1-6-2017

Effects of Errors of Commission on Student Performance During Discrete Trial Tasks

Carina M. De Fazio
Georgia State University

ACCEPTANCE

This dissertation, EFFECTS OF ERRORS OF COMMISSION ON STUDENT PERFORMANCE DURING DISCRETE TRIAL TASKS, by CARINA M. DE FAZIO, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree, Doctor of Philosophy, in the College of Education and Human Development, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chairperson, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty.

_____
Paul Alberto, Ph.D.
Committee Chair

_____
Laura Fredrick, Ph.D.
Committee Member

_____
Nicole Patton-Terry, Ph.D.
Committee Member

_____
Lauren Boden, Ph.D.
Committee Member

_____
Date

_____
Laura Fredrick, Ph.D.
Chairperson, Department of Educational Psychology, Special Education, and Communication Disorders

_____
Paul Alberot, Ph.D.
Dean
College of Education
and Human Development

AUTHOR'S STATEMENT

By presenting this dissertation as a partial fulfillment of the requirements for the advanced degree from Georgia State University, I agree that the library of Georgia State University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote, to copy from, or to publish this dissertation may be granted by the professor under whose direction it was written, by the College of Education and Human Development's Director of Graduate Studies, or by me. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without my written permission.

_____

Carina Marie De Fazio

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Carina Marie De Fazio
Department of Educational Psychology, Special Education, and Communication Disorders
College of Education and Human Development
Georgia State University

The director of this dissertation is:

Paul Alberto
Department of Educational Psychology, Special Education, and Communication Disorders
College of Education and Human Development
Georgia State University
Atlanta, GA 30303

CURRICULUM VITAE

Carina M. De Fazio

ADDRESS: 60 Oakridge Ave SE
Atlanta, GA 30317

EDUCATION:

Ph.D.	2016	Georgia State University Department of Educational Psychology, Special Education, and Communication Disorders
M.Ed.	2005	Georgia State University Multiple and Severe Disabilities, Autism
B.S.	1997	Smith College Psychology

PROFESSIONAL EXPERIENCE:

2012-2104	Fellowship on Disability and Disparity 2nd Century Initiative, Georgia State University
2008-2012	Editorial Assistant *Focus on Autism and Other Developmental Disabilities*
2005-2011	Trainer, Technical Assistance for Severe Behavior Cobb County School District, Marietta, GA

PRESENTATIONS AND PUBLICATIONS:

Duchaine, E.L., & **De Fazio, C. M.,** (2010, December). *Behavior specific praise statements: small teacher change leads to big classroom changes.* Positive Behavior Interventions and Supports: Making a Difference in Students' Lives, Atlanta, Georgia

Duchaine, E. L., & **De Fazio, C. M.** (2010, October). *Response cards: Increase student engagement & decrease challenging behaviors in juvenile justice settings*. Teacher Education for Children with Behavior Disorders, Tempe, Arizona

Duchaine, E.L., **De Fazio, C. M.,** & McDaniel, S. C. (2010, April). *Motivate & engage students with behavior specific praise: Easy as 1-2-3!* Council for Exceptional Children, Nashville Tennessee

De Fazio, C.M., Whitmarsh, E.L., & Jefferson, A.E. (2008, February). *Writing a BIP with an escape from demand function based on an undifferentiated FBA.* Poster presented at Applied Behavior Analysis International –Autism Conference, Atlanta, GA.

Whitmarsh, E.L, **De Fazio, C.M**., Jefferson, A.E., & Friedly, N. (2008, February). *The application of a technical assistance model within a Metro-Atlanta area school district*. Poster presented at Applied Behavior Analysis International –Autism Conference, Atlanta, GA.

Hayes, G.R., Heflin, J., Abowd, G.D., Gardere, L.M., **DeFazio, C.M**., Pittman, C., Pirouz, N, and Hooda, S. (February, 2007). *Evaluating a selectively archived video recording system for functional behavior assessment in schools*. Poster presented at Applied Behavior Analysis International –Autism Conference, Boston, MA.

