. If no errors occurred, the experimenter stated this and played the video without pausing. One week after ending the feedback conditions, a follow up probe was conducted. The key findings were that, although video modeling improved treatment integrity, when they added performance feedback, treatment integrity improved to higher levels, and the combination was found to be socially acceptable. Individualized video modeling combined with direct performance feedback was not only successful, but more successful than either method was alone, and therefore should

be used in combination. Limitations of this study that may have affected results are participant reactivity and that there was an inconsistent delay between the participant viewing the video model and when the observation was conducted.

Rehearsal is another method that has been evaluated for its effectiveness in increasing treatment integrity. Jenkins and DiGennaro-Reed (2016) conducted a parametric analysis of rehearsal opportunities for implementation of a FA by 18 undergraduate students without prior experience. The primary dependent variable was procedural integrity of FA implementation, specifically for attention, tangible, and escape conditions. Integrity was measured using an eight-step FA task analyses. For each step in the task analyses, correct implementation included appropriate delivery of an antecedent, consequence, or presence of specific stimuli. Procedural integrity was calculated as a percent by dividing the number of correctly implemented steps by the total number of possible steps and multiplying by 100. FA conditions were implemented for 5 min at a time during 60-min lab visits, 2 to 3 times a week. A multi-element design within a multiple-baseline design across participants was used to evaluate experimental control. Five phases were included to analyze different levels of rehearsal within a BST package for conducting functional analyses. These five phases were instructions, video modeling, rehearsal analysis, supplemental rehearsal, and follow up. In the instructions phase, participants read procedural instructions, completed a nine-item quiz, and read a client description. Participants implemented FA conditions with the client, but no feedback was provided. For video modeling, participants viewed a 1-min video model of each FA condition and then implemented FA conditions with the client. No feedback was provided. During rehearsal analysis, participants rehearsed each condition with the client and received feedback on their performance. Participants were assigned to one of three rehearsal analysis conditions, each condition requiring a different

number of rehearsal opportunities, specifically, one, three or ten times. Participants were provided feedback after each rehearsal opportunity unless they met mastery criterion. A supplemental rehearsal with feedback was provided for each condition in which participants did not meet the mastery criterion in the rehearsal analysis phase. The mastery criterion was set at 100% procedural integrity across two consecutive sessions for each condition. Follow up was then provided 2 weeks after the mastery criterion was met. Results of this study were that rehearsal with feedback effectively increased treatment integrity regardless of the rehearsal condition to which participants were assigned, almost all (16 out of 18) participants required supplemental rehearsals to meet and maintain criterion level for at least one condition, and integrity was maintained once the criterion was met. It is important to note that some participants met the criterion with little training and others required the full BST package. Therefore, trainings should be individualized. An efficiency analysis was also conducted to identify the most efficient rehearsal condition. This analysis was conducted by adding up the total number of rehearsal opportunities and number of seconds each participant spent in rehearsal and feedback until they met the set criterion. The most efficacious and efficient training across participants was BST with single rehearsals until participants met the set criterion. Therefore, BST is effective in increasing treatment integrity when including a rehearsal component. However, a reported limitation of this study is that the mastery criterion may have been set too high, leading to additional rehearsal and feedback sessions needed to reach the criterion. Second, the lack of feedback in the supplemental rehearsal condition may have served as indirect feedback, alerting the participant that they met mastery criteria for the other conditions. Third, some participants worked with a single client and others worked with more than one client, causing an inconsistency that may have affected treatment integrity. Another inconsistency that may have

affected results is that the number of rehearsal opportunities varied across participants allowing some participants to have more training than others.

Pantermuehl and Lechago (2015) conducted a study comparing feedback, another component of BST, on the treatment integrity of staff working with children with autism through different observation methods. Participants for this study were three therapists working at an outpatient ABA clinic. Methods of observation were baseline or covert observations, *in vivo* observations, and Skype observations. During baseline, a camcorder was set up in the therapists' classrooms, but they did not know when they were being recorded. During *in vivo* observations, a supervisor told therapists that they would be observed for 15 min but the therapist was unaware of data collection. Feedback was provided during *in vivo* observations. During Skype observations, a supervisor called therapists from a remote location and the therapist knew they were being observed, but similarly to the *in vivo* condition, they were unaware that data was being collected on treatment integrity. Feedback was provided during Skype observations. The experimental design used was a multiple-baseline, across-participants design with an embedded multi-element design. The primary dependent variable was therapist treatment integrity of an error-correction procedure using a most to least prompt hierarchy. The independent variable was the intervention consisting of the different methods of observation. A global treatment-integrity score was calculated by recording data on the number of steps implemented correctly for each trial and dividing this by the total number of steps of the error-correction procedure. Results were that both *in vivo* and Skype observation methods resulted in similarly high levels of treatment integrity for participants; with low treatment integrity in baseline. This is important because both telehealth and in person methods of observation and feedback may be effective for increasing treatment integrity levels for practitioners, which may increase flexibility and reduce response

effort required for supervisors. A reported limitation of this study is that Skype observations were never conducted more than 45 min away, which may reduce issues with connectivity and clarity.

DiGennaro, Martens, and Kleinmann (2007) evaluated a package intervention to increase treatment integrity in which four special-education teachers participated. Their students had various diagnoses, behavior concerns, and were on different medications, but all engaged in off-task behavior. Teachers participated in interviews, training, implementation of plans, and performance feedback on individualized behavior plans for their students. The dependent variables were treatment integrity, treatment effectiveness, and treatment acceptability. Treatment integrity was observed daily during 15-min observations of the teacher implementing the intervention. The number of treatment steps that were implemented accurately during the observation was divided by the total number of steps and converted to a percentage for a single treatment-integrity measure. Treatment effectiveness was measured by collecting data on student problem behavior in daily 15-minute observations divided into 10-s intervals and recording if the behavior occurred at any time during the interval. This was then converted to a single percent for each session by dividing the number of intervals in which problem behaviors occurred by the total number of intervals. Treatment acceptability was measured at the end of this study by having teachers complete the Intervention Rating Profile-15 (IRP-15; Martens, Witt, Elliot, & Darveaux, 1985). A multiple-baseline design across pairs was utilized to evaluate results. Training consisted of didactic instruction, modeling, coaching, and immediate corrective feedback, and was conducted until teachers implemented their assigned plan with 100% integrity on two consecutive occasions. Following initial training, teachers implemented the plan without assistance or feedback from the consultant. Next, a goal setting and performance feedback phase

was put in place in which teachers received feedback regarding their student's performance but not their own implementation. Next, a phase was implemented in which teacher performance feedback was provided as well as directed rehearsal with meeting cancellation. If a teacher did not implement the plan with 100% integrity, they were required to attend a meeting where they practiced missed steps; however, if they implemented the plan with 100% accuracy, they did not have to attend this meeting. Fading was conducted where, if integrity was maintained at 100% for three consecutive sessions, the scheduled was thinned to once every other day, then once per week, and finally once every two weeks. Results were that goal setting and performance feedback for student performance increased treatment integrity for only one out of four teacher participants. Teacher feedback and directed rehearsal with meeting cancellation resulted in improved treatment integrity for teachers with all four teachers meeting mastery criteria within five sessions. When the intervention was faded, all teachers maintained high levels, above 90%, of treatment integrity. All students demonstrated improvement in their target behavior with introduction of the intervention and decreases in problem behavior were maintained during fading of the intervention for teachers for three out of four students. The intervention was rated as generally accepted on the IRP-15 and goal setting and student performance feedback was rated as slightly more acceptable than teacher performance feedback and directed rehearsal with meeting cancellation. An important result to highlight is that teachers could implement plans with 100% integrity on two consecutive occasions following training consisting of didactic instruction, modeling, answering questions, coaching, and immediate corrective feedback, which is a training package similar to BST. However, when this training ended, and teachers only received feedback regarding their student's performance, not their own implementation, treatment integrity decreased substantially. When teacher feedback and directed rehearsal with

meeting cancellation was implemented, treatment integrity increased, reaching mastery criteria of three consecutive days with 100% integrity. Results were maintained despite fading of performance feedback. This strengthens the point that ongoing performance feedback, even when faded to as little as once every other week or once every 2 weeks, may be required to maintain treatment integrity for teachers. A limitation of this study is that it is unknown which components of the intervention specifically produced increases in treatment integrity as they were combined in an intervention package.

