# AN EVALUATION OF COMMON TREATMENT INTEGRITY ERRORS BASED ON LEVEL OF EXPERIENCE AND THEIR RELATIONSHIP TO KNOWLEDGE OF APPLIED BEHAVIOR ANALYSIS

By

ERICA M. DASHOW

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Psychology

Written under the direction of

G. Terence Wilson

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October 2019

ABSTRACT OF THE DISSERTATION

AN EVALUATION OF COMMON TREATMENT INTEGRITY ERRORS BASED ON LEVEL OF EXPERIENCE AND THEIR RELATIONSHIP TO KNOWLEDGE OF APPLIED BEHAVIOR ANALYSIS

by ERICA M. DASHOW

Dissertation
Director:
G. Terence Wilson, Ph.D.

Treatment integrity (TI) refers to the extent to which the independent variable (e.g. treatment) has been implemented as intended and is concerned with both the accuracy and consistency of implementation. Without high TI, obtained results are ambiguous and not easily interpretable, and low integrity may result in poor outcomes for students. The existing literature suggests that teachers often fail to implement interventions with high integrity despite having received intensive training. However, it is unclear whether level of experience plays a role in the amount and types of TI errors made. The present study examined a) whether the types and number of TI errors differ depending on staff's level of experience, b) whether errors in specific components (e.g. reinforcement or prompting) are related to a lack of understanding of those components, and c) whether the number of errors made on a TI quiz are related to level of experience. Forty-two expert staff and 28 novice staff working at a university-based school specializing in applied behavior analysis for individuals with autism spectrum disorder participated in the study. Each participant was observed during a discrete trial instruction (DTI) session with a student and their TI on several components was coded. Participants then completed a brief quiz

containing procedural and theoretical questions related to each TI component as well as a survey about their level of experience. The results of the current study indicated that there was no significant difference in the number of TI errors made between groups; however, novice staff made a significantly greater number of errors on implementing controlling prompts than experienced staff. Individuals who made a greater number of TI errors also tended to make a greater number of errors on the quiz, but there were no significant correlations between individual quiz and TI components. Finally, results indicated that the novice group made significantly more errors on both procedural and theoretical quiz questions when compared to the experienced group. Implications for training staff are discussed.

# ACKNOWLEDGMENTS

Thank you to Bob LaRue for all of his encouragement, mentorship, and support over the last six years. I could always count on him for advice and assistance and could not have made it to this point without him. Many thanks to Terry Wilson for acting as the chair of my committee, and to Lara Delmolino-Gatley, who generously offered to step in halfway through this project. Thank you to Stacy Lauderdale for taking the time to assist me with the analysis portion of my dissertation. Special thanks to Sandy Harris, who supported me throughout my graduate education and helped guide me through the conceptualization of my graduate school research.

Thank you to thank Kate Fiske-Massey for her guidance, mentorship, and unwavering optimism that everything would work out. I feel so grateful to have had the opportunity to work with her. Special thanks to Meredith Bamond and Rob Isenhower for their friendship and advice throughout my time at the DDDC. I could not have asked for two better people to share an office with. Many thanks to Megan Dent, Nishi Kadakia, Ashley George, Ellie Mohanty, Areizo Ahmad, and Amalia Hubal for their assistance with data collection and coding.

Finally, thank you to my friends and family for their unwavering love and support throughout my time in graduate school.

TABLE OF CONTENTS

LIST OF TABLES

## I. Introduction

Applied Behavior Analysis (ABA) is a scientific approach that focuses on systematically manipulating environmental variables to determine their effects on socially significant behavior (Cooper, Heron, & Heward, 2007, p. 3). Inherent in this definition is one of the core features of ABA – the systematic manipulation of variables in order to produce significant changes in behavior. One of the most fundamental goals of behavior analysis is to demonstrate that changes in behavior are functionally related to particular stimuli in the environment (Baer, Wolf, & Risley, 1968). To do so, environmental variables are manipulated in a systematic way such that other extraneous variables are controlled for.

In order to demonstrate that a behavior has been changed as a result of a treatment, we must be able to show that the treatment itself has been implemented properly. Treatment integrity (TI) refers to the extent to which the independent variable (e.g. treatment) has been implemented as intended and is concerned with both the accuracy and consistency of implementation (Peterson, Homer, & Wonderlich, 1982). Behavior analysts most frequently utilize single case designs where the independent variable is systematically and repeatedly applied over a period of time. Therefore, measurement of the accuracy of the independent variable is a crucial component of the science of applied behavior analysis. Without high TI, obtained results are ambiguous and not easily interpretable. One cannot say an intervention has occasioned behavior change if the intervention was not implemented properly. A causal relationship can only be inferred if an intervention is implemented with high accuracy and the desired outcomes are produced (McIntyre, Gresham, DiGennaro, & Saini, 2007). It is important

to be able to distinguish between ineffective interventions and effective interventions that have been implemented with low integrity (Gresham, 1989).

Treatment integrity can be broken down into three components: therapist competence, treatment adherence, and treatment differentiation. Therapist competence refers to the level of judgment and skill shown by the therapist in delivering the intervention and may include factors such as inflexibility in the use of techniques or difficulty communicating effectively with clients. Treatment adherence refers to how accurately and consistently the therapist utilizes the specific components and procedures of the intervention and includes how well a therapist follows a protocol for implementing tasks. Finally, treatment differentiation refers to whether the treatments being studied differ enough along crucial components (i.e. actively implementing procedures for one type of treatment and avoiding implementing procedures for another type of treatment) (Perepletchikova & Kazdin, 2005). Low integrity in any of these components can affect the success of the treatment as well as the validity of the study.

**Measurement and Calculation of Treatment Integrity**

Measurement of the independent variable can be achieved through several methods, such as direct observation, self-report, permanent products, and behavioral interviews (Wilkinson, 2007). TI can be measured through direct observation, in which a rater records the occurrence or non-occurrence of specific treatment components. While this method is the most accurate way of measuring integrity, it is time consuming and often not feasible (Gresham, 1989).

Self-report is a method where an implementer reports the extent to which they have implemented each component of the treatment as intended, and does not require

significant resources or time to complete. However, self-report is often inaccurate and is prone to overestimation of treatment fidelity (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000).

Another method of measuring TI is the examination of permanent products. Permanent products refer to naturally occurring products of the treatment and can be used as indicators of accurate implementation (e.g. examining whether or not school notes were sent home and signed by parents as a measure of parent/teacher correspondence). This method is advantageous in that it is less time consuming, less reactive, and more accurate than assessing TI via self-report (Gresham et al., 2000). Unfortunately, many procedures do not result in permanent products.

Treatment integrity scores can be calculated in several ways. One commonly used measure of TI is the percent of opportunities in which the procedure was correctly implemented (Vollmer, Sloman, & Pipkin, 2008). This percentage is calculated by dividing the sum of accurately implemented components by the total number of opportunities to implement the components, and then multiplying this number by 100 in order to yield a percentage. However, this measure does not allow for differentiation of integrity between specific components. Another method used is to calculate TI for each of the individual treatment components. This percentage is calculated by dividing the sum of a single accurately implemented component (e.g. reinforcement) by the total number of opportunities to implement the component, and then multiplying this number by 100 in order to yield a percentage. This measure allows for more targeted feedback and improvement in individual TI components (Vollmer et al., 2008).

**The Status of Treatment Integrity in Autism Research**

While many methods exist for measuring and calculating TI, there is relatively little literature focusing on this topic. Although it is standard practice to record data reliability (e.g. through interobserver agreement), it is not standard practice to record TI. Progar, Perrin, DiNovi, and Bruce (2001) rated a selection of articles from 1968-1972 and 1996-2000 from the Journal of Applied Behavior Analysis (JABA) for whether TI was measured. They found that, although the number of studies reporting a measure of the independent variable did increase, the total number of studies also increased. Overall, they found that the percentage of articles reporting TI did not increase over the course of three decades.

Neely, Davis, Davis, and Rispoli (2015) conducted a more recent review of autism focused journals analyzing the extent to which TI data had been reported over a time period of 30 years, by sampling articles from specific years over three decades. One hundred and nineteen articles were included in the review. Neely and colleagues found that TI data were collected in 17% of studies in 1992, 38% of studies in 2002, and 59% of studies in 2012. However, after evaluating the number of studies that met a quality threshold (integrity above 80%, as is the standard when collecting data reliability), the percentage of studies that met the standard decreased to 6% in 1992, 19% in 2002, and 30% in 2012. Although these results are somewhat promising in that they show an increasing trend in TI reporting, these numbers reflect a need for improvement in the reporting of high-quality TI data. Fifty-nine percent of studies with reported TI means that the results of 41% of studies just a few years ago cannot be interpreted without caution. Additionally, it is important to consider that the number of publications on autism using single case design increased 295% from 1992 to 2012, indicating that the

field still requires improvement in this area, as TI is a variable that is still being overlooked.