PROFESSIONAL SOCIETIES AND ORGANIZATIONS

2009-present	Council for Exceptional Children - Teacher Education
2008-present	Council for Exceptional Children - Autism and Developmental Disabilities
2008-present	Council for Exceptional Children - Teacher Education Division
2002-present	National Education Association

Effects of Errors of Commission on Student Performance During Discrete Trial Tasks

by

Carina M. De Fazio

Under the Direction of Paul Alberto

ABSTRACT

The extent to which interventions are implemented as intended is called treatment integrity (TI). Given that it is unreasonable to expect 100% TI in applied settings, a through understand of TI is essential to appropriately train teachers. This understanding must include the types of TI errors that may be committed and how these effect student learning. It is essential to study TI errors because of the real-world implications for students, including eligibility decisions for special education services, which are based upon students' responses to inventions. It is not possible to make educational decisions on intervention effectiveness unless it is clear that evidence-based practices have been implemented accurately. If TI is low, it is impossible to determine which services and interventions a student requires.

Further, measuring the fidelity with which interventions are applied allows for a more thorough and accurate understanding of which components of an intervention are effective, necessary, and feasible. A broader understanding of which TI errors are most significant, as well as measuring the necessary levels of TI, will lead to more accurate information about how to implement evidence-based practices.

The purpose of this study is to gain a more nuanced understanding of TI failures in the form of errors of commission and the role commission errors have on participant responsiveness (Power, 2005). This study extends the results of DiGennaro Reed et al. (2011) by including an intermediate (80%) level of TI which may more accurately represent an attainable level of TI in

applied settings. For two of four students, more errors of commission were related to lower skill acquisition. For the other two students, idiosyncratic patterns of responding emerged.

INDEX WORDS: Treatment integrity, Fidelity, Errors of commission

EFFECTS OF ERRORS OF COMMISSION ON STUDENT PERFORMANCE DURING DISCRETE TRIAL TASKS

by

CARINA DE FAZIO

A Dissertation

Presented in Partial Fulfillment of Requirements for the

Degree of

Doctor of Philosophy

in

Education of Students with Exceptionalities

in

Department of Educational Psychology, Special Education, and Communication Disorders

in

the College of Education and Human Development

Georgia State University

Atlanta, GA

2016

DEDICATION

This dissertation is dedicated to my grandparents, Peter and Teresa De Fazio and Joseph and Rita DellaCamera. I was lucky to have had you as long as I did. Your example of how to be good inspires me still.

ACKNOWLEDGMENTS

This dissertation would not have been possible without the support and guidance of many people. Each of you has provided something essential along the way, and for that I am grateful. First, thank you to Dr. Paul Alberto for taking me on so late in my program. Your good humor and encouragement have been the tonic I've needed to sustain myself through the more anxious moments. Thank you to Dr. Lauren Boden, Dr. Laura Fredrick, and Dr. Nicole Patton-Terry for your time and valuable input on this dissertation. And of course, Dr. Juane Heflin will always have my deepest thanks and admiration for her commitment to quality work, her dedication to her students, and for the care, love, and enthusiasm she has for the ones she prepares us to teach. The lessons she taught me will last a life time. I can only hope to pass on some of what she has given me.

Thank you to the students with whom I have worked over the years, including the beautiful children in this study. I am indebted to you, and to your families for the hard work you do each day to learn and to teach those like me who want to help. Your patience and perseverance are a model to me every day.

Thank you to my GSU friends Ellen Duchaine, Christy Fain, Jackie Isbell, Brooks Peters, Nicole Mays, and Ginny Van Rie for friendship, laughter, advice and help. To all of women in my life who have helped with childcare, friendship, honesty, and wine, you have each given me gifts that I will treasure forever. Thank you to Sheilagh Cullen, Tatjana Krause, Lynn Santiago-Calling, Theresa Stowe, Mary Watkins, and Amycla Webb for being my people.

Thank you to all of my family- Susan and John Cole, Justin and Jamie Cole, Perry and Julie De Fazio, Kath and Steve Robb, and to all of my nieces and nephews. All of you have supported me in many ways during this process. No words can express my gratitude to my

parents Elaine and Anthony De Fazio for never giving up on me, and to my husband, Jeremy Cole for never letting me give up on myself. To my daughter Elena Cole, let this be a lesson to you – you CAN do whatever you put your mind to, my love. YOU are the reason Mama finished!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

IMPLEMENTING INTERVENTIONS WITH INTEGRITY: DON'T *JUST DO IT*, DO IT RIGHT

Introduction

Current educational law and policies require teachers to use evidence-based practices (EBPs) when providing instruction to individuals with disabilities (Alexander, Ayers, & Smith, 2015; Odom, Brantling, Gersten, Horner, Thompson, & Harris, 2005; Simpson, 2005). Although various terms have been used to refer to EBPs (e.g., empirically validated practices, scientifically based strategies, empirically supported treatments), the concept, in general, refers to practices based on professional judgment by practitioners to select and apply instructional strategies that have been demonstrated to be effective through empirical research (Slocum et al., 2014). EBPs are being advocated for in schools due to legal mandates, the wealth of underlying research supporting them, and because of the face validity of using them.