Codding, Livanis, Pace, and Vaca (2008) conducted a study to assess if performance feedback would have a positive effect on treatment integrity. Three teachers in a self-contained program participated in this study. The variables measured were the percentage of behavior plan components that were implemented as written, and the presence or absence of the observer across sessions. The observer was located inside the classroom during 50% of the observations (observer present) and was absent from the classroom but observed through a one-way window for the other 50% of observations (observer absent). Baseline consisted of observing each teacher and completing an integrity data sheet under both observer-present and observer-absent conditions, without providing feedback. Performance feedback was implemented after stable performance in baseline was demonstrated for both conditions. The observer provided feedback on all steps observed. Feedback included praise for steps followed as written and corrective feedback for steps that were followed sometimes or not at all. A multiple-baseline design across staff members with alternating treatments was selected to evaluate the effectiveness of performance feedback across observer-present and observer-absent conditions. Treatment integrity was measured by recording if each component of the behavior plan was implemented as written, not implemented as written, or if the teacher did not have an opportunity to implement

the component and therefore it could not be observed. The number of components implemented as written was then divided by the total number of components that the teacher had an opportunity to implement. This was then converted to a percentage by multiplying by 100%. Results of this study were that performance feedback led to improvements in correct implementation for all teachers reaching 100% treatment integrity. Whether the observer was present or absent had no effect on results. Therefore, reactivity caused by the presence of the observer did not influence on the positive effects of performance feedback leading the researchers to conclude that performance feedback is important in improving plan implementation regardless of whether or not the consultant or observer is present. Additionally, a 10-item social validity questionnaire was provided to teachers at the end of this study and indicated positive views of performance feedback among teachers. Limitations of this study as reported by the researchers are that the teachers and the observer were not blind to the purpose of the study and that student behavior was not observed at the same time as teacher behavior.

Data collection by staff has also been used as a method to increase treatment integrity. Howard, Burke, and Allen (2013) evaluated whether having supervisors collect data on their staff's treatment integrity increases supervisor treatment integrity for delivery of behavior specific praise. A type of reactivity termed the "observer effect" has been documented in previous literature and relates to the researcher's purpose. By simply observing and collecting data on a behavioral intervention, treatment integrity of implementation of the intervention can improve. Supervisors for this study were five individuals working at a day treatment program responsible for implementing academic curriculum, conducting daily activities, providing direct ABA services to six to eight children. Supervisors were also responsible for staff observations, collecting data, and monitoring staff's treatment integrity of implementation of ABA services to

children. The primary dependent variable was frequency of the delivery of behavior specific praise calculated in rate per minute. Behavior specific praise was chosen as the primary dependent variable due to its importance at the day treatment program and for children's BIPs. Secondary dependent variables were the frequency of general praise per minute, frequency of time-out, and praise-to-correction ratios. A multiple baseline with reversal design was used to evaluate a multistep procedure for this study. First, baseline was conducted to assess the typical level of behavior specific praise for staff. Data were collected by researchers in staff's classrooms during a minimum of three 10-min observations across separate days. Next the intervention on data collection for supervisors took place. This intervention was broken into three components. The first component consisted of individual preintervention meetings with the administrator and each supervisor to discuss the importance of the behavior specific praise In the second component, the participant or supervisor was provided a data sheet with examples, operational definitions, goals, and areas for the supervisor to complete each day. In the third component, the participant or supervisor collected treatment-integrity data on staff's delivery of the behavior specific praise. When the intervention was introduced, three participants showed marked, stable, and replicable changes with an increase in behavior specific praise representing a high level of treatment integrity.

Following the third component, the intervention was withdrawn (Howard et al., 2013). When the intervention was withdrawn, two participants (of the three who showed improvements in behavior specific praise from the intervention) demonstrated decreased rates for behavior specific praise. After withdrawal, there was a return to intervention that was the same as the original procedures except that the script used in the preintervention meeting was slightly modified to reduce redundancy. During the return to intervention, the same two participants

demonstrated recovered rates of behavior specific praise. Across components there were corresponding changes in praise to correction ratios (to behavior specific praise rates).

Last, the supervisors went through a debriefing and consent (Howard et al., 2013). During the debriefing participants were provided a social validity assessment. This assessment indicated that data collection was an acceptable approach to increasing treatment integrity. Therefore, overall, the results of this study indicated that data collection can be an effective way to improve treatment integrity for supervisors without intervening on them directly. However, it is important to note that participants did not reach the target rate of BSP and variable results were reported among staff. External variables that may have contributed to the results are behaviors of students and staff value in their employment. There are many benefits of using data collection to improve treatment integrity in an applied setting. For example, data collection was completed by employees independent of research involvement and also reduced the amount of time and effort required by the administrator.

These studies demonstrated methods that are effective in increasing treatment integrity levels for staff implementation of behavior intervention programs. Self-monitoring alone and self-monitoring combined with other procedures (such as video modeling, DRO, collection of permanent product data, and monitoring of data collection by family members) and performance feedback has been demonstrated as effective to increase treatment integrity (Craig, 2010; Pelletier et al., 2010; Plavnick et al., 2010). Self-monitoring has also been demonstrated to be successful in maintaining levels of treatment integrity while fading consultant support (Mouzakitis, Codding, & Tryon, 2015). By adding a permanent product measure or data collection checks by another individual, such as a family member, accuracy can be better ensured. Performance feedback has also been demonstrated consistently as effective for

increasing treatment integrity, both alone and in combination with other methods such as individualized video modeling (Codding et al., 2008; DiGennaro-Reed et al., 2007, 2010; Pantermuehl & Lechago, 2015). When performance was combined with individualized video modeling, the combination was more successful in increasing treatment integrity than either was alone (DiGennaro-Reed et al., 2010)

BST has also consistently been an effective method of increasing treatment integrity and when compared to rehearsals and performance feedback, the full package of BST was the most effective and efficient to increase treatment integrity of FA implementation (Jenkins & DiGennaro-Reed, 2016). Additionally, performance feedback has been demonstrated as an acceptable method among staff and as effective for maintaining increases in treatment integrity while consultant support is faded out (DiGennaro-Reed et al., 2007). Some limitations of these studies include: (a) lack of examination of specific intervention components as compared to examining whole interventions, (b) lack of consistent background variables such as experience implementing BIPs and value in employment, (c) limited generality due to small sample sizes or behavior problems maintained by participant reactivity, (d) mastery criteria that is set too high (for example, at 100%), (e) inconsistencies in terms of how many clients each participant is responsible for (f) lack of client observations to monitor behavioral progress, and (g) inconsistencies in length and frequency of training sessions for participants (Codding et al., 2008; Craig, 2010; DiGennaro-Reed et al., 2007, 2010; Howard et al., 2013; Jenkins & DiGennaro-Reed, 2016; Miller et al., 2014; Pantermuehl & Lechago, 2015; Pelletier et al., 2010; Plavnick et al., 2010).

Social Validity of Treatment Integrity

Sterling-Turner and Watson (2002) conducted a study examining the correlation between treatment acceptability and treatment integrity. They described five variables that may be related to treatment integrity: difficulty of the intervention, time requirement, required number of implementation agents, the strain on resources, and acceptability of the intervention. *Treatment acceptability*, a type of social validity, includes how fair, reasonable, appropriate, and unobtrusive an individual perceives a treatment to be (Kazdin, 1980). Positive, or *accelerative,* interventions, such as reinforcement procedures, are more acceptable than negative, or *reductive,* interventions, such as punishment procedures (Blampied & Kahan, 1992; Elliott, Witt, Galvin, & Peterson, 1984; Kazdin, 1984; Kalfus & Burk, 1989; Miller & Kelley, 1992; Spreat, Lipinski, Dickerson, Nass, & Dorsey, 1989). Other factors that may affect acceptability of interventions are how severe the problem behavior is, the amount of time required to implement the intervention with integrity, how complex the treatment is, and how effective the individual perceives the intervention will be. Studies that have measured the correlation between acceptability and integrity of interventions have been methodologically limited as they have relied heavily on self-report. The Treatment Acceptability Rating Form (TARF) has been used to measure acceptability of an intervention by parents. The TARF assesses several areas related to treatment acceptability including the effort and time required for intervention implementation, how effective individuals perceive the intervention to be, their willingness to carry out the intervention, and by what degree the intervention is accepted by implementation agents.

Sterling-Turner and Watson (2002) had 75 psychology undergraduate students rate the acceptability of a multicomponent intervention including DRO, mild punishment, increasing client awareness, data collection, and data collection using the IRP-15. Questions asked on this

rating profile related to participants' willingness to implement the intervention, how appropriate and fair he or she perceived the intervention to be, cost–benefit ratio for the client, and what others would think of the intervention. General acceptability was also measured as a single factor and Spearman rank-order correlations were calculated to determine the magnitude of the relationship between treatment acceptability and treatment integrity. Results indicated that treatment acceptability did not have a strong correlation with treatment integrity. However, a limitation of this study that may have affected these results is the use of rating scales to measure treatment acceptability.

Strohmeier, Mulé, and Luiselli (2014) also addressed social validity issues in relation to treatment integrity. The researchers surveyed 44 direct-service staff at a day and residential school setting for children and youth with autism and related developmental disabilities about their experiences implementing behavior support plans and recommendations that could improve treatment integrity. Demographics questions asked about age, gender, ethnicity, education, years of teaching experience, experience with students with autism, and experience implementing behavior support plans were asked in addition to questions about performance feedback, online training, ratings of positive and negative reinforcement, and ratings of effectiveness of the four listed interventions. Most participants reported that they had a lot of experience and understood the behavior support plans of their students very well. Their ratings of the feedback they received were variable, but the majority reported that they would respond very well to receiving performance feedback and rated performance feedback as the most effective training procedure. This indicates that performance feedback is the most acceptable and socially valid procedure. This may partially explain why, in previous research, performance feedback has been

so effective; therefore, it may be beneficial to have staff choose the training method they would like to experience for increasing treatment integrity.