**Treatment Integrity in Applied Behavior Analysis**

The assessment of TI is particularly important in ABA because the training of other intervention agents (such as consultation to schools) is a large area of service delivery in the field. In this role, the behavior analyst typically develops an intervention and subsequently trains staff to implement the treatment. The staff are then responsible for implementing the intervention as intended and keeping the consultant informed in regard to student outcomes (DiGennaro, Martens, & Kleinmann, 2007). Consultants frequently assume that staff will implement plans as intended based on verbal agreement and good intentions. However, the behavioral consultation process does not provide a built-in method of monitoring integrity (Wilkinson, 2007). In fact, only 11% of school professionals report consistent evaluation of TI during the consultation process (Cochrane & Laux, 2008). This limited focus on assessing TI may be due to time constraints or lack of resources, especially since priority is typically given to integrity of data collection of outcomes, or dependent variables (Mouzakitis, Codding, & Tryon, 2015).

The existing literature suggests that teachers often fail to implement interventions with high integrity despite having received intensive training (DiGennaro, Martens, & McIntyre, 2005). Initial TI following training tends to be high but decreases over time as supervision and support decrease (Codding, Feinberg, Dunn, & Pace, 2005). This is unsurprising, as interventions may require teachers to learn new behavior management and instructional skills, as well as to learn new ways of interacting with students, and to incorporate and use these skills during their work day (DiGennaro et al., 2007). Research

suggests that in order to maintain high integrity, staff require ongoing supervision, support, and clinical consultation over an extended period of time (Addis, 2002).

It is a waste of time and resources for both the consultant and the teacher if treatments are not implemented properly. If a treatment is not effective in changing behavior, additional time may be wasted in modifying the treatment, or even designing, training, and implementing new treatment plans that would have been unnecessary had the original intervention been implemented as intended. This is particularly important to consider when providing services to young children, as research has indicated that early intervention is critical (Corsello, 2005). If a behavior plan is deemed ineffective, consultants may move to more restrictive interventions (Vollmer et al., 2008). This is an issue that has important ethical implications because consultants may move to punishment-based procedures or medication when such procedures may not actually be warranted. Treatment integrity is of paramount importance while conducting assessments such as functional behavioral assessments, as all treatments put into place should be based on the function of the behavior. Poorly conducted assessments can lead to misidentification of behavioral function (false positives or negatives), and therefore, lead to ineffective treatment.

Low integrity may also result in poor outcomes for students. If a teacher does not implement a behavior plan as written, it is unlikely that the intervention will be successful in prompting behavior change (Sanetti, Luiselli, & Handler, 2007). High integrity has been implicated as a vital component of successful interventions, and higher integrity usually produces better outcomes (DiGennaro et al., 2007; DiGennaro et al., 2005; Gresham, 1989).

**Treatment and Staff Variables**

There are several treatment and staff related factors that impact how well TI is maintained. Treatment-related factors include intervention complexity, access to necessary materials, and time constraints. Yeaton and Sechrest (1981) recognized that the complexity of the treatment directly impacts the degree of integrity, such that more complex treatments result in lower TI. Interventions that require resources or materials beyond what is typically available are also more likely to have lower integrity than interventions that do not require specific or special resources (Gresham, 1989). The time required to implement an intervention also plays a large role in how accurate implementation will be. Time also interacts with complexity, in that more complex treatments often require more time to implement accurately (Gresham, 1989).

Staff-related factors that influence integrity include the number of treatment agents, the motivation of the treatment agent, and perception of the intervention. Treatments that are more complex or time consuming may require more than one individual to implement the program. The more individuals involved in implementing a program, the more possible points of integrity break down exist. This is particularly relevant considering therapist drift, in which treatment agents gradually alter the treatment over time. This can be seen in either staff not implementing particular components of the treatment, or in adding additional components that are not included in the intervention plan (Peterson et al., 1982).

The motivation of the teacher may also play a role. If a teacher's goal is for the child in question to be referred out or to be placed elsewhere (such as transitioning from a regular classroom into a special education classroom), then TI may suffer because the

teacher does not want to treatment to be effective. (Gresham, 1989). In contrast, if the purpose of the intervention is to solve a problem for the implementer (e.g. targeting the disruptive behavior of a student in the classroom), then it is more likely that the individual will be motivated to implement the treatment with high integrity (Detrich, 2014).

Finally, a teacher's perception of an intervention may have an impact on their TI. If a treatment is perceived to be effective or to result in rapid behavior change, treatment may be continued with higher integrity than a treatment that staff perceives to be slow acting or ineffective. Further, interventions are more socially valid and match the beliefs, experiences, or values of the individuals implementing them are more likely to be perceived as reasonable and credible and are therefore more likely to be implemented with higher integrity (Detrich, 1999).

**Strategies to Improve Treatment Integrity**

In addition to examining factors relating to poor TI, researchers have also focused on methods that can be used to improve TI. One major factor related to TI is type of training procedure. Indirect training involves procedures in which individuals are trained via written materials (e.g. manuals) or verbal instruction (e.g. lectures). Although these procedures are commonly used, they are less effective than other methods, such as direct training (Sterling-Turner, Watson, Watkins, & Little, 2001).

Direct training involves procedures such behavioral skills training, which includes modeling, role playing, rehearsal, and feedback. Compared to indirect training procedures, direct training procedures provide greater opportunities to practice skills, which increases the probability of use and generalization (Sterling-Turner et al., 2001).

Behavioral skills training is an empirically supported training method, and research has supported its use in the maintenance of skills over time (Parsons & Reid, 1995). Sterling-Turner and colleagues (2001) compared the effects of different training methods on TI for a complex behavior protocol. Participants included 64 undergraduate students who were randomly assigned to one of three conditions: didactic training, modeling training, or rehearsal/feedback training. During didactic training, participants received verbal training (an explanation of the procedures) and three examples of the target behavior were demonstrated. In the modeling training, participants watched a 5-minute video of accurate implementation of the procedure, during which the trainer provided verbal explanations for each component of the protocol. During rehearsal/feedback training, participants received a 5-minute training where verbal prompts were delivered for each component of the protocol, and participants received praise for accurate implementation and corrective feedback for mistake. The results indicated that participants with rehearsal/feedback training had the highest TI scores, whereas the participant with didactic training had the lowest TI scores. Overall, this study showed that more direct training leads to more accurate implementation. These results are particularly important to keep in mind when training new staff with little to no experience in the field.

Other researchers have examined TI within an actual classroom setting. Noell et al. (1997) examined the effect of performance feedback on three teachers' TI in an elementary school. The first phase consisted of consultation only, in which the consultant explained how to implement the intervention and provided data collection forms. Contact between the teacher and consultant was then discontinued, mimicking the typical consultation process. In the performance feedback phase, the consultant met with the

teacher each morning and presented graphs depicting student performance data and teacher integrity data. Additionally, the consultant provided specific feedback regarding intervention components with low integrity, praise for components implemented accurately, and discussed how performance could be improved that day. During the maintenance phase, data collection was continued but performance feedback was discontinued. Noell and colleagues found that, although all teachers initially exhibited high TI, performance decreased after two to four days. When the performance feedback was introduced, all teachers demonstrated a subsequent increase in TI. These results indicate that performance feedback can be an effective strategy to increase TI during the consultation process.

In an extension of previous research, Codding and colleagues (2005) examined the effect of performance feedback on five teachers' implementation of antecedent versus consequence-based procedures. This is particularly important as the accuracy of teachers' implementation may differ depending on the type of procedure, as each requires a different skill set. During baseline, TI data sheets were completed based on direct observation and without the teacher's knowledge, after which feedback was provided. Treatment consisted of performance feedback, where the consultant provided an average of 12 minutes of feedback on the day of the observation, and was provided every other week. Performance feedback was removed once improved performance was stable. Finally, maintenance and follow up sessions were conducted five weeks after the last feedback session, and at further five-week intervals. The effects of this study indicated that performance feedback was effective in increasing TI across both antecedent and consequence-based behavior plans. The results of the intervention were maintained

across teachers for over three months. The findings of Codding and colleagues support and extend previous research findings indicating that performance feedback can increase TI for both antecedent and consequence-based interventions, even when feedback is given on a less frequent basis.