Increasing awareness of the need to use EBPs stems from both the No Child Left Behind Act of 2001 (NCLB) and the Individuals with Disabilities Education Act of 2004 (IDEA, 2004). NCLB specifically mentions "scientifically-based research" as a tool that should be used to devise educational interventions. In fact, this component is important enough to be mentioned more than 100 times in the Act (Browder & Cooper-Duffy, 2003). The *What Works Clearinghouse* (funded by the Institute for Education Science) was created out of this legislation as a repository for EBPs (Burns & Ysseldyke, 2009). Teachers are accountable to society, taxpayers, and most importantly, their students, for preparing the latter to participate in society in a meaningful way. In fact, this accountability has come to the fore as of late with teacher evaluations and compensation increasingly being tied to student performance (Hallinger, Heck,

& Murphy, 2014). The central idea behind pushing teachers to use EBPs is that these will lead to better outcomes for students.

When IDEA first became law in 1975, it brought much-needed attention to the education of students with disabilities and mandated that these students receive instruction tailored to their needs. While IDEA has changed names over time, the focus has remained the same – to provide a "free and appropriate" education for students with disabilities. In its current form, IDEA focuses on training teachers in EBPs as a significant component for improving education for students with disabilities (Cook, Tankersley, Cook, & Landrum, 2008). Perhaps due to this focus from IDEA, school districts are under pressure to provide teacher education about EBPs to teachers of individuals with disabilities (Alexander et al., 2015).

NCLB (2001) requires EBPs to be used in determining the allocation of resources and services for *all* students (Erchul, 2011). For students with disabilities, EBPs are required by IDEA (2004). Translating policy into practice in schools requires systematizing processes to adhere to the law. One method for systematizing EBPs in school systems is the Response to Intervention (RTI) model, which was written into federal law with the 2004 reauthorization of IDEA. RTI underscores the importance of using EBPs in instructional settings. In simple terms, RTI is a process that applies to all students in public schools and serves as a framework for educators in determining what learning strategies are effective for which students (Barnes & Harlacher, 2008; Fuchs & Fuchs 2005; Johnson, Mellard, Fuchs, & McKnight, 2006). The pivotal idea behind RTI is that no student should be labeled as requiring special education services, or presumed to have a disability, until it is clear that he or she has been taught using sound, evidence-based practices (Erchul, 2011; Fuchs & Fuchs, 2005). This is to protect a student who has bad teachers, ineffective curriculum, and/or classrooms with poor behavior

management from being labeled as needing *special* education, when, in truth, he or she has not received a sound *general* education. The current reauthorization of NCLB, now renamed the Every Student Succeeds Act (ESSA, 2015) made many changes, most notably a move from federal control of education to greater state control. However, the requirement to use EBPs remains.

In an environment where test scores and school and teacher accountability are the focus, it is not unheard of for students with problems to simply leave school, at times with encouragement from schools because of the phenomenon known as "pushout" (Doll, Eslami, & Walters 2013). Unfortunately, under current law, far too often schools are not incentivized to apply consistent methodology (EBPs) to student achievement. Rather, there are benefits to pushing students out of school, as students who leave a school therefore no longer "count against" a school. There is thus potentially an incentive for teachers and schools to not apply EBPs for low-performing students and students with disabilities. This underscores the need for the verification of use of EBPs in all classroom settings, and particularly in schools with high numbers of low-performing students (Glennie, Bonneau, Vandellen & Dodge, 2012).

In addition to system-wide challenges to implementing EBPs, there is also the challenge of EBPs being carried out, classroom by classroom. For example, there is evidence that teachers require significant time and support to adequately implement new EBPs (Boardman, Argüelles, Vaughn, Tejero Hughes, & Klingner, 2005; Stahmer, Reed, Lee, Reisinger, Connell, & Mandell, 2015). It is not hard to imagine a scenario in which a teacher is provided with an evidence-based reading program, yet does not have the time, training, or sufficient materials to implement the program. There is also evidence that, far too often, EBPs are not followed in classrooms (Borders & Bock, 2015). The reasons for this are many and well-debated in the literature with money,

time, and training all factoring in to what is commonly referred to as the research to practice gap (Carnine, 1997). The problem is not a new one, and new solutions continue to be proposed (See, for example, the March 2016 special issue of *Intervention in School and Clinic* which is dedicated to this topic).