These studies demonstrated methods that have been effective in increasing treatment integrity levels for staff implementation of behavior intervention programs. Self-monitoring alone and self-monitoring combined with other procedures (such as video modeling, DRO, collection of permanent product data, and monitoring of data collection by family members, and performance feedback; Craig, 2010; Pelletier et al., 2010; Plavnick et al., 2010) have been demonstrated as effective methods for increase treatment integrity. Self-monitoring has also been shown to be successful in maintaining levels of treatment integrity while fading consultant support. By adding a permanent product measure or data collection checks by another individual, such as a family member, accuracy can be better ensured. Performance feedback has also been demonstrated consistently as effective for increasing treatment integrity, both alone and in combination with other methods such as individualized video modeling (Codding et al., 2008; DiGennaro-Reed et al., 2007, 2010; Pantermuehl & Lechago, 2015). When performance was combined with individualized video modeling, the combination was more successful in increasing treatment integrity than either was alone (DiGennaro-Reed et al., 2010)

BST has also been demonstrated as an effective method of increasing treatment integrity and when compared to rehearsals and performance feedback, the full package of BST was the most effective and efficient to increase treatment integrity of FA implementation (Jenkins & DiGennaro-Reed, 2016). Additionally, performance feedback has been demonstrated as an acceptable method among staff and as effective for maintaining increases in treatment integrity while consultant support is faded out (DiGennaro-Reed et al., 2007). Other effective methods of increasing treatment integrity found in the previous literature are data collection on other

individuals' behavior and a lottery-based incentive system (Howard et al., 2013; Miller et al., 2014).

Limitations of these studies include lack of examination of specific intervention components as compared to examining whole interventions, lack of consistent background variables such as experience implementing BIPs and value in employment, limited generality due to small sample sizes or behavior problems maintained by a single function, participant reactivity, mastery criteria set too high (for example, at 100%), inconsistencies in terms of how many clients each participant was responsible for, lack of client observations to monitor behavioral progress, and inconsistencies in length and frequency of training sessions for participants (Codding et al., 2008; Craig, 2010; DiGennaro-Reed et al., 2007, 2010; Howard et al., 2013; Jenkins & DiGennaro-Reed, 2016; Miller et al., 2014; Pantermuehl & Lechago, 2015; Pelletier et al., 2010; Plavnick et al., 2010).

Treatment Integrity Research Populations

Although there remains a limited amount of research on how to increase treatment integrity for implementation agents, especially when compared to the amount of research on how to implement behavioral procedures, most of the research is on staff behavior, including the behavior of teachers and individuals working with students with disabilities. The available literature has demonstrated that for staff behavior there are several ways to increase levels of treatment integrity for behavioral procedures and the importance of doing so. High levels of treatment integrity can be obtained using BST, performance feedback, rehearsal, modeling, and other procedures and leads to more effective and efficient interventions for clients. Staff and teachers can be successful implementing interventions when the right procedures are put in place. However, as opportunities remain for BCBAs in the home setting, parents also need to be

taught how to implement behavioral interventions with high levels of integrity. When ABA services are provided in client's homes, BCBAs are often provided only a small number of hours for direct work with clients and spend a majority of their time training parents to implement interventions. However, to train parents successfully to implement interventions with high integrity, research needs to be conducted in this area. Behavior analysts need to know if the same interventions that are successful for increasing integrity levels for staff implementation work for parent interventions and the unique variables associated with doing so. It is likely that parents will be able to implement interventions with high levels of integrity; however, research needs to be conducted in this area to develop the best practices available. There is a limited amount of research on treatment integrity for staff implementation of interventions; however, research on treatment integrity is even more limited for parent implementation. As of the writing of this paper, only one article on treatment integrity for parent implementation of intervention was found.

Arkoosh et al. (2007) evaluated treatment integrity outcomes for functional communication training (FCT) with parents. Parents of five participants participated in this study and were included because their children engaged in problem behavior (self-injury, aggression, tantrums, or destruction) maintained by positive or negative reinforcement contingencies (including attention and escape). Parents implemented FCT and sessions were videotaped. Three categories of parents' behavior were recorded to assess treatment integrity: reinforcement of mands, reinforcement of other behaviors, and application of the BIP to reduce problem behaviors. Treatment integrity was measured by: contingent delivery of a reinforcer within 12 s of a trained mand, contingent delivery of a reinforcer within 6 s of an appropriate social interaction, or contingent delivery of an aversive event within 6 s of a child's inappropriate

response. The results of this study were long-term reduction of problem behaviors for 3 out of 5 participants. Children who did not achieve long term reduction also received low integrity of treatment. Additionally, treatment integrity levels for problem behavior reduction was low for all clients, and much higher for increasing positive behaviors. The implications of this study were that this was the first analysis of treatment integrity for parent implementation of all procedures in the home setting, that results demonstrated that higher treatment integrity was related to long-term reduction of behavior problems, and that treatment integrity of parent implementation was higher for increasing positive behaviors as compared to decreasing negative or problematic behaviors. A limitation of this study is that different levels of treatment integrity were not manipulated during this study so effects of levels of treatment integrity are descriptive rather than experimental. Areas for future research as suggested by Arkoosh et al. (2007) include evaluating appropriate levels of treatment integrity for different interventions and the effectiveness of the experimental manipulation of treatment integrity levels.

Summary and Transition

As previously discussed, it is imperative to evaluate treatment integrity when conducting studies where an intervention is implemented. Without doing so, it is difficult to replicate research and to determine if a functional relation was obtained. Methods to increase treatment integrity are also important to evaluate so that practitioners can ensure that parents and staff are implementing interventions as intended. This is the only way that behavior analysts can ensure that behavior changes are due to interventions. Additionally, behavior interventions are written based on thorough assessments by behavior analysts and therefore should be implemented as written. Unfortunately, the research continues to be limited and there is substantial room for growth.

Chapter 3: Research Design and Method

Chapter Overview

The main purpose of this study is to evaluate the effects of BST on treatment integrity for parent implementation of their child's skill acquisition programming. A secondary purpose is to determine if there is a correlation between child skill acquisition and treatment integrity of parent implementation. This chapter will review the participants and setting for this study, materials, response measurement, interobserver agreement, experimental design, and procedures. The primary researcher, a board certified behavior analyst (BCBA), will conduct training with three parent–child dyads using BST and a research assistant will collect IOA data to ensure reliability of data and procedures. A multiple-baseline design across participants will be used to evaluate the effects of BST on treatment integrity.

Methods

Participants and Setting

Research team. One BCBA served as the lead researcher for this study and implemented ongoing training and data collection procedures with parents. One additional BCBA served as a research assistant and collected reliability data, which was compared to the research lead's data. All members of the research team completed the human subjects protection training and were trained on all procedures for this study including data collection for BIPs, DTT, and treatment integrity.

Parent–child dyads. There were three parent–child dyads who participated in this study. Inclusion criteria included: parents had a child who demonstrated behavior deficits such as difficulty with social skills, attention, verbal behavior, and emotional regulation and benefited from applied behavior analysis (ABA) programming, particularly DTT. High-intensity and high-

risk behaviors such as severe self-injurious behavior were not included for this study due to the dangerous nature of these behaviors and the need to intervene immediately to reduce risk of harm. Additionally, parents had less than one year of experience implementing ABA programs and demonstrated treatment-integrity scores below 70%. Parents were recruited from an ABA company based out of Florida. Each parent was responsible for implementation of their child's skill acquisition procedures or DTT program. There were no restrictions regarding the amount of prior ABA therapeutic services received by the child.

All trainings and intervention sessions took place in the participants' homes in an area or room designated for behavioral training or specified by the parent. Depending on the behavior plan in place for children, guidelines existed regarding the set-up of the environment, such as limiting distracting stimuli, removal of dangerous objects, and ensuring the area is well lit and temperature regulated.

Materials

Behavior intervention plans (BIPs). Children had individualized BIPs that were written by the BCBA. Each included at least one preparation category (necessary materials for plan), one antecedent strategy, and one reactive strategy. Each behavior plan also included an operational definition for the target behavior and directions on procedures for implementation. A paper copy of the BIP was provided to parents.

Task materials and reinforcers. During all training sessions, individualized task materials (such visual stimulus cards and data collection forms) and reinforcers (such as edibles and tokens) were present.

Dependent Variable, Response Measurement, and Interobserver Agreement

The primary dependent variables were global and component treatment integrity for parent implementation. Treatment integrity was defined as the accurate implementation of all components of an intervention based on the treatment integrity checklist. *Global treatment integrity* was defined as the degree to which the parent implemented the entire intervention as intended. *Component treatment integrity* was defined as the degree to which the parent implemented each component response of the intervention. Data were collected by the BCBA using the treatment integrity checklist on both global and component integrity.