DiGennaro-Reed, Codding, Catania, and Maguire (2010) examined the effects of video modeling and performance feedback on TI of three teachers of individuals with developmental disabilities in the classroom. After a brief training period, individualized video modeling was introduced. Participants viewed an instructional video of a role play in which an experienced teacher modeled accurate implementation of the intervention, while a voice-over explained relevant details of the procedure. Teachers were instructed to implement the procedure within 45 minutes, during which TI data were collected. In the video modeling plus feedback phase, the consultant paused the video at points where the teacher had not implemented a step correctly and prompted them to pay attention. A follow up probe was conducted 1 week prior to termination of the intervention. DiGennaro-Reed and colleagues found that accuracy increased immediately following introduction of video modeling, although performance was highly variable. When performance feedback was introduced, integrity rose to 100% across all participants, and these gains were maintained at the one week follow up probe. These findings indicate that while video modeling alone can increase integrity, it is not sufficient to promote stable performance. The addition of performance feedback was crucial to maintain treatment gains. It is important to note, however, that the follow up period was brief (one week) and that it remains unclear whether accurate implementation would maintain over longer periods of time.

In the previous studies, performance feedback was given in the forms of graphs (Noell et al., 1997) and verbal feedback (Codding et al., 2005). Sanetti et al. (2007) compared the effects of verbal and graphic performance feedback on a four-person teaching team's intervention implementation in an elementary school. In the verbal feedback condition, the trainer met with the teachers and verbally reported the percentage of components implemented correctly. The trainer also provided corrective feedback for components implemented incorrectly and answered questions. In the verbal and graphic feedback condition, the trainer presented a graph of the teachers' percentage of components implemented correctly in addition to providing verbal feedback as described above. The results indicated that while TI did not improve when verbal performance feedback was given, a combination of both procedures resulted in over a 100% increase in TI. Furthermore, improved TI led to better student performance, such that the student displayed increased levels of appropriate behavior during the verbal and graphic feedback condition.

Although performance feedback has been shown to be effective at increasing accurate treatment implementation, it has several disadvantages. Performance feedback is time and resource intensive for both teachers and consultants and can cause discomfort to those receiving such feedback. Additionally, there is literature indicating that TI does not remain high once performance feedback is removed, indicating a lack of sustainability with this method. Therefore, it is important to continue to research additional methods to improve TI (Gresham, 1989; Noell et al., 2005).

Self-monitoring is another strategy that has been suggested to increase TI, and, unlike performance feedback, is a more sustainable model over time. Self-monitoring

differs from self-report in that self-monitoring involves a person systematically observing and recording their own behavior, as opposed to simply attempting to recall their own behavior after the fact. This method would allow teachers to become more independent in managing their own performance, therefore freeing up time and resources of supervisors or consultants. Mouzakitis and colleagues (2015) examined the use of performance feedback and self-monitoring strategies in schools to promote increased implementation accuracy among four special education teachers in an inclusion program for children with ASD. Specifically, they examined the use of self-monitoring alone vs. self-monitoring plus performance feedback, and how these interventions would impact TI for the delivery of treatment with a target student and generalization student, as well as percentage of on task behavior for the target and generalization students. As in previous studies, the data showed that TI decreased within 3 days of initial training. Although TI improved with self-monitoring, it did not improve to criterion levels until performance feedback was added. When performance feedback was removed, two of three teachers maintained high levels of TI. Improvements in TI generalized to the implementation of other students' behavior plans for three of the four teachers, supporting the hypothesis that performance feedback and self-monitoring may generalize to other students in the classroom with similar intervention plans. Mouzakitis and colleagues also found a moderate to high association between TI and student behavior outcomes. Overall, these results suggest that although self-monitoring does increase TI, performance feedback may be a necessary component in order for integrity to reach acceptable levels.

Other strategies to augment the benefits of performance feedback have been studied as well. DiGennaro et al. (2005) examined the effects of a negative reinforcement

contingency relating to directed rehearsal coupled with performance feedback. Four elementary school teacher-student dyads participated in the study. The negative reinforcement contingency worked as follows: teachers could avoid meeting with the consultant to practice program components (directed rehearsal) that were implemented incorrectly by maintaining 100% TI. Following training and a baseline phase in which teachers were not given feedback or assistance, performance feedback and the negative reinforcement contingency were introduced. Performance feedback consisted of daily written feedback as well as graphs of their TI as well as their student's off task behavior. If a teacher did not maintain 100% integrity, the consultant scheduled a meeting for the next day during which they reviewed and practice missed steps several times. As stated previously, teachers were able to avoid these meetings by maintaining high integrity. Finally, dynamic fading was placed into effect, such that intervention procedures were thinned over time from daily to once every 2 weeks. The results of this study show that while implementation was variable during the baseline phase (eventually decreasing to 0-18%), all but one teacher met performance criteria (100% accuracy) within approximately three sessions. When the intervention was faded, all teachers averaged 91-100% TI. These results indicate that performance feedback and a negative reinforcement contingency can rapidly increase implementation accuracy, and, similar to Noell (2000), that a dense schedule of meetings with a consultant may not be necessary to maintain these gains. In addition, for some students, increased TI was associated with a decrease in student off task behavior.

DiGennaro et al. conducted another study in 2007, where they compared the extent to which two different sets of strategies affected teacher TI. Four student-teacher

dyads consisting of special education teachers and students with brain injuries participated in the study. The first strategy involved student performance feedback and goal setting, whereas the second strategy was comprised of teacher performance feedback and a negative reinforcement directed rehearsal contingency. After initial training (didactic instruction, modeling, coaching, and corrective feedback) and an implementation baseline, goal setting and student performance feedback were introduced. This particular strategy is unique in that goals were set for student behavior and teachers were unaware that TI data were being taken. Teachers received daily written feedback and graphs of student performance. The second intervention phase involved teacher performance feedback and directed rehearsal with meeting cancellation. In this phase, teachers were provided with daily written feedback and graphs of their own performance and were able to avoid meeting with a consultant to practice missed steps if they maintained 100% TI. In the fading phase, teacher performance feedback and directed rehearsal with meeting cancellation was faded over time to once every 2 weeks as long as integrity was maintained at 100%. The results of this study show that positive reinforcement contingencies such as goal setting and student performance feedback were not sufficient to maintain high TI. Teacher performance feedback and directed rehearsal meeting cancellation was a more effective strategy for increasing teacher integrity. The results also support previous findings that suggest that performance feedback can be faded to as little as once every two weeks without teachers demonstrating a decrease in treatment implementation accuracy (DiGennaro et al., 2005). Student behavior data are also consistent with other research findings, indicating a negative correlation between teacher TI and student problem behavior for 3 of 4 student/teacher dyads (DiGennaro et

al., 2005, Mouzakitis et al., 2015, Sanetti et al., 2007).

**The Effects of Varying Levels of Treatment Integrity on Behavioral Interventions**

Although it is clear that TI affects student outcomes, it is also important to examine the levels in which TI errors actually impact student performance. A review conducted by Gresham, Gansle, Noell, and Cohen (1993) showed a correlation between integrity and outcomes, such that higher integrity was associated with more positive treatment outcomes.

Vollmer, Roane, Ringdahl, and Marcus (1999) evaluated the effects of different levels of TI on students' appropriate and inappropriate behavior. Three students who had been referred by teachers and parents for treatment of severe behavioral problems participated in the study. Vollmer and colleagues sought to develop a methodology to test how successful a particular treatment would be in the face of treatment errors (e.g. less than optimal integrity). This is particularly important, as such a methodology would allow clinicians to identify critical treatment components. The treatment used as part of the integrity protocol was differential reinforcement of an alternative behavior (DRA), in which the inappropriate behavior contacted extinction and an alternative behavior was reinforced. Data were collected on aggression, self-injurious behavior, compliance, and appropriate mands. During baseline sessions, the condition of the functional analysis that produced the highest rate of problem behavior (escape condition for two participants and the tangible condition for one participant). During baseline, appropriate behavior was never reinforced and inappropriate behavior was reinforced 100% of the time. During the full implementation phase, appropriate behavior was reinforced 100% of the time and inappropriate behavior contacted extinction. In the partial implementation phases,

integrity was intentionally lowered to mimic typically occurring TI failures. For two participants, different levels of partial implementation included 25/75, 50/50, and 75/25, where the first value indicates percentage of reinforcement for appropriate behavior, and the second value the percentage of reinforcement for inappropriate behavior. For one participant, these levels included 20/0, 100/100, and 40/0, in order to examine effects of integrity errors when one treatment component (reinforcement or extinction) was always conducted with perfect integrity.

For the first participant, in the 25/75 condition, appropriate behavior decreased over time and inappropriate behavior increased over time. In the 50/50 condition, relatively stable rates of appropriate behavior were seen (comparable to full implementation), but inappropriate behavior occurred at higher rates than during full implementation. In the 75/25 condition, the participant demonstrated relatively stable rates of appropriate behavior that were comparable to full implementation, but inappropriate behavior occurred at higher rates than full implementation and lower rates than in the 50/50 condition. For the second participant, the 50/50 condition evoked relatively stable but low rates of appropriate behavior along with low levels of inappropriate behavior, whereas in the 25/75 condition appropriate behavior decreased over time and inappropriate behavior increased over time. For the last participant, variable rates of appropriate behavior and inappropriate behavior were seen in the 20/0 condition, with higher rates of appropriate behavior overall, but with a decreasing trend. In the 100/100 condition, appropriate behavior was high and no inappropriate behavior was exhibited. In the 40/0 condition, the participant exhibited a variable but increasing trend of appropriate behavior, as well as a variable but overall lower rate of inappropriate

behavior.