Therefore, there is one persistent problem with RTI, as with many classroom interventions: how do we determine that the required evidence-based curricula and strategies are being implemented correctly? In other words, what is the role of treatment integrity (TI), defined as interventions being implemented as intended (Codding, Livanas, & Pace, 2007; Dane & Schneider, 1998; Gresham, 2005; Gresham, Gansle, Noell, Cohen, & Rosenblum, 1993; Perepletchikova & Kazdin, 2005). Sanetti & Fallon (2011) note the impossibility of drawing conclusions about the effectiveness of an intervention without assessing treatment integrity. Treatment integrity is an issue that affects both general and special education. When a new method or technology is introduced into the classroom, it is only as effective as the correct implementation of that method or technology by the teacher.

It is clear that the only way to understand if an intervention has contributed to student outcomes is by evaluating the implementation of all components of the intervention (Keller-Margulis, 2012). This is where assessing treatment integrity (TI) becomes necessary. Crucially, only when EBPs are implemented correctly can a determination of intervention effectiveness be made. In fact, one of the underlying assumptions behind EBPs is TI: "Because evidence-based practices assume fidelity of implementation (FOI), failure to establish FOI limits the conclusions that can be drawn from any outcome evaluation" (Missett & Foster 2015, p 1.).

In terms of special education, there is arguably an even greater need for TI. Educators regularly use information from RTI as well as behavior intervention plans (BIPs) to determine

educational placement. Placement and service decisions are supposed to be based upon the interventions that a student requires in order to make progress. If TI is low, there is no way to determine if services and interventions have been effective for that student, and no meaningful way to determine if services should be continued or modified, and in what manner. Further making the case for the need for TI, once a student has a documented need for special education support, the student is already behind his or her peers; there is no time to waste guessing about the effectiveness of interventions. Interventions must be shown to be implemented appropriately so that students receive the most benefit.

Treatment Integrity (TI) Review

The brief definition of TI offered earlier (an intervention being used as intended) is used often when discussing this complex concept. While having a succinct definition is useful in facilitating dialogue, the multi-faceted nature of TI can be lost if an overly-simplistic definition is the only one used. Contributing to the complexity of understanding the concept, the literature on TI is both widespread and discrete. TI research takes place in many fields that have significant overlap in subject matter as well as the populations served (behavioral health, psychotherapy, juvenile justice, substance abuse prevention, ABA, special education). However, the language in each field can be quite distinct. In child psychology and gifted education, the abbreviation FOI, for fidelity of implementation, is used (Century Rudnick, & Freeman, 2010; Missett & Foster, 2015). In behavioral health, fidelity monitoring and implementation monitoring are both used (Rosenheck, 2001). And in juvenile justice, treatment fidelity (Sprague et al. 2013) is a key term. Applied Behavior Analysis uses both TI (Fryling, Wallace, & Yassine, 2012) and procedural reliability (Gresham, 2009), while special education researchers may refer to intervention implementation (Johnson, Wehby, Symons, Moore, Maggin, & Sutherland 2014),

and general educators may use curriculum fidelity (Vartuli & Rohs, 2009). These few examples document not only the consensus across disciplines that it is important to implement treatments and interventions as designed, but also hint at the notion that TI is a multi-faceted concept.