Global treatment integrity was calculated by recording data on the number of steps implemented correctly for each trial and dividing this by the total number of steps of the error-correction procedure. An individualized treatment integrity checklist was created for each intervention. The treatment integrity checklist included: preparation (whether all necessary materials were readily available and where they should be), antecedent strategies (whether they were carried out with accuracy), and reactive strategies (whether they were carried out with accuracy) to be completed by the BCBA for parent implementation of interventions. One of four levels of implementation integrity were scored for each component on the checklist. These included implemented as written, implemented with deviation, not implemented, and no opportunity to observe. A percentage was calculated by dividing the number of components implemented as written by the total number of components observed and multiplied by 100% for the global treatment-integrity score.

Component treatment integrity was calculated by dividing the number of correctly implemented responses within a single component by the total number of opportunities to implement each component response within a session and converting to a percent. Component

integrity was calculated by dividing the number of correctly implemented responses within a single component by the total number of opportunities to implement each component response within a session and converting to a percent.

Data were also collected on child behavior by the BCBA, specifically progress made for the specific behavior intervention chosen. Specific child behavior depended on the program put in place (for example correct and incorrect responding for a DTT program). For Child Participants 1 and 2, the target behaviors were sight word recognition (i.e. touching a target word on stimulus card when vocally presented). For Child Participant 3, the target behavior was recognition of Spanish words (i.e. touching a target word or phrase on stimulus card written in Spanish when vocally presented in English). Component integrity data were graphed with the intervention component along the *x*-axis and percentage of opportunities implemented correctly along the *y*-axis (see Cook et al., 2015). Each step was represented by individual bars and baseline, intervention, and maintenance phases were color-coded in the key or legend. Global-integrity data were graphed with sessions along the *x*-axis and percent treatment integrity. Child accuracy data were graphed with sessions along the *x*-axis and percent accuracy along the *y*-axis. Each graph was labeled by parent–child dyad (see DiGennaro et al., 2007).

Interobserver agreement (IOA). All sessions were video recorded so that a second trained observer could independently collect data during a minimum of 30% of sessions for each phase to calculate interobserver agreement (IOA). IOA was calculated using the point-by-point method by comparing each observer's data on the treatment integrity checklist per component. Agreements for each component were defined as each observer scoring the same response for each component. Components scored as agreements were summed, divided by the total number of components and multiplied by 100%. IOA for this study was 96.85%.

Experimental Design

A multiple-baseline design across parent–child dyads was used to evaluate experimental control. A multiple-baseline across participants was used to eliminate the requirement of a withdrawal strategy under the assumption that treatment removal would be unlikely to result in a return to levels of treatment integrity below its baseline level. A minimum of three sessions were conducted for baseline. Phase changes were made once stability in data had been achieved for the previous dyad. Specifically, baseline began with the first dyad at the start of the study and once behavioral stability was demonstrated BST began for that dyad. Subsequently, once BST began for the first dyad, baseline continued for the second dyad for at least two more sessions to measure what changes, if any, occurred between the first dyad and the second dyad (comparison between the two baselines). When data remained stable, BST was then implemented with the second dyad and the process repeated for the third dyad. BST was ended when dyads met mastery criteria. Once mastery criteria were achieved, dyads moved into the maintenance phase. Internal validity of this design was ensured by multiple replications of the intervention implemented at different times across dyads.

Procedure

Sessions were conducted two to three times per week for 1 hr with the parent, child, and BCBA present. The times were set based on parental availability and approval. For portions of the sessions, the child was not required to be in the training area, for example, when the BCBA reviewed the BIP with the parent. The training room was set up with all materials required for the intervention (for example program sheet, table and chairs, data forms, and any materials to be used with the child). Except for during baseline, the BCBA provided opportunities for the parent

to ask questions at the beginning and end of each session. During each observation, the BCBA collected treatment-integrity data using the checklist described above.

Pre-behavioral-skills training (BST; baseline). During baseline sessions, parents read their child's BIP and implemented procedures to the best of their ability for 10 trials. Parents were not provided any additional feedback or training on behavior plans. Opportunities to ask questions were not provided.

Behavioral skills training (BST). All steps of the BST package (described below) were implemented until the parents meet mastery criteria of 90% or higher treatment integrity for three consecutive sessions. Performance feedback, including corrections and praise, were delivered across all steps of training and implementation. Parents were provided the opportunity to ask questions at the end of each treatment component.

Instruction presentation. During the instruction component, the BCBA vocally reviewed and described each component of the child's BIP with the parent.

Modeling. During the modeling component, the BCBA demonstrated how to correctly implement the procedures with the child for three consecutive trials.

Role-play. During the role-play component, the parent demonstrated how to correctly implement the procedures by implementing three consecutive trials with the BCBA pretending to be the child. Feedback on their performance was provided after each trial.

Practice. During the practice component, the parent implemented the procedures by implementing ten consecutive trials with their child. Immediately after demonstration with the client, the BCBA delivered performance feedback based on the parent's implementation of all treatment components. The full BST package continued until the parent achieved 90% or higher correct implementation for three consecutive sessions.

Post-behavioral-skills training (BST). During posttraining, the BCBA instructed the parent to implement 10 trials with his or her child. The BCBA did not provide any additional training or feedback during this time. Once a parent achieved 90% or higher during this session, BST training no longer occurred unless treatment integrity fell below 80% for two consecutive sessions. If this occurred, a remedial BST session was conducted by the BCBA.

Validity

Threats to internal validity are those factors that have the potential to provide alternate explanations for the observed effects. Threats to internal validity for single-subject design include history, maturation, testing, instrumentation, and mortality. History includes intervening effects that may influence measurement outcomes. For example, if a participant is learning a similar skill in school as they are being taught in the home setting, this may have been a potential threat to the validity of the study; therefore, before a skill was chosen for acquisition, the BCBA ensured that the participant was not learning it elsewhere. Maturation includes changes in participant behavior that is extraneous to manipulations of the teaching program. This study was completed within two months, therefore significant maturation is unlikely. Testing includes influences of testing, observation, or measurement of the dependent variable. To ensure that testing did not affect internal validity, the baseline in which parents implemented procedures to the best of their ability was kept as short as possible while still abiding by the multiple-baseline design parameters. To avoid issues related to instrumentation, in which changes or inconsistencies in how phenomena are assessed in a study negative impact validity, IOA was calculated for a minimum of 30% of sessions for each phase. Direct observation was used as measurement to avoid issues related to credibility.

Data Processing

Microsoft Excel was used to record and graph data. Visual analysis of data was conducted to determine if a functional relation existed between the dependent and independent variables, specifically BST and treatment integrity for parent implementation.

Ethical Assurances

Participants were recruited through Creative Solutions in Home Health, a home-based ABA provider located in Florida. Participants were required to provide written consent before starting this study and all procedures and were provided at no cost to participants. Additionally, participants were free to drop out of the study at any point. Sessions were video-recorded for IOA purposes, but all recordings were stored on a password-protected drive. The primary investigator, research assistant, and research supervisor had access to these video recordings only for the purpose of this research study. Recordings will be destroyed once the information needed for research manuscript have been obtained and no identifying information will be associated with the audio or video recording or the transcript. These data were kept secure and will be destroyed by the primary researcher. An associated risk of this study was an invasion of participants' privacy as the researcher observed the parent and their child's behavior within their personal home. However, this risk was minimized by the parent controlling when and where the researcher conducted sessions. Additionally, identifiable information remained confidential with only the primary researcher having access to this information. Only personal information pertaining to this study were collected and all identifiable information was coded as early in the activity as possible and securely stored so that only the researcher had access to it. The identity of participants was not and will not be released without their consent. In order to guard the anonymity and confidentiality of participant responses and personal information, the following

steps were taken. Personally identifiable information was replaced with research identification codes. Contact lists, recruitment records, assessments, and any other documents that contained personally identifiable information were kept separately from shared data forms and graphs and will be destroyed when no longer needed for the research. Files containing electronic data were password-protected and encrypted when data were transferred or transported. Research data were stored securely in locked cabinets. Electronic data were stored in password-protected computers and files. Files containing electronic data were closed when computers are left unattended. Consent forms were stored securely in locked cabinets or rooms, separately from the research data. Research staff was trained in the Institutional Review Board-approved methods for managing and storing research data and specimens. Video recordings were stored on a password-protected drive and encrypted if transferred or transported. The two researchers and their research supervisor had access to video-recordings only for the purpose of this research study. Recordings will be destroyed once the information needed for research has been obtained. These data will be kept secure and destroyed by the primary researcher. Research materials will be kept for a minimum of five years after publication per the guidelines of the American Psychological Association. All procedures were approved by The Chicago School of Professional Psychology's Institutional Review Board.