The overall results of the study indicate the following: at full implementation of the intervention, appropriate behavior almost fully replaced inappropriate behavior. At lower levels of implementation where the schedule of reinforcement favored inappropriate behavior, efficacy of the intervention was reduced. Unexpectedly, there was a general bias towards appropriate behavior during the 100/100 condition. This may have been due to response effort, such that less time or effort was necessary to engage in the appropriate response. Although this may not generalize to all individuals, it would be beneficial to keep response effort or time to a minimum when developing an appropriate alternative response. It is also interesting that, in conditions where integrity was low, participants initially showed a bias towards appropriate responding for several trials before finally switching over to the inappropriate response. These data suggest that if these earlier attempts to access reinforcement through appropriate behaviors had been reinforced with high integrity, the appropriate form of responding would persist.

Northup, Fisher, Kahang, Harrell, and Kurtz (1997) also examined the effects of different levels of TI on a treatment for problem behavior (i.e., DRA and time out). They implemented two components of the treatment, delivery of reinforcement and time out, at 100%, 50%, and 25% TI for three participants with developmental disabilities. When time out was reduced to 50% strength (time out implemented for every other instance of problem behavior rather than every time), the treatment was just as effective for all participants as compared to full 100% integrity. For two participants, treatment effects were maintained even when both components, treatment and reinforcement, were implemented only 50% of the time. Additionally, levels of problem behavior rose only

slightly during the 25% integrity phase of both reinforcement and time out. For the third participant, when time out only was implemented at 25% integrity (and reinforcement was maintained at 100% integrity), the treatment was just as effective as when all components were implemented with 100% accuracy. These results indicate that DRA with time out may be effective to a certain point even when implemented at with low integrity, and that the time out component in particular does not need to be implemented on a continuous schedule in order to be effective. However, it is also important to note that the individuals of this study responded differently to the varying levels of treatment strength, indicating that these lower levels may not maintain acceptable treatment levels for all individuals, and that these levels should still be individualized. Furthermore, different interventions may require higher levels of TI in order to maintain effectiveness. For example, Pence and Peter (2015) found that when mand-training was implemented with 40% TI and below, no participants were able to acquire the trained mands.

Leon, Wilder, Majdalany, Myers, and Saini (2014) also examined the effects of differing levels of TI on DRA effectiveness, but they examined both errors of omission (failing to deliver a reinforcer) and errors of commission (delivering a reinforcer erroneously). Two children participated in the first experiment, which examined errors of commission in the treatment of noncompliance when asked to give up a toy. Four conditions were evaluated using a reversal design – baseline, 100% integrity, 60% integrity, and 20% integrity. In the baseline condition, compliance was not reinforced, whereas in the 100% integrity condition, compliance always resulted in praise and the delivery of a preferred edible item. In the 60% and 20% conditions, compliance resulted in the delivery of an edible item and praise for only 60% and 20% of trials, respectively.

For both participants, the level of integrity was positively associated with level of compliance, such that compliance was lowest in baseline and the 20% condition, and highest in the 100% integrity condition. Compliance in the 60% condition was high, but levels depended on the condition that preceded it. For both participants, compliance during the 60% integrity condition was low and variable when it followed a 100% condition, and higher when it followed a baseline condition. This is an important finding to keep in mind when delivering and fading treatment, in that lower integrity may be more effective when following a baseline condition. These results support those of Northup et al. (1997), in that it may not be necessary to implement a DRA with 100% integrity in order to cause clinically significant behavior change. Additionally, these conditions may be more representative of the natural environment, where, depending on who the child is working with each day (e.g. different staff or home with parents), the treatment may be implemented at various and lower levels of TI than initially recommended.

The second experiment consisted of an examination of errors of commission on a DRA treatment for compliance and included three children as participants. A reversal design was used to evaluate three conditions – baseline, 0% integrity, 50% integrity (for one participant only), and 100% integrity. In the baseline phase, compliance resulted in praise. In the 0% integrity condition, praise and an edible item were delivered on every trial, regardless of compliance. In the 50% integrity phase, praise and an edible were delivered during 50% of trials where the child did not comply. In the 100% integrity phase, praise and an edible item were delivered only as programmed; that is, only after compliance. For all three participants, compliance was low in the 0% and 50% integrity

phases and high during the 100% integrity phase. Overall, these results indicate that level of integrity may matter more in regard to errors of commission than errors of omission, and that integrity of errors of omission may be dependent on the level of integrity immediately preceding them.

While the previous studies focused on the effects of various levels of TI in the reduction of problem behavior, Carroll, Kodak, and Fisher (2013) examined the effects of TI errors in the presentation of instructions, prompts, and reinforcement on skill acquisition in three experiments. In the first experiment, the researchers observed the children and teachers in a typical academic setting to identify the most common TI errors. Participants included five children and ASD and nine teachers or paraprofessionals, consisting of four special education teachers, one regular education teacher, three paraprofessionals, and one speech pathologist. Carroll and colleagues found that the most common errors included: failure to deliver a tangible item or praise after a correct response, not delivering a controlling prompt, and presenting the instruction more than once. However, the authors also found that teachers were more likely to secure attention, present clear instructions, and establish ready behavior, indicating that TI errors are not consistent.

In their second experiment, Carroll and colleagues (2013) evaluated the effects of these three most common errors when conducting discrete trial instruction (DTI). Six children with ASD participated in this study. Conditions in this alternating design study included baseline, control, high integrity, and low integrity. In the baseline condition, differential consequences were not provided for incorrect or correct responses. The control condition was similar to baseline and was used to monitor correct responding

without training. In the high integrity condition, procedures were implemented with 100% accuracy, whereas in the low integrity condition, integrity errors were programmed into 67% of trials. All participants in this study acquired and mastered the skill when the high integrity procedures were used, whereas only one participant mastered the stimuli when the low integrity condition was in place, and in double the time. These results indicate that integrity errors in the delivery of reinforcement, controlling prompts, and instructions can decrease treatment efficacy.

The third experiment was designed to parse out the effects of each of the three types of integrity errors. In this experiment, the effects of each type of integrity error (instructions, controlling prompt, reinforcement) were evaluated separately. Three children with ASD participated in this study. An alternating treatments design was used to compare acquisition of target stimuli during several low TI conditions. The baseline, control, and high integrity conditions were identical to those used in experiment two. Three low integrity conditions were included, where integrity errors were programmed into 67% of trials. The three conditions included low integrity instructions, low integrity controlling prompts, and low integrity reinforcement. For each of the three conditions, all other DTI components except for the target (e.g. reinforcement) were implemented with high integrity. For the first participant, the skill acquisition criterion was reached most quickly in the low integrity reinforcement condition, followed by the low integrity instruction conditions and finally the low integrity prompt conditions. The second participant acquired the skill most quickly in the low integrity instructions condition, followed by the low integrity prompt and low integrity reinforcement conditions. For the third participant, the least amount of training time was required for the low integrity

reinforcement condition, and he did not master the target stimuli in the low integrity instructions and prompt conditions until integrity was increased. For two participants, skill acquisition of target stimuli occurred most quickly in the high integrity condition. Overall, responding was idiosyncratic, in that different integrity errors affected skill acquisition in each participant differently. For all participants, integrity errors detrimentally affected effectiveness or efficiency of skill acquisition.

**Gap in the Literature**

It is clear from the literature that TI is a vital component in applied behavior analysis. Although there have been many studies examining various aspects of this topic, significant gaps still remain. One area that has not yet been examined is whether level of experience plays a role in the type of errors made. Although we know that TI errors are common (whether due to a lack of training or therapist drift once training has ended), it is unclear whether the types of errors made differ based on level of experience. This information has implications for the focus of initial trainings, supplemental booster training sessions, as well as how frequently TI should be assessed for newer and more experienced staff.

**Current Study**

This study is designed to extend the literature on TI by answering several questions. First, do the types and number of TI errors in DTI differ depending on staff's level of experience? Second, does knowledge or understanding of rationale for each component of DTI correspond to errors in each area? That is, are errors in specific DTI components (e.g. reinforcement or prompting) related to a lack of understanding of that component? The final goal of this investigation was to examine whether the number of

errors made on a TI quiz are related to level of experience.