However, the disparity of search terms causes significant difficulty when attempting to review the TI literature (DiGennaro Reed & Codding, 2014). It is possible that much relevant work is being ignored or simply cannot be found because, as with the research to practice gap, "people don't talk with one another" (DeAngelis, 2010, p. 42). Hence there are multiple terms for TI, and work in related fields that do not reference each other. Having such disparate language makes it difficult to talk to each other. Some key examples of this difficulty are apparent within the related fields of school psychology and behavioral education. In 2009, *School Psychology Review* dedicated a special issue to the topic of TI. In that one issue, *TI* is used in five articles (Greenwood, 2009; Gresham, 2009; McLeod, Southam-Gerow, & Weisz; Sanetti & Kratochwill, 2009; Schulte, Easton, & Parker, 2009): *fidelity measurement* in is used in one (Sheridan, Swanger-Gagne, Kwon, & Garbacz, 2009), *level of implementation/implementation* is used in two (McKenna, Rosenfield, & Gravois; Ransford, Greenberg, Domitrovich, Small, & Jacobson, 2009), and the terms *treatment fidelity*, *intervention integrity*, and *procedural reliability* are acknowledged in the introduction by Sanetti and Kratochwill (2009). A second special issue on the topic was published in 2014 by *The Journal of Behavioral Education.* While the editors acknowledge the difficulty of having a disparate language (DiGennaro Reed & Codding, 2014), in this one issue the term *procedural fidelity* is used in four articles (Barnett et al.; DiGennaro Reed & Codding; Pence, St. Peter, & Giles; Sanetti & Colliler-Meek, all 2014), *TI* is used in two (Gross, Duhon, & Doerksen, 2014; Leon, Wilder, Majdalany, Myers, & Saini, 2014) *treatment fidelity* in one (Suess et al., 2014),

and *implementation/treatment implementation* used in three (Noell, 2014; Noell et al., 2014; Reinke, Stormoht, Hevman, & Newcomer, 2014). The need for linguistic clarity about TI, as well as information about whether these disparate terms mean the same things or point to variances in meaning, while needed, are outside the scope of this paper. For the purpose of this paper, the term TI will be used. The following review of the literature examines how the field has parsed the concept of TI, how it is measured in the professional literature, the importance of TI, and the persistent under-reporting of TI in the literature.

Components of TI

Several teams of researchers have attempted to tease out the dimensions or components of TI. In a thorough review of TI in the childhood prevention literature, Dane and Schneider (1998) identified five dimensions of TI that should be examined. Dane and Schneider's five dimensions are: adherence, quality, exposure, participant responsiveness, and program differentiation. These five dimensions have served as an enduring theoretical basis with which to analyze TI (see for example Caroll et al., 2007; Nelson, Cordray, Hulleman, Darrow & Sommer, 2012; Pas & Bradshaw, 2012; Power et al., 2005). However, it does not appear that there are any studies that address all five dimensions in anything other than a theoretical context. Further, there is not any research to support these five dimensions individually as the critical pieces of TI. However, due to the enduring use of these dimensions, at least in theoretical contexts, it is important to understand them.

Adherence. Adherence is the extent to which a written procedure is followed exactly as intended. Adherence is often measured through checklists that count the number of intervention components that are completed accurately. Adherence is the easiest component to measure because it is so concrete. Subsequently, adherence is the most commonly measured dimension of

treatment integrity (Durlak & Dupree, 2008; Sanetti & Kratochwill, 2009; Fryling et al., 2012). It is important to note that when researchers report TI, they are almost always exclusively referring to adherence TI. There are two main types of adherence studies: studies that include feedback to teachers about their level of TI in an effort to improve adherence, and those that simply examine adherence on its own. Both types of studies will be described here and in later sections of the paper.

There are abundant examples of adherence TI monitoring in general education and behavioral science. Many of these do not include a feedback component. For instance, Stage et al. (2008) monitored teachers' implementation of behavior plans created using a consultation-with-functional behavioral assessment (FBA) model and a consultation-only model and measured teachers' adherence to the behavior plans. They found that high levels of adherence were required for behavior change strategies to be effective for students.

DiGennaro, Martens, and McIntyre (2005) also studied teacher adherence to behavior plans for students with behavior problems, and DiGennaro, Martens, and Kleinman (2007) had teachers implement a multi-step intervention to address off-task and attentional problems of four students with emotional behavioral disorders (EBD). In both cases, the researchers employed observers to watch teachers implement the intervention and counted how many of the prescribed steps were completed. Once again, adherence to the behavior plans (which they simply called "TI") was found to have a beneficial effect on student outcomes.

Arkoosh, Derby, Wacker, Berg, McLaughlin, and Barretto (2007) described an evaluation of long-term treatment integrity. The purpose of their study was twofold. First, they described the relationship between treatment integrity levels and treatment effectiveness. Second, they highlighted the importance of reporting the treatment integrity in outcome-based research.

Arkoosh et al. described adherence treatment integrity and measured it in the same way as the researchers above. They found that high levels of treatment integrity were required for overall treatment success. However, very low levels of integrity may be required for behavioral reduction procedures (i.e., extinction, aversive events) if high levels of reinforcement are provided. Arkoosh et al. posited that is possible that necessary levels of TI change depending on the intervention.