Summary

This chapter reviewed the methods that were used for my dissertation study. The primary investigator implemented parent training procedures using BST to increase treatment integrity of parent implementation of their children's BIPs. A research assistant independently collected data for at least a third of sessions for IOA purposes. Parents implemented all behavioral procedures with their children. A multiple-baseline design was used to evaluate the effects of BST on

treatment integrity. Potential threats to the validity of this study were examined and strategies are in place to eliminate these threats. Additionally, ethical assurances were put in place and confidentiality was obtained from participants before the start of this study.

Chapter 4: Results

Data showing global treatment integrity, component treatment integrity, and child accuracy on skill acquisition programs for the three dyads who participated in this study are presented in Figures 1 through 3. Overall, participants improved in their accuracy (both parent and child behavior) across phases.

Figure 1 shows the results for global treatment integrity across the three parents. Global treatment integrity was calculated by dividing the number of components implemented as written by the total number of components observed per session and converting to a percent. All three dyads had higher treatment integrity during BST as compared to baseline and met mastery criteria. Performance was also maintained, with treatment-integrity scores above 80%, measured 1 and 2 weeks post-BST for all participants. Global treatment integrity for Parent 1 in baseline was stable, ranging from 26% to 31% (30% in Session 1, 26% in Session 2, 31% in Session 3, and 29% in Session 4; *M = 29*). Upon the introduction of BST in Session 4, Parent 1's level of treatment integrity increased over 30%. Although Parent 1's treatment integrity was somewhat variable throughout the intervention, his performance gradually improved (55%, 73%, 76%, 86%, 91%, 96%, 94%; *M* = 81.6). In maintenance, 1 week after meeting the mastery criteria, treatment integrity above 90% across three sessions, Parent 1's treatment integrity level was maintained (96%); however, his score decreased to 82% during the second week after BST (*M* = 89). Global treatment integrity for Parent 2 in baseline was stable, ranging from 24% to 28% (27%, 26%, 28%, 27%, 27%, 24%; *M* = 22.7). Upon the introduction of BST in Session 7, Parent 2's level of treatment integrity quickly increased to above 90% (92%). Subsequently, in the second and third sessions of BST, Parent 2 met 100% accuracy across both sessions (*M* = 97.3). Her behavior was maintained in weeks one and two following BST with treatment integrity in

each session at 98% (*M = 98)*. Global treatment integrity for parent three in baseline was consistent across all eight baseline sessions (20%; *M* = 20). Upon the introduction of BST in Session 9, Parent 3's treatment integrity jumped to 100% across all three sessions (*M* = 100). Perfect treatment integrity was maintained at 1 and 2 weeks post-BST (*M* = 100).

Figure 2 shows average individual-component scores for all three participants. Average scores per component, measured by percentage of opportunities implemented correctly, are shown by the colored bars. Average scores for global treatment integrity are shown by the horizontal lines. Component integrity was calculated by dividing the number of correctly implemented responses within a single component by the total number of opportunities to implement each component response within a session and converting to a percent. Component treatment integrity for parent one was variable throughout study phases. In baseline, Parent 1 had high rates of treatment integrity for attention (92.5%) and completion (100). He had a moderate rate of treatment integrity for token or edible (68.8%). All other areas were incorrectly implemented 100% of the time (0%). During BST, all areas increased (materials, 94.3; attention, 97.1; instruction, 61.7; praise, 85.5; token or edible delivery, 93.7; error correction, 21.7, completion, 100). Parent 1 demonstrated a weakness with error correction, on average only implementing correctly 21.7% of opportunities). During maintenance, Parent 1 improved demonstrating higher integrity scores for materials (100%), instruction (75%), token or edible delivery (100%), and error correction (50%). Scores decreased or remained stable for attention (95%), praise (77.8%), and completion (100%). Component treatment integrity for Parent 2 either improved or remained stable for all areas across intervention phases. In baseline Parent 2 had high rates of treatment integrity for attention (100%) and completion (100%), low rates of treatment integrity for token or edible delivery (21.4%), and materials, instruction, praise, and

error correction were implemented incorrectly 100% of opportunities (0% accuracy). During BST, Parent 1's scores either remained stable or increased (materials, 100%; attention, 100%; instruction, 100%; praise, 90%; token or edible delivery, 96.7%; completion, 100%). The student made no errors during BST; therefore, there were no opportunities to implement error correction. Scores were similar in maintenance (materials, 100%; attention, 100%; instruction, 100%; praise, 100%; token or edible delivery, 94.5%; completion, 100%). Error correction in maintenance was implemented correctly 83.4% of opportunities, reflecting the lowest score across both BST and maintenance for Parent 2. Component treatment integrity for Parent 3 improved for all areas to 100% accuracy, with the exception of attention, which was already at 100% accuracy in baseline. For example, in baseline, Parent 3's scores were as follows: materials, 0%; attention, 100%, instruction, 0%; praise, 0%; token or edible delivery, 0%; error correction, 0%; completion, 0%. In intervention and maintenance scores were the same and were as follows: materials, 100%; attention, 100%, instruction, 100%; praise, 100%; token or edible delivery, 100%; error correction, 100%; completion, 100%.

Figure 3 shows the results for child skill acquisition, specifically progress made for the specific behavior intervention chosen. Overall, participants improved across phases. However, across participants behavior was variable. For Participant 1, in baseline, percent accuracy ranged from 20–40% (20%, 20%, 30%, 40%; $M = 27.5$). In BST, percent accuracy ranged from 50–100% (60%, 80%, 90%, 100%, 50%, 60%, 80%; $M = 74.3$). In maintenance, percent accuracy was 90% both 1 week and 2 weeks after BST. For Participant 2, in baseline, percent accuracy ranged from 50-80% (60%, 50%, 60%, 80%, 80%, 50%; $M = 63.3$). In BST, percent accuracy was 100% in all three sessions ($M = 100$). In maintenance, percent accuracy was 90% 1 week post-BST and 80% 2 weeks post-BST. For Participant 3, in baseline, percent accuracy ranged

from 33-67% (33%, 33%, 44%, 56%, 67%, 67%, 67%, 67%; *M* = 54.3). In BST, percent accuracy was 100%, 100%, and 40%. In maintenance, percent accuracy was 80% in both 1 week and 2 weeks after BST.

Chapter 5: Discussion and Conclusions

The purpose of this study was to examine the effect of BST as a method for increasing treatment integrity of parents' implementation of behavioral interventions for their children. A secondary purpose was to assess any correlation between child skill acquisition and treatment integrity of parent implementation. All three dyads displayed higher treatment integrity during BST as compared to baseline and met mastery criteria quickly. Performance was also maintained, with treatment-integrity scores above 80%, measured 1 and 2 weeks post-BST for all participants. These data replicate that of previous research showing that BST is an effective method for increasing skills (Barnett et al., 2014; DiGennaro-Reed & Codding, 2013; Jenkins & DiGennaro-Reed 2016; Pantermuehl & Lechago, 2015). These results suggest that parents can be trained to implement skill-acquisition programs with their children despite having limited-to-no experience or formal training in behavior analysis. Additionally, this training can be completed in a short amount of time (i.e., weeks). The shortest amount of time needed to meet mastery criteria for the current participants was three sessions and the longest was seven. Sessions were conducted on average two to three times a week. Therefore, all parents met mastery criteria within a month. If sessions had been conducted more often, the parents may have met the mastery criteria much faster.

In addition to the timing of the training sessions, it is also important to note that it may take parents longer to reach high levels of treatment integrity if their children demonstrate a high rate of maladaptive behaviors. For example, Parent 1 took the longest to reach mastery criteria, and worked with his son who demonstrated the highest rate of problem behavior out of the three children in this study—specifically yelling and screaming, noncompliance, and getting out of his seat during instructional times (escape). The other two children had no maladaptive behaviors

during the course of this study. This may also account for some of the variability in child one's behavior progress.

Another variable to consider is the type of procedure being trained. It is interesting to note that parents consistently scored lowest in error correction and highest in gaining attention, even in baseline. This may suggest that error correction is more difficult than other program components for parents and individuals with limited skills in behavior analysis to implement, and therefore, require more training. Errors made in implementation of error correction often involved both errors of omission (i.e. not touching the correct card) and errors of commission (i.e. adding unspecified words such as stating, "you're wrong," "this is the correct card," etc.). Future studies may consider examining more targeted training to address frequent errors, rather than continuing BST for all components.

It may also be important to consider the child's behavior and performance when training a parent in the programs. In the current study, child performance was somewhat variable; however, all children demonstrated higher accuracy on programs as compared to baseline. Variability may be attributed to several factors, including difficulty of the words paired in sets, but could have affected how well the parents were able to continue implementing training with high integrity. It may be important to consider initially training parents using a maintenance program and then moving into more difficult programs such as skill acquisition.