## II. Method

### Participants and setting

Seventy individuals working at the Douglass Developmental Disabilities Center (DDDC) participated in this study. The DDDC is a . Participants were recruited into two groups. The first group contained 28 novice staff and was recruited from a group of undergraduate students who took a course in ASD and ABA that included hands on training in the classroom along with relevant course lectures. Novice staff were 75% female and 25% male with a mean age of 21.9 (SD=2.31) years. The mean number of months that novices provided direct ABA services was 3 months. Individuals in the novice group received behavioral skills training, didactics, and verbal instruction in the implementation of discrete trial instruction (DTI) from their classroom teachers prior to participation in the study. The second group contained 42 experienced staff, defined as having one or more years of ABA experience, whether at the DDDC or elsewhere. Experienced staff trained at the DDDC also received a combination of behavioral skills training, didactics, and verbal instruction. Experienced staff were 90.5% female and 9.5% male with a mean age of 29.9 (SD=6.92) years. The mean number of months that experienced staff provided direct ABA services was 73 months, and the median was 53 months (Table 1). Participants who completed the study procedures were entered into a raffle to win one DDDC t-shirt, which was awarded when recruitment ended.

All sessions took place during the regular school day in the staff member's classroom. IRB approval was obtained for this study, and informed consent was obtained for each participant.

**Materials**

Materials for the study included a video camera to record observation sessions as well as several measures. A TI check sheet was used in order to code TI from the behavioral observations (Appendix A). The check sheet listed all possible implementation components with boxes to check off implementation as correct, incorrect, or not applicable. Each component was listed ten times for a total of ten trials of DTI.

Staff behaviors included a) establishing ready behavior, b) secure attending, c) presenting clear instructions, d) praising and/or providing a tangible or edible contingent on a correct response, e) using a controlling prompt following an incorrect or no response, and f) continuing with demands and ignoring or blocking problem behavior.

Ready behavior was defined as staff waiting to present an instruction until the student is not engaging in any disruptive movements of the limbs and is oriented towards the teacher for at least one second. Secure attending was defined as requiring the student to look at the teacher or materials before presenting the instruction. Presenting clear instructions was defined as presenting the instruction once in a concise way that clearly specifies the target behavior. Praising and/or providing tangible or edible reinforcement contingent on a correct response was defined as delivering praise and/or a tangible or edible item within 5 seconds of a correct unprompted or prompted response. Using a controlling prompt following an incorrect or no response was defined as providing a prompt that evokes a correct response within 3 seconds of the instruction following an incorrect response and within 10 seconds of the instruction following no response. Finally, continuing with demands and ignoring or blocking problem behavior was defined as the staff member removing verbal or physical attention and minimizing facial

expressions following problem behavior while continuing with the current trial. If the behavior was dangerous, the staff member was required block the behavior (or inform a teacher who is qualified to do so) and rearrange the environment or use the minimum amount of physical interaction necessary to keep the student safe. The TI data sheet and definitions were adapted from Carroll et al., (2013) based on standard DDDC practices.

Treatment integrity was calculated by dividing the total number of correctly implemented components by the total number of opportunities to implement the components over the course of the observation period and multiplying the quotient by 100 in order to obtain a percentage. Percent errors were calculated for each of the six TI categories by dividing the total number of incorrectly implemented components by the total number of opportunities to implement the components over the course of the observation period and multiplying the quotient by 100 in order to obtain a percentage.

A twelve-question multiple choice quiz was designed to examine procedural and theoretical knowledge of the components of DTI (Appendix B). The quiz contained two questions for each of the six TI components listed on the TI check sheet. For each set of two questions per component, one question probed the individual's theoretical knowledge of the component (e.g. the definition of reinforcement) and one question probed the individual's procedural knowledge of the component (e.g. how to use reinforcement). The quiz was constructed and validated based on the literature on TI as well as the expert consensus of five Board Certified Behavior Analysts (BCBAs) who determined that each question reflected knowledge of the designated TI component. The percentage of incorrectly answered quiz questions for each staff member was calculated by dividing the total number of incorrect answers by the total number of questions on the quiz and

multiplying the quotient by 100. The percentages of incorrectly answered procedural and theoretical questions were calculated in an identical manner, except that only the six procedural or six theoretical questions were used, rather than the total number of questions. Total errors for specific TI components on the quiz (e.g. reinforcement) were calculated for each component by adding the total number of questions answered incorrectly for that component.

Finally, a brief self-report questionnaire was designed to assess level of direct ABA experience and training, as well as to gather basic demographic information such as age and gender. Information obtained by the questionnaire included number of years/months the participant had provided direct ABA services, whether the individual has obtained their BCBA credential, whether the participant had taken any classes in ABA, and the type of training they had received (e.g. coursework, behavioral skills training, verbal instruction, or video instruction) (Appendix C).

**Procedure**

Each participant was observed for a single DTI session with a student consisting of ten trials. Each participant was familiar with both the student and DTI program used for the session. All sessions were video recorded for coding purposes. The staff member was informed that the purpose of the observation was to examine teacher-student interactions during skill acquisition and was instructed to conduct their teaching procedures as they normally would. Staff were not told the true purpose of the observation in order to control for reactivity. Treatment integrity data were coded from the video recordings using the TI check sheet after the sessions had concluded. At the end of the observation, participants were asked to complete a brief quiz and survey and to

return these to the primary investigator when finished. The purpose of the survey was to collect information on the staff member's direct ABA experience and training and to verify that they were indeed novice or experienced staff based on number of years of ABA experience. The purpose of the quiz was to examine the staff member's procedure and theoretical knowledge of DTI.

**Response measurement and Interobserver agreement**

Each staff member was observed while working individually with a student for ten trials of DTI instruction during regularly scheduled time for academic instruction. Observers collected data on several components of staff behavior by coding the teacher's response as either correct or incorrect for each trial on TI data sheets

The primary investigator, as well as undergraduate research assistants with at least 8 months of experience coding TI, coded TI for each individual staff member from video recordings of the observation sessions. Undergraduate research assistants were trained in the collection of TI data using verbal instruction, modeling, and sample videos until they achieved a minimum of 90% IOA with the principal investigator. Undergraduate research assistants also served as independent observers in order to obtain interobserver agreement (IOA). IOA was collected for 35% of observation sessions. Exact IOA was calculated by comparing the TI data sheets for observer 1 and observer 2 for each trial. The number of components where both observers agree were divided by the total number of components. An agreement was defined as both observers recording correct implementation, incorrect implementation, or as not applicable for the same component.

**Data Analysis**

A statistical power analysis was performed for sample size estimation. With an

alpha of .05 and a power of 0.95, the required final sample size needed to detect a large sized effect was estimated to be approximately 84 participants (G*Power 3.1).

Chi square tests for independence were conducted to compare group with demographic variables such as gender and whether individuals had obtained their BCBA credential. Independent-samples t-tests were conducted to compare group with demographic variables such as age and number of months they had provided direct ABA services.

Independent-samples t-tests were conducted to determine whether the type of TI error (i.e. the six categories defined and listed on the TI check sheet) made differed based on group (i.e. level of experience). An independent sample t-test was also conducted in order to determine whether the total number of TI errors made was related to level of experience.

Bivariate Pearson correlations were conducted in order to determine whether TI errors were related to quiz errors on each DTI component. Additional bivariate Pearson correlations were conducted in order to determine whether the correlation between quiz errors and TI errors was significant for each group. Bivariate Pearson correlations were therefore conducted for both the experienced group and novice group separately.

Independent-samples t-tests were conducted to determine whether novice staff and experienced staff differed on their quiz performance, as well as whether the two groups differed on their quiz performance on procedural vs. theoretical questions.

Finally, paired t-tests were conducted to determine whether there were differences within groups on theoretical and procedural quiz questions.

**III. Results**

**Data Analysis**

The results of the chi square tests indicated that there were no significant differences between groups for gender or number of individuals who had received the BCBA credential. The results of the independent-samples t-tests indicated that individuals in the experienced group were significantly older ($t(68) = 6.959$, $p<.01$) and had provided direct ABA services for a significantly greater length of time ($t(68) = 7.803$, $p<.01$) than individuals in the novice group.

The results of the independent-samples t-tests examining the relationship between type of TI errors and group indicated that there was a significant difference in the percentage of errors made on implementing controlling prompts, such that experienced staff made significantly fewer errors than novices ($t(68) = -2.066$, $p<.01$). There were no significant differences between groups for percent of errors made on the following categories: ready behavior, secure attending, clear instructions, or problem behavior. Furthermore, results of an independent-samples t-test indicated that there was no significant difference in number of TI errors made between groups $t(68) =-1.387$, $p=.170$ (Table 2).

Results of the Pearson correlation indicated that there was a significant positive correlation between the total percentage of TI errors and the total percentage quiz errors $r(68)=.386$, $p<.01$ (Table 3). However, there were no significant correlations between errors on a particular quiz component (e.g. questions on reinforcement) and the corresponding TI component (e.g. performance on providing reinforcement).