A limitation of this study is that parents worked with children who had different levels of functioning and behavioral difficulties. For example, Child 1 had a diagnosis of ADHD and global delays and demonstrated yelling or screaming (on average 20 per week), noncompliance, and aggression (approximately two per week). Child 2 was typically developing and demonstrated no behavior problems. Child 3 had a diagnosis of autism and ADHD, and

demonstrated a low rate of yelling or screaming (on average three per week). Functioning levels and behavioral difficulties of the individuals being worked with greatly impact the amount of response effort required to implement intervention programs. Additionally, a parent or therapist may require a greater skill set and expertise in program implementation when working with a more challenging case. They should be trained in managing problem behaviors as they arise as well as teaching the skill acquisition program. Despite our best efforts in initial training, it may still be difficult for parents and staff to manage a child who engages in high rates of problem behavior, requiring more frequent training and maintenance. Even though parents in this study required different amounts of training sessions in order to meet mastery criteria, BST was effective in training parents to implement skill acquisition program with their children as well as maintaining desired effects. Therefore, BST is an effective option for training parents despite behavioral difficulties of their children, but it should not be constrained to a specified amount of time or sessions, as this may differ on variables outside of the training itself.

Another limitation is that a social-validity questionnaire was not included. Including a questionnaire assessing parent's acceptability and preference of the training programs may provide valuable information for development of future trainings. Additionally, parents were not provided an option regarding the type of training they would receive. It is possible that parents would have mastered the skill faster if it was taught using a procedure they chose. It would also be interesting so see if mastery criteria would be met faster with the addition of other strategies such as a lottery-based incentive system, self-monitoring, token economy, or having the implementation agent engage in data collection.

Also, generalization measures were not included. It would have been beneficial to provide the parents with a novel behavior plan to measure the generalization of their skills.

Based on the data presented, we are only able to make assertions based on one program per parent–child dyad.

However, regardless of the limitations, this research is still valuable because it includes a component integrity measure, skills taught to both parents and children were relevant to future skill development, parents had to meet a high score across three sessions for training to end, a maintenance phase, a simple skill acquisition procedure for parent implementation, and a reachable mastery criterion (i.e. mastery criteria not set at 100%) were used. Based on these important inclusions, this work adds to treatment integrity research for parent implementation. In conclusion, as BCBAs continue to work with clients in the home setting, it is imperative that parents are able to implement interventions with their children. Additionally, it is important that they are able to do so with high levels of integrity and without time-consuming training procedures. In this study BST was demonstrated to be quick and effective in doing so. Research should continue to be conducted in this area. Similar research should also be conducted evaluating the use of BST to increase treatment integrity for parent implementation of behavior reduction procedures.

References

Arkoosh, M. K., Derby, K. M., Wacker, D. P, Berg, W., McLaughlin, T. F., & Barretto, A. (2007). A descriptive evaluation of long-term treatment integrity. *Behavior Modification, 31*(6), 880-895. doi:10.1177/0145445507302254

Barnett, D., Hawkins, R., McCoy, D., Wahl, E., Shier, A., Denune, H., & Kimener, L. (2014). Methods used to document procedural fidelity in school-based intervention research. *Journal of Behavioral Education, 23*(1), 89-107. doi:10.1007/s10864-013-9188-y

Blampied, N. M., Kahan, E. (1992). Acceptability of alternative punishments: A community survey. *Behavior Modification*, *16*(3)*,* 400–413. doi:10.1177/01454455920163006

Carroll, R. A., Kodak, T., & Adolf, K. J. (2016). Effect of delayed reinforcement on skill acquisition during discrete-trial instruction: Implications for treatment-integrity errors in academic settings. *Journal of Applied Behavior Analysis, 49*(1), 176-181. doi:10.1002/jaba.268

Carroll, R. A., Kodak, T., & Fisher, W. W. (2013). An evaluation of programmed treatment-integrity errors during discrete-trial instruction. *Journal of Applied Behavior Analysis, 46*(2), 379-394. doi:10.1002/jaba.49

Codding, R. S., Livanis, A., Pace, G. M., & Vaca, L. (2008). Using performance feedback to improve treatment integrity of classwide behavior plans: An investigation of observer reactivity. *Journal of Applied Behavior Analysis, 41*(3), 417–422. doi:10.1901/jaba.2008.41-417

Cook, J. E., Subramaniam, S., Brunson, L. Y., Larson, N. A., Poe, S. G., & St. Peter, C. C. (2015). Global measures of treatment integrity may mask important errors in discrete-trial training. *Behavior Analysis in Practice, 8*(1), doi:10.1007/s40617-014-0039-7

Craig, A. R. (2010). Self-administered behavior modification to reduce nail biting: Incorporating simple technology to ensure treatment integrity. *Behavior Analysis in Practice, 3*(2), 38–41. doi:10.1007/BF03391763

DiGennaro, F. D., Martens, B. K., & Kleinmann, A. E. (2007). A comparison of performance feedback procedures on teachers' treatment implementation integrity and students' inappropriate behavior in special education classrooms. *Journal of Applied Behavior Analysis, 40*(3), 447–461. doi:10.1901/jaba.2007.40-447

DiGennaro-Reed, F. D. & Codding, R. S. (2013). Advancements in procedural fidelity assessment and intervention: Introduction to the special issue. *Journal of Behavioral Education, 23*, 1-18. doi:10.1007/s10864-013-9191-3

Digennaro-Reed, F. D., Codding, R., Catania, C. N. & Maguire, H. (2010). Effects of video modeling on treatment integrity of behavioral interventions. *Journal of Applied Behavior Analysis*, *43*(2), 291–295. doi:10.1901/jaba.2010.43-291

Elliott, S. N., Witt, J. C., Galvin, G. A., & Peterson, R. (1984). Acceptability of positive and reductive behavioral interventions: Factors that influence teachers' decisions. *Journal of School Psychology*, *22*(4)*,* 353–360. doi:10.1016/0022-4405(84)90022-0

Fazzio, D., Amal, L., & Martin, G. (2010). Discrete-trials teaching evaluation form (DTTEF) scoring manual. Unpublished manuscript. Retrieved from: http://www.dtteaching.com/.

Fiske, K. E. (2008). Treatment integrity of school-based behavior analytic interventions: A review of the research. *Behavior Analysis in Practice*, *1*(2), 19–25. doi:10.1007/BF03391724

Fryling, M. J., Wallace, M. D., & Yassine, J. N. (2012). Impact of treatment integrity on intervention effectiveness. *Journal of Applied Behavior Analysis*, *45*(2), 449–453. doi:10.1901/jaba.2012.45-449.

Gresham, F. M., Gansle, K. A., & Noell, G. H. (1993). Treatment integrity in applied behavior analysis with children. *Journal of Applied Behavior Analysis, 26*(2), 257–263. doi:10.1901/jaba.1993.26-257

Gross, T. J., Duhon, G. J., & Doerksen-Klopp, B. (2014). Enhancing treatment integrity maintenance through fading with indiscriminable contingencies. *Journal of Behavioral Education, 23*(1), 108–131. http://dx.doi.org/10.1007/s10864-013-9185-1

Grow, L. L., Carr, J. E., Gunby, K. V., Charania, S. M., Gonsalves, L., Ktaech, I. A., & Kisamore, A. N. (2009). Deviations from prescribed prompting procedures: Implications for treatment integrity. *Journal of Behavioral Education, 18,* 141–156. doi:10.1007/s10864-009-9085-6

Hagermoser-Sanetti, L. M. & Collier-Meek, M. A. (2014). Increasing the rigor of procedural fidelity assessment: An empirical comparison of direct observation and permanent product review methods. *Journal of Behavioral Education, 23*(1), 60–88. doi:10.1007/s10864-013-9179-z

Hagermoser-Sanetti, L. M. & Kratochwill, T. R., (2008). Treatment integrity in behavioral consultation: Measurement, promotion, and outcomes. *International Journal of Behavioral Consultation, 4*(1), 95-114. doi:10.1037/h0100835

Hagermoser-Sanetti, L. M., Dobey, L. M., & Gritter, K. L. (2012). Treatment integrity of interventions with children in the *Journal of Positive Behavior Interventions* from 1999 to 2009. *Journal of Positive Behavior Interventions, 14*(1), 29–46. doi:10.1177/1098300711405853

Howard, M. R., Burke, R. V., & Allen, K. D. (2013). An evaluation of the observer effect on treatment integrity in a day treatment center for children. *Behavior Modification, 37*(4), 490–515. doi:10.1177/0145445513486801

Jenkins, S. R., & DiGennaro-Reed, F. D. (2016). A parametric analysis of rehearsal opportunities on procedural integrity. *Journal of Organizational Behavior Management, 36*(4), 255–281. doi:10.1080/01608061.2016.1236057

Kalfus, G. R., & Burk, B. K. (1989). The influence of case history and treatment mediator on ratings of acceptability of child treatment. *Child and Family Behavior Therapy*, *11*(2), 45–55. doi:10.1300/J019v11n02_04

Kazdin, A. E., Bellack, A. S., & Hersen, M. (Eds.). (1980). *New perspectives in abnormal psychology*. New York, NY: Oxford University Press.