Bivariate Pearson correlations were conducted for both the experienced group and novice group separately in order to determine whether the correlation between quiz errors

and TI errors was significant for each group. Results of the Pearson correlation indicate that there was a significant positive correlation between the total percentage of TI errors and the total percentage quiz errors in the experienced group $r(40)=.467$, $p<.01$. However, there was no significant correlation between quiz errors and TI errors for the novice group $r(26)=.274$, $p=.158$.

Due to the fact that the eligibility criteria for the experienced group was having provided at least 1 year of direct ABA services, participants in this group had varying levels of expertise ranging from 1 year of experience to 21.5 years of experience. The median number of months that experienced staff provided direct ABA services was 53 months. The experienced group was split into two subgroups based on the median split of length of ABA experience, such that one subgroup contained individuals with 12-53 months of experience and one subgroup contained individuals with 53.1+ months of experience. Two bivariate Pearson correlations of quiz errors and TI errors were conducted to determine whether one of these subgroups was responsible for the significant correlation found earlier between these two variables. Results of the Pearson correlation for the 12-53 month subgroup (n=21) indicated that there was no significant correlation between quiz errors and TI errors $r(19)=.314$, $p=.166$. However, results of the Pearson correlation for the 53.1+ month subgroup (n=21) indicated that there was a significant positive correlation between quiz errors and TI errors $r(19)=.596$, $p<.01$.

Next, an independent-samples t-test was conducted to determine whether novice staff and experienced staff differed on their quiz performance. Results of the t-test indicated that novice staff made significantly more errors on the quiz than experienced staff $t(68)=-6.641$, $p<.01$ (Table 4).

Further independent-samples t-tests were conducted to determine whether novice staff and experienced staff differed on their quiz performance on procedural vs. theoretical questions. Results of the t-tests indicated that novice staff made significantly more errors than experienced staff on both the procedural questions ($t(68)=-4.952$, $p<.01$) and theoretical questions $t(68)=-4.983$, $p<.01$ (Table 5).

Paired t-tests were conducted to determine whether there were differences within groups on theoretical and procedural quiz questions. The results indicated that there was no significant difference in the percentage of errors on theoretical vs. procedural questions in the experienced group $t(41)=1.374$, $p=.177$ (Table 6). However, there was a significant difference in the novice group, such that novice staff made significantly more errors on procedural questions than theoretical questions $t(27)=2.217$, $p<.05$ (Table 7).

**Interobserver Agreement**

Interobserver agreement data were collected for 35% of sessions on TI using an electronic coding document. Average IOA for these sessions was 97%, with a range of 85%-100%.

**IV. Discussion**

TI has become a topic of increasing interest in the field of ABA in recent years (Neely et al., 2015). In the current investigation, the goal was to extend the literature on TI by examining the extent to which types and number of TI errors in skill acquisition differ depending on level of experience. The findings for each research question will be discussed in turn.

**Question 1: Do the types and number of TI errors in DTI differ based on level of experience?**

The results of the current study indicated that there was no significant difference in the number of TI errors made between groups. That is, participants in the novice and experienced groups performed equally well in their implementation of skill acquisition programs in the classroom. On average, the novice group made TI errors 11% of the time while the experienced group made TI errors 8% of the time. This lack of a significant difference may be due to a number of factors. On one hand, the experienced group may have performed as poorly as the novice group due to therapist drift, in which treatment agents gradually alter the treatment over time, through failure to perform the important elements of the treatment or with adding in additional components that are not included in the intervention plan. On the other hand, the novice group may have performed as well as the experienced group due to the fact that they had received one-on-one training on how to implement skill acquisition programs through lectures and behavioral skills training within 1-2 months of data collection. In addition, the training the novices received may represent a more thorough and personalized approach than what is typically found in other settings, such as public schools. It is difficult to determine which hypothesis may be accurate, as no research has been published on the average percentage of TI errors made by staff during typical skill acquisition programming.

In addition to examining the total number of TI errors across groups, each individual TI component (e.g. secure attention, clear instructions) was examined in order to determine whether there were any differences between groups. The results indicated that novice staff and experienced staff performed equally well on all components except

one. There was a significant difference in the percentage of errors made on implementing controlling prompts, such that novice staff made a significantly greater number of errors than experienced staff. This finding replicates the results from Carroll et. al (2013), who found that controlling prompts were often implemented with lower integrity when compared to other components of DTI. The difference in errors in only one component of the current study may be due to the nature of this TI category, as implementing controlling prompts is seen as more challenging than any of the other measured TI variables. Unlike the other TI variables (establishing ready behavior, securing attending, presenting clear instructions, providing reinforcement, and working through problem behavior), which typically require the same procedure regardless of which skill acquisition program is being implemented, controlling prompts often vary. Depending on the skill acquisition program, the controlling prompt may consist of a verbal prompt, gestural prompt, or physical prompt. This lack of standardization across programs may result in a greater number of errors when implemented by novice staff.

**Question 2: Does knowledge or understanding of the rationale for each component of DTI reflect errors in each area?**

Several analyses were conducted in order to determine whether TI errors were related to quiz errors on each DTI component (e.g., secure attending, clear instruction, etc.). A significant positive correlation was found between the total percentage of TI errors and total percentage of quiz errors, indicating that individuals who made a greater number of TI errors also tended to make a greater number of errors on the quiz. However, there were no significant correlations between individual quiz and TI components. This is a particularly important finding, considering that it is common for professional

development trainings to use a pre-post quiz to measure performance. These results suggest that although general quiz performance is related to performance in the classroom, knowledge of a specific component does not necessarily impact performance on that component. Therefore, trainings that focus on a particular TI component (e.g., how to provide reinforcement) and use pre-post quizzes to measure attained knowledge may not translate to actual improved performance of that component in the classroom.

Additional analyses were conducted in order to determine whether the significant correlation between quiz errors and TI errors was significant for each group. The results indicated that quiz errors and TI errors were significantly correlated for the experienced group only. However, the lack of significance in the novice group may have been impacted by the smaller sample size (n=28) compared to the experienced group (n=42). In addition, the experienced group contained participants with a large range of experience varying from 1-21.5 years. Therefore, the experienced group was split into two subgroups based on a median split of 53 months providing direct ABA services. Additional analyses of quiz errors and TI errors were conducted in order to determine whether one of these subgroups was responsible for the significant correlation found earlier between these two variables. Although there was no significant correlation in the 12-53 month subgroup, there was a significant positive correlation between quiz errors and TI errors in the 53.1+ group, indicating that quiz and TI performance are only correlated in staff that have been working in the field for a number of years. This may be due to the fact that longer-term staff may have a greater amount of exposure to trainings and professional development workshops on both knowledge of as well as implementation of TI.

**Question 3: Are the number of quiz errors related to level of experience?**

The results indicated that novice staff made significantly more errors on the quiz than the experienced staff. The quiz questions were divided into procedural and theoretical questions to determine whether the two groups differed on performance on these two sets of questions. The results indicated that the novice group made significantly more errors on both procedural and theoretical questions when compared to the experienced group. This finding is unsurprising, as all novices were enrolled in an introductory ABA class. Novices therefore had only a few months' worth of knowledge and exposure to ABA, whereas the experienced group has a range of 1-21.5 years of experience, and attended several required trainings each year.

Finally, analyses were conducted in order to determine whether there were differences within groups on the number of errors made on theoretical and procedural questions. The results indicated that there was no significant difference in the number of errors made on theoretical vs. procedural questions in the experienced group. However, there was a significant difference in the novice group, such that novices made significantly more errors on the procedural questions as compared to the theoretical questions. As mentioned above, participants in the novice group were all enrolled in an introductory ABA class, which focused primarily on theory and used quizzes to test knowledge. Therefore, participants in the novice group may have had their theoretical knowledge tested using a quiz format previously, whereas their procedural knowledge of ABA was tested exclusively in vivo, rather than through a written quiz. In addition, some of the theoretical questions may have been covered in previous undergraduate courses

(e.g. General Psychology, Learning Processes), whereas it was only within the start of this ABA course that they were trained on TI procedures.

**Strengths, Limitations and Future Directions**

The present study extends previous research on TI in a number of ways. This study was the first to examine the relationship between TI and level of experience, and was the first to examine how knowledge of the components of TI relates to actual performance in the classroom. Furthermore, this study examined TI through individual treatment components rather than a single global integrity measure, which allowed for detailed analyses of integrity errors. Although global integrity scores are often used in TI research, they may mask poor performance on individual components as these errors may be averaged out (Cook, Subramaniam, Brunson, Larson, Poe, & Peter, 2015). Certain limitations, however, require that these findings be interpreted with caution.

This study only included staff that worked at the DDDC, a university-based school specializing in ABA for individuals with ASD. Both the novice and experienced staff were recruited from the school. Staff at the DDDC may differ from staff in the general community in that they receive rigorous training, have frequent contact with BCBAs, and work in a setting with a high level of resources. Further investigation is necessary to determine whether the results from this study would generalize to staff in other settings, such as public school special education classrooms or clinic-based settings.

A power analysis indicated that the required sample size needed to detect a large sized effect was estimated to be approximately 42 individuals per group. Although this recruitment goal was achieved with the experienced group, the novice group contained only 28 participants. Therefore, these results should be interpreted with caution, as the

sample size was smaller than what the power analysis indicated was necessary. Future studies could address this weakness by recruiting and evaluating a larger number of participants.

This study was also limited by the fact that each participant was observed for TI only once. According to Gresham, Dart, and Collins (2017), a minimum of four behavior samples are necessary in order to produce reliable TI data. Furthermore, participants in this study were not required to conduct the same DTI task during the behavior sample. Some DTI programs may be more difficult than others to implement, which may have skewed the results. In addition, student differences (such as the extent to which staff needed to manage challenging behavior) may have influenced the results. Future studies could examine differences in TI errors among staff with varying levels of experience using a standardized DTI task with a specific set of students.

**Training Implications**

Despite its limitations, the results of this study may provide guidelines for improving TI among staff. First, although a number of studies have examined the effects of varying levels of TI on student skill acquisition, the base rate of TI errors in schools (as well as other settings) is still unknown. Knowing the average rate of TI errors after training new staff would allow supervisors to determine when staff require additional or remedial trainings. This could be assessed via quick behavior samples of staff TI in order to determine what components of TI need to be targeted. Second, the results of this study indicate that quiz scores on particular components of TI may not generalize to practical skills in the classroom. This finding suggests that the use of pre and post tests in trainings and workshops may be a less effective manner to determine whether knowledge has been

attained. However, if quizzes are still used, the results of this study suggest that a larger focus should be placed on teaching the procedures involved in TI, as novices made significantly more errors on procedural questions than theoretical questions. Lastly, this study found that novice staff tend to make significantly more TI errors than experienced staff on implementing controlling prompts correctly. This finding suggests that a greater emphasis should be placed on this particular component when training new staff. Overall, these findings should be taken into consideration when training new or experienced staff as well as when conducting booster trainings, as this may increase the efficacy of these trainings and promote higher levels of TI in the long term.

Appendix A*: Treatment Integrity Check Sheet*

**Treatment Integrity Check Sheet**

**Staff Member: _____** **Date: _____**

**Observer: _____** **Time Observed: _____**

| **Criteria** | **Trial #** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Establish ready behavior:<br>*Teacher waits to present an instruction until the student is not engaging in any disruptive movements of the limbs and is oriented towards the teacher for at least one second* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Secure attending:<br>*Teacher requires the student to look at the teacher or materials before presenting the instruction* | | | | | | | | | | |
| Present clear instructions:<br>*Teacher presents the instruction once in a concise way that clearly specifies the target behavior* | | | | | | | | | | |
| Praise and/or provide a tangible or edible contingent on a correct response:<br>*Teacher delivers praise and/or a tangible or edible item within 5 seconds of a correct unprompted or prompted response* | | | | | | | | | | |
| Use a controlling prompt following an incorrect or no response:<br>*Teacher provides a prompt that evokes a correct response within 3 seconds of the instruction following an incorrect response and within 10 seconds of the instruction following no response* | | | | | | | | | | |
| Continue with demands and ignore or block problem behavior:<br>*Teacher does not provide verbal or physical attention and minimizes facial expressions following problem behavior while continuing with the current trial.*<br>*If the behavior is dangerous, the teacher should block behavior* (or inform a teacher who is qualified to do so) *and rearrange the environment or use the minimum amount of physical interaction necessary to keep the student safe* | | | | | | | | | | |

*Code "+" for correct implementation, "-" for incorrect implementation, and "N/A" if the component is not applicable*

$$\text{Treatment Integrity} = \frac{\text{Total \# correctly implemented components}}{\text{Total \# applicable components}} \times 100 =$$

Appendix B: *DTI Quiz*

**For each question, select the best answer.**

1. When establishing ready behavior, the teacher should:
    a. Wait until the surrounding area is free of distraction
    b. Wait until the student is sitting appropriately
    c. Present a clear instruction to the student
    d. Present a preferred item in order to entice the child to sit quietly

2. In order to secure attending, the teacher must:
    a. Instruct the student to sit appropriately with hands on lap and feet on the floor
    b. Require the student to look at the teacher or training materials
    c. Establish stimulus control
    d. Present the instruction

3. When presenting an instruction, the teacher should:
    a. First provide an edible to maintain motivation
    b. Repeat the instruction until the child responds
    c. Present the instruction only once
    d. Use very different wording each time in order to promote generalization

4. With regard to reinforcement all of the following rules should be followed EXCEPT:
    a. Deliver reinforcement immediately following a correct response
    b. Reinforcers should always be delivered non-contingently.
    c. Pair yourself with reinforcement.
    d. Appropriately use and fade reinforcement.

5. How should prompts be used in skill acquisition?
    a. Prompts should always be used so that the student does not get upset
    b. Prompts should be introduced and faded systematically
    c. Physical prompts should be used consistently
    d. The teacher should use different prompts depending on the mood of the student

6. When using extinction:
    a. Make sure it is used consistently
    b. Ignore the student as often as possible
    c. Never combine it with other procedures
    d. Provide aversive stimuli upon each instance of problem behavior

7. Which of the following is not a pre-attending skill?
    a. Visually attending
    b. Orienting body towards the teacher
    c. Appropriate sitting
    d. Smiling

8. Why is it important to get attending before you deliver the SD?
    a. It is extremely difficult to establish stimulus control if a student is not attending
    b. The reinforcer may not work
    c. Attending decreases the probability of problem behavior occurring in the moment
    d. Attending is necessary for motivation

9. What is a discriminative stimulus (SD)?
    a. A stimulus in the presence of which a particular response will not be reinforced
    b. A stimulus that alters the effectiveness of some object or event as reinforcement
    c. A stimulus that signals the availability of the reinforcer
    d. A stimulus that alters the frequency of behavior

10. How can you tell if an item functioned as a reinforcer?
    a. If the individual smiled while interacting with the item
    b. If the child appeared to enjoy consuming the item
    c. If the probability of the child requesting the item increases in the future
    d. If the behavior was actually repeated in the future after receiving the item

11. Why is it important to use prompts when you first start teaching a new skill?
    a. Prompts help provide errorless learning so students don't practice mistakes
    b. Prompts eliminate distractions
    c. Prompts signal the availability of reinforcement
    d. Prompts assist with generalization of responses to other environments over time

12. A procedure in which reinforcement of a previously reinforced behavior is discontinued and as a result, occurrences of that behavior decrease in the future is:
    a. Punishment
    b. Extinction
    c. Reinforcement
    d. Establishing Operation

Appendix C: ABA experience survey
**Please complete the following survey.**

1. Age: _____

2. Gender: _____

3. Position in classroom (circle one):
    a. Fieldworker
    b. Classroom assistant
    c. Lead instructor
    d. Head teacher

4. Number of years/months you've provided direct ABA services:
   Years:_____ Months:_____

5. Number of years/months you've provided direct ABA services at the DDDC:
   Years:_____ Months:_____

6. Do you have your BCBA credential?
    a. Yes
    b. No

7. Have you taken any classes in ABA?
    a. Yes
        i. If so, how many?
        ii. Were they part of a certified BCBA course sequence?
    b. No

8. What sort of training have you received (circle all that apply):
    a. Coursework
    b. Behavioral skills training (instruction modeling, role play/rehearsal, corrective feedback)
    c. Verbal instruction
    d. Video instruction
    e. Other
        i. Specify:
           _____

Table 1
*Demographic information*

| | Experienced | Novice | $t/\chi^2$ |
|---|---|---|---|
| *n* | 42 | 28 | - |
| Age (M) | 29.9 | 21.9 | t=6.959** |
| Age (SD) | 6.92 | 2.31 | - |
| % Male | 9.5 | 25.0 | $\chi^2$=3.038 |
| Months provided ABA (M) | 73.24 | 3.16 | t=7.803** |
| Months provided ABA (SD) | 58.17 | 1.66 | - |
| % with BCBA | 4.76 | 0.00 | $\chi^2$=1.373 |

* p < .05 **p < .01

Table 2
*Independent-samples t-tests for TI errors and group*

| | Expert | | Novice | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | t (68) |
| Ready Behavior | 6.476 | 15.31974 | 4.6429 | 9.61563 | .615 |
| Secure Attending | 5.4762 | 12.53335 | 10.7143 | 18.64454 | -1.303 |
| Clear Instructions | 1.6905 | 4.94106 | 1.7857 | 6.11832 | -.069 |
| Reinforcement | 15.5452 | 21.85926 | 25.7143 | 34.14876 | -1.397 |
| Controlling Prompts | 4.2381 | 16.64524 | 17.5714 | 31.32202 | -2.066* |
| Problem Behavior | 2.3810 | 15.43033 | 0.0000 | 0.00000 | 1.000 |
| Total Errors | 7.8262 | 9.59535 | 11.4214 | 12.02243 | -1.326 |