Kazdin, A. E. (1984). Acceptability of aversive procedures and medication as treatment alternatives for deviant child behavior. *Journal of Abnormal Child Psychology*, *12*(2), 289–301. doi:10.1007/BF00910669

Leon, Y., Wilder, D. A., Majdalany, L., Myers, K., & Saini, V. (2014). Errors of omission and commission during alternative reinforcement of compliance: The effects of varying levels of treatment integrity. *Journal of Behavioral Education, 23*(1), 19–33. doi:10.1007/s10864-013-9181-5

Martens, B. K., Witt, J. C., Elliott, S. N., & Darveaux, D. X. (1985). Teacher judgements concerning the acceptability of school-based interventions. *Professional Psychology: Research & Practice, 16*(2), 191-198. doi:10.1037/0735-7028.16.2.191

McIntyre, L. L., Gresham, F. M., DiGennaro, F. D., & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the *Journal of Applied Behavior Analysis* 1991-2005. *Journal of Applied Behavior Analysis, 40*(4), 659–672. doi:10.1901/jaba.2007.659-672

Miller, D. L., & Kelley, M. L. (1992). Treatment acceptability: The effects of parent gender, marital adjustment, and child behavior. *Child and Family Behavior Therapy*, *14,* 11–23. https://www.tandfonline.com/toc/wcfb20/current

Miller, M. V., Carlson, J., & Sigurdsson, S. O. (2014). Improving treatment integrity in a human service setting using lottery-based incentives, *Journal of Organizational Behavior Management, 34*(1), 29–38. doi:10.1080/01608061.2013.873381

Mouzakitis, A., Codding, R. S., & Tryon, G. (2015). The effects of self-monitoring and performance feedback on the treatment integrity of behavior intervention plan implementation and generalization. *Journal of Positive Behavior Interventions, 17*(4), 223–234. doi:10.1177/1098300715573629

Noell, G. H., Gresham, F. M., & Gansle, K. A. (2002). Does treatment integrity matter? A preliminary investigation of instructional implementation and mathematics performance. *Journal of Behavioral Education, 11*(1), 51–67. doi:10.1023/A:1014385321849

Pantermuehl, R. M., & Lechago, S. A. (2015). A comparison of feedback provided in vivo versus an online platform on the treatment integrity of staff working with children with autism. *Behavior Analysis in Practice, 8*(2)*,* 219–222. doi:10.1007/s40617-015-0059-y

Pelletier, K., McNamara, B., Braga-Kenyon, P., & Ahearn, W. H. (2010). Effect of video self-monitoring on procedural integrity. *Behavioral Interventions, 25*(4)*,* 261–274. doi:10.1002/bin.316

Pence, S. T., & St. Peter, C. C. (2015). Evaluation of treatment integrity errors on mand acquisition. *Journal of Applied Behavior Analysis, 48*(3), 575-589. doi:10.1002/jaba.238

Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15*(4), 477–492. doi:10.1901/jaba.1982.15-477

Plavnick, J. B., Ferreri, S. J., & Maupin, A. N. (2010). The effects of self-monitoring on the procedural integrity of a behavioral intervention for young children with developmental disabilities. *Journal of Applied Behavior Analysis, 43*(2), 315–320. doi:10.1901/jaba.2010.43-315

Plumb, J. C. & Vilardaga, R. (2010). Assessing treatment integrity in acceptance and commitment therapy: Strategies and suggestions. *International Journal of Behavioral Consultation and Therapy, 6*(3), 261–295. https://eric.ed.gov

Reed, D. D., Fienup, D. M., Luiselli, J. K., & Pace, G. M. (2010). Performance improvement in behavioral health care: Collateral effects of planned treatment integrity observations as an applied example of schedule-induced responding. *Behavior Modification, 34*(5), 367–385. doi:10.1177/0145445510383524

Spreat, S., Lipinski, D., Dickerson, R., Nass, R., & Dorsey, M. F. (1989). A paramorphic representation of the acceptability of behavioral programming. *Behavioral Residential Treatment*, *4*(1)*,* 1–13. doi:10.1002/bin.2360040102

Sterling-Turner, H. E. & Watson, T. S. (2002). An analog investigation of the relationship between treatment acceptability and treatment integrity. *Journal of Behavioral Education, 11*(1), 39–50. doi:10.1023/A:1014333305011

St. Peter, C. C., Byrd, J. D., Pence, S. T., & Foreman, A. P. (2016). Effects of treatment-integrity failures on a response-cost procedure. *Journal of Applied Behavior Analysis, 49*(2), 308–328. doi:10.1002/jaba.291

St. Peter Pipkin, C., Vollmer, T. R., & Sloman, K. N. (2010). Effects of treatment integrity failures during differential reinforcement of alternative behavior: A translational model. *Journal of Applied Behavior Analysis, 43*(1), 47–70. doi:10.1901/jaba.2010.43-47

Strohmeier, C., Mulé, C., & Luiselli, J. K. (2014). Social validity assessment of training methods to improve treatment integrity of special education service providers. *Behavior Analysis in Practice, 7*(1), 15-20. doi:10.1007/s40617-014-0004-5

Vollmer, T. R., Sloman, K. N., & St. Peter Pipkin, C. (2008). Practical implications of data reliability and treatment integrity monitoring. *Behavior Analysis in Practice, 1*(2), 4–11. doi:10.1007/BF03391722

Weck, F., Grikscheit, F., Höfling, V., & Stangier, U. (2014). Assessing treatment integrity in cognitive-behavioral therapy: Comparing sessions segments with entire sessions. *Behavior Therapy, 45*(4)*,* 541–552. doi:10.1016/j.beth.2014.03.003

Wilder, D. A., Atwell, J., & Wine, B. (2006). The effects of varying levels of treatment integrity on child compliance during treatment with a three-step prompting procedure. *Journal of Applied Behavior Analysis, 39*(3), 369–373. doi:10.1901/jaba.2006.144-05

Wilkinson, L. A. (2007). Assessing treatment integrity in behavioral consultation. *International Journal of Behavioral Consultation and Therapy, 3*(3), 420–432. https://eric.ed.gov

Appendix A: Figures

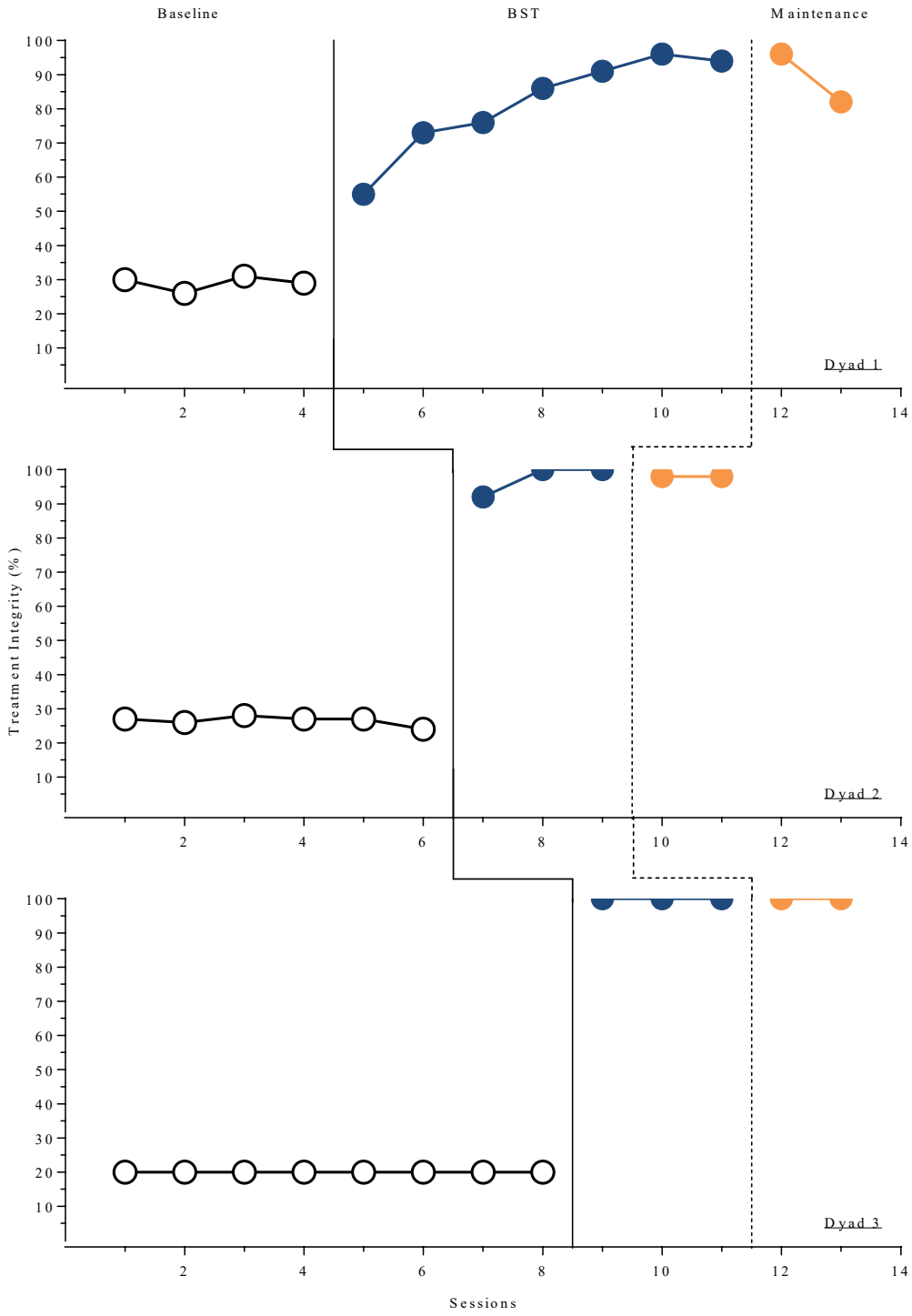

Figure 1. Global treatment-integrity scores for all parent participants.