* $p < .05$

Table 3
*Bivariate Pearson correlations for TI errors and quiz errors*

| | Ready Behavior (TI) | Secure Attending (TI) | Clear Instructions (TI) | Reinforcement (TI) | Controlling Prompts (TI) | Problem Behavior (TI) | Total Errors (TI) |
|---|---|---|---|---|---|---|---|
| Ready Behavior (quiz) | -.088 | .035 | .024 | -.005 | .097 | .213 | .004 |
| Secure Attending (quiz) | .069 | -.039 | .060 | .290* | .058 | .089 | .209 |
| Clear Instructions (quiz) | .333** | .237* | .081 | .085 | .092 | -.076 | .235 |
| Reinforcement (quiz) | .115 | .008 | -.145 | .201 | .182 | -.089 | .213 |
| Controlling Prompts (quiz) | .059 | .369** | .195 | .083 | .083 | -.046 | 239* |
| Problem Behavior (quiz) | .152 | .123 | -.121 | .358** | -.019 | -.065 | .347** |
| Total Errors (quiz) | .233 | .203 | .019 | .319** | .163 | -.002 | .386** |

* $p < .05$ **$p < .01$

Table 4
*Independent-samples t-tests for quiz errors and group*

| | Expert | | Novice | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | t (68) |
| Errors on Quiz (%) | 9.8333 | 10.13065 | 28.3214 | 13.11785 | -6.641* |

* $p < .05$

Table 5
*Independent-samples t-tests for type of quiz error and group*

| | Expert | | Novice | | |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | t (68) |
| Procedural Questions (%) | 11.5952 | 12.99354 | 32.7857 | 20.00436 | -4.952** |
| Theoretical Questions (%) | 8.3571 | 12.29876 | 23.8214 | 12.99303 | -4.983** |

* $p < .05$ **$p < .01$

Table 6

*Paired-samples t-tests for type of quiz error within the expert group*

| | Mean | SD | t (41) |
|---|---|---|---|
| Procedural Questions (%) | 11.5952 | 12.99354 | 1.374 |
| Theoretical Questions (%) | 8.3571 | 12.29876 | |

* $p < .05$

Table 7

*Paired-samples t-tests for type of quiz error within the novice group*

| | Mean | SD | t (27) |
|---|---|---|---|
| Procedural Questions (%) | 32.7857 | 20.00436 | 2.217* |
| Theoretical Questions (%) | 23.8214 | 12.99303 | |

* $p < .05$

**References**

Addis, M. E. (2002). Methods for disseminating research products and increasing evidence-based practice: Promises, obstacles, and future directions. *Clinical Psychology: Science and Practice, 9*(4), 367–378. https://doi.org/10.1093/clipsy/9.4.367

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of applied behavior analysis, 1(*1), 91-97. https://doi.org/10.1901/jaba.1987.20-313

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. https://doi.org/10.1002/jaba.49

Cochrane, W. S., & Laux, J. M. (2008). A survey investigating school psychologists' measurement of treatment integrity in school-based interventions and their beliefs about its importance. *Psychology in the Schools, 45*, 499–507. https://doi.org/10.1002/pits.20319

Codding, R. S., Feinberg, A. B., Dunn, E. K., & Pace, G. M. (2005). Effects of immediate performance feedback on implementation of behavior support plans. *Journal of Applied Behavior Analysis*, *38*(2), 205-219. https://doi.org/10.1901/jaba.2005.98-04

Cook, J. E., Subramaniam, S., Brunson, L. Y., Larson, N. A., Poe, S. G., & Peter, C. C. S. (2015). Global measures of treatment integrity may mask important errors in discrete-trial training. *Behavior analysis in practice, 8*(1), 37-47. https://doi.org/10.1007/s40617-014-0039-7

Cooper, J. O. Heron, T. E., & Heward, W. L. (2007). c. Upper Saddle River, New Jersey: Pearson Merrill Prentice Hall.

Corsello, C. M. (2005). Early intervention in autism. *Infants & Young Children, 18*(2), 74-85. https://doi.org/10.1097/00001163-200504000-00002

Detrich, R. (1999). Increasing treatment fidelity by matching interventions to contextual variables within the educational setting. *School Psychology Review, 28*(4).

Detrich, R. (2014). Treatment integrity: Fundamental to education reform. *Journal of Cognitive Education and Psychology, 13*(2), 258-271. https://doi.org/10.1891/1945-8959.13.2.258

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. https://doi.org/10.1901/jaba.2010.43-291

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. https://doi.org/10.1901/jaba.2007.40-447

DiGennaro, F. D., Martens, B. K., & McIntyre, L. L. (2005). Increasing treatment integrity through negative reinforcement: Effects on teacher and student behavior. *School Psychology Review*, *34*(2), 220.

Goense, P. B., Boendermaker, L., & van Yperen, T. (2016). Support systems for treatment integrity. Research on Social Work Practice, 26(1), 69-73. https://doi.org/10.1177/1049731515579205

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School psychology review*.

Gresham, F. M., Dart, E. H., & Collins, T. A. (2017). Generalizability of multiple measures of treatment integrity: Comparisons among direct observation, permanent products, and self-report. *School Psychology Review, 46*(1), 108-121. https://doi.org/10.17105/spr46-1.108-121

Gresham, F. M., Gansle, K. A., Noell, G. H., & Cohen, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychology Review*.

Gresham, F. M., MacMillan, D., Beebe-Frankenberger, M. E. & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented*? Learning Disabilities Research and Practice, 15,* 198-205. https://doi.org/10.1207/sldrp1504_4

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. https://doi.org/10.1007/s10864-013-9181-5

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. https://doi.org/10.1901/jaba.2007.659-672

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. https://doi.org/10.1177/1098300715573629

Neely, L., Davis, H., Davis, J., & Rispoli, M. (2015). Review of reliability and treatment integrity trends in autism-focused research. *Research in Autism Spectrum Disorders*, *9*, 1-12. https://doi.org/10.1016/j.rasd.2014.09.011

Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly*, *12*(1), 77. https://doi.org/10.1037/h0088949

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., . . .Resetar, J. L. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review, 34*, 87–106.

Northup, J., Fisher, W., Kahang, S. W., Harrell, R., & Kurtz, P. (1997). An assessment of the necessary strength of behavioral treatments for severe behavior problems. *Journal of Developmental and Physical Disabilities*, *9*(1), 1-16. https://doi.org/10.1007/s10882-017-9563-z

Parsons, M. B., & Reid, D. H. (1995). Training residential supervisors to provide feedback for maintaining staff teaching skills with people who have severe disabilities. *Journal of applied behavior analysis, 28*(3), 317-322. https://doi.org/10.1901/jaba.1995.28-317

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. https://doi.org/10.1002/jaba.238

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice*, *12*(4), 365-383. https://doi.org/10.1093/clipsy.bpi045

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. https://doi.org/10.1901/jaba.1982.15-477

Progar, P. R., Perrin, F. A., DiNovi, B. J., & Bruce, S. S. (2001). Treatment integrity: Some persistent concerns and some new perspectives. *The Behavior Analyst Today*, *2*(1), 28. https://doi.org/10.1037/h0099912

Sanetti, L. M. H., Luiselli, J. K., & Handler, M. W. (2007). Effects of verbal and graphic performance feedback on behavior support plan implementation in a public

elementary school. *Behavior Modification*, *31*(4), 454-465. https://doi.org/10.1177/0145445506297583

Sterling-Turner, H. E., Watson, T. S., Wildmon, M., Watkins, C., & Little, E. (2001). Investigating the relationship between training type and treatment integrity. *School Psychology Quarterly*, *16*(1), 56. https://doi.org/10.1521/scpq.16.1.56.19157

Vollmer, T. R., Roane, H. S., Ringdahl, J. E., & Marcus, B. A. (1999). Evaluating treatment challenges with differential reinforcement of alternative behavior. *Journal of Applied Behavior Analysis*, *32*(1), 9-23. https://doi.org/10.1901/jaba.1999.32-9

Vollmer, T. R., Sloman, K. N., & Pipkin, C. S. P. (2008). Practical implications of data reliability and treatment integrity monitoring. *Behavior analysis in practice*, *1*(2), 4. https://doi.org/10.1007/bf03391722

Wilkinson, L. A. (2007). Assessing treatment integrity in behavioral consultation. *International Journal of Behavioral Consultation and Therapy*,*3*(3), 420. https://doi.org/10.1037/h0100816

Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: strength, integrity, and effectiveness. *Journal of consulting and clinical psychology, 49*(2), 156. https://doi.org/10.1037/10109-050