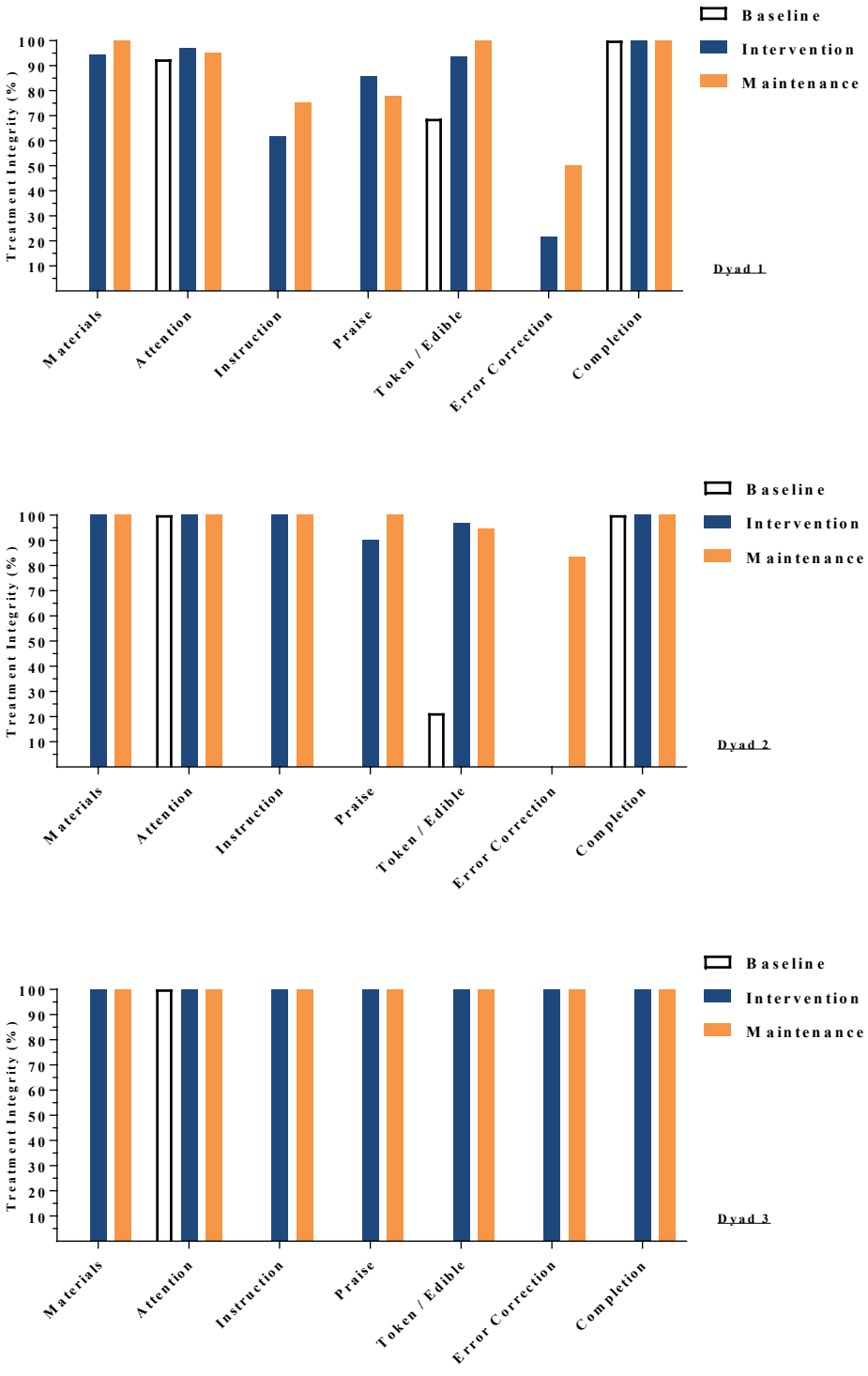

Figure 2. Component treatment-integrity scores for all parent participants.

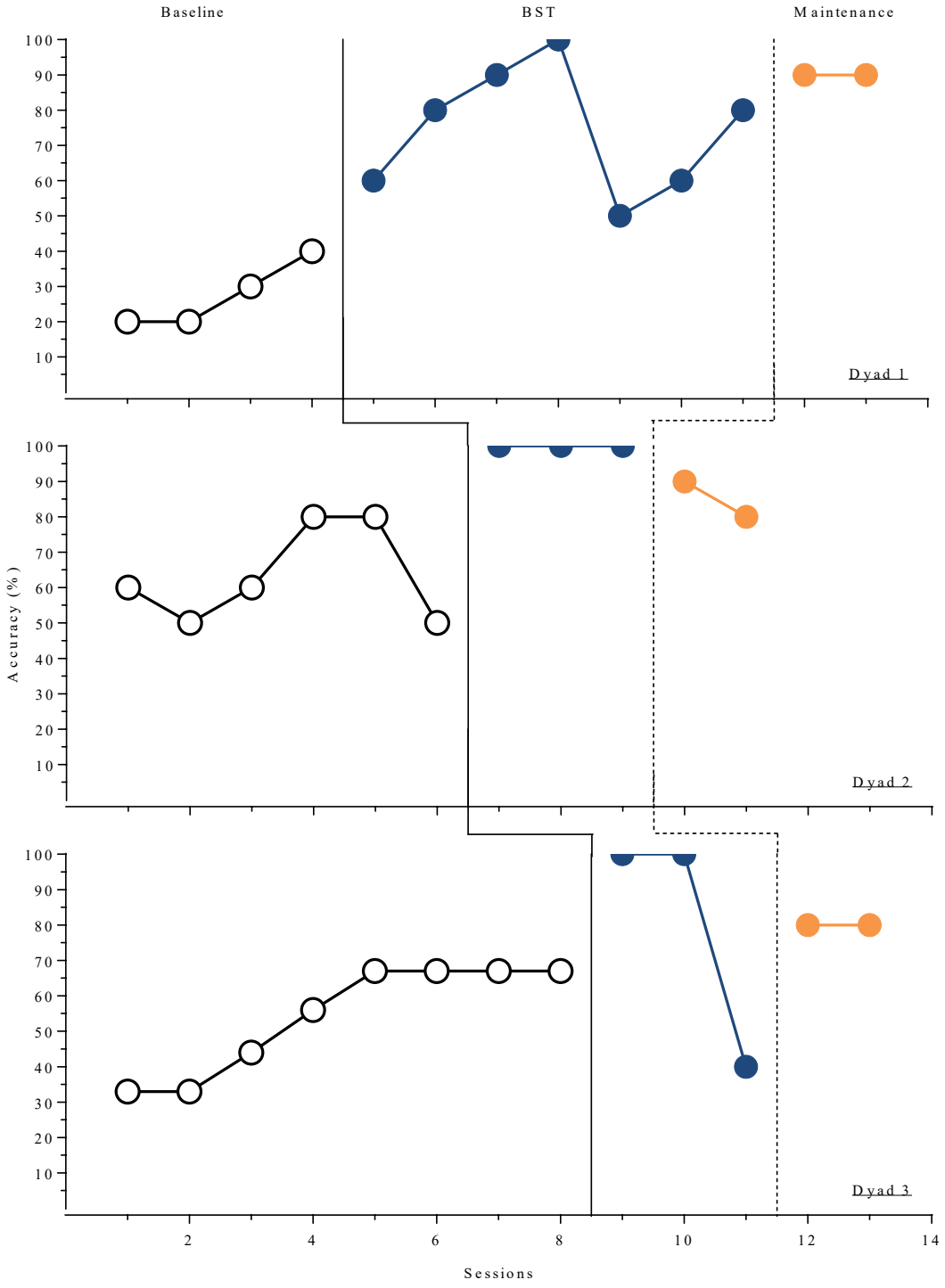

Figure 3. Child behavior progress for skill-acquisition programs for all children participants.