The effect of performance feedback on implementer TI was examined by Codding, Feinberg, Dunn, and Pace (2005) and Codding et al. (2008). This team also used direct observation to monitor the implementation of the behavior plans for five and eight students (respectively) with a variety of behavior difficulties and calculated the number of prescribed steps that were completed correctly during the observation period. Both of the teams headed by Codding included feedback to teachers as part of their research.

In the 2005 study, Codding et al. observed teachers implementing the behavior intervention plans of five students with acquired brain injuries and significant behavior problems. During each observation, teachers were rated on whether or not each component of a behavior plan was implemented in its entirety as intended, not implemented in its entirety as intended, or if there was no opportunity to observe the component. For any step that was not implemented correctly, or was missed, the researchers provided verbal feedback to the teachers in meetings that took place every other week. Codding et al. found that every other week feedback increased teachers' accuracy of intervention implementation. In this study, the effects of feedback were measured on TI. The teachers in the study reported that they felt the study had a beneficial effect on their students. However, Codding et al. did not measure how the different levels of TI affected student behavior. This occurs frequently in TI work – ways to increase TI

are measured, but whether the increased TI had the desired effect on student behavior is often left unexamined.

In the 2008 study, Codding et al. extended the results of the previous work on direct observation with performance feedback to examine changes to teacher behavior when observers were present versus when observers appeared to be absent. This was done in an attempt to control for any effect an observer being in the classroom might have on teachers' implementation of behavior plans. However, it seems unlikely that teachers were not aware that they were being observed. In this study, eight students with a variety of emotional and behavioral disorders were taught by a master's level teacher and two aides. The classroom was connected to an office that had a one-way mirror and audio speaker that allowed those in the office to hear what was going on in the classroom. Codding et al. rated treatment integrity based on the same three criteria used in the 2005 study. Performance feedback was provided within five to ten minutes of the time the teacher behavior took place. In one condition, the researchers were in the room and the teachers knew they were being observed. In the second condition, the researchers were in the office and the teachers had no idea if they were being observed at any given time (although they knew it was a possibility). Once again, performance feedback provided on adherence to the steps of a behavior intervention plan increased adherence to the intervention plan. Interestingly, teachers did not show a difference in implementation when they knew they were being observed versus being unaware of the observation. Again, student response to increased TI was not reported.

Watson, Ray, Sterling Turner, and Logan (1999) trained 2 teachers, and a teacher's aide to implement functional analysis to assess, and implement treatment for the self-injurious behavior of one student who had multiple and severe disabilities. The researchers measured the effects of direct observation and immediate performance feedback of both teacher-implemented functional

analysis and teacher-implemented intervention. As in the other adherence studies above, adherence to the steps of the procedure were measured. In this case, accurate implementation of 90% of steps in each functional analysis condition was considered sufficient. However, Watson et al. did not provide a basis for the 90% cutoff. As with all arbitrary levels of TI, it is impossible to know if a critical component might have been in the 10% that was allowed to be missed.

Jones, Wickstrom, and Friman (1997) also used direct observation with performance feedback to attempt to increase the TI of behavior plans. Jones, et al. used a multiple baseline across participants design. In this study, as in all of the studies of adherence TI with feedback to practitioners, the participants were the teachers. Three conditions were used: baseline, teacher consultation, and teacher consultation with performance feedback. Baseline consisted of operationally defining on-task behavior and reviewing the school-wide behavior plan with the teachers for a classic "business as usual" implementation of behavior plans with no intervention on the part of the researchers; teacher consultation, which consisted of the researchers reviewing the student on-task behavior from baseline with the teachers and developing a behavior plan. In all cases, the behavior plan included the teacher marking the student's data sheet and providing positive feedback at least every two minutes for on-task behavior. All "passive" off-task behavior (staring into space, non-participation in work) would be ignored. These techniques meshed with the Boys Town model used by the school. If low TI was observed in the teacher consultation phase (which it was in all cases), teachers entered phase three, performance feedback. This phase was similar to phase two, with the addition of daily feedback about both the student's behavior, and the teacher's behavior in providing praise and marking the data sheet at two minute intervals. It is not clear how many minutes or hours feedback was provided after the

observations. While TI improved slightly in the consultation alone phase, TI was vastly improved in the performance feedback phase. In the teacher consultation only phase, no teacher exceeded 37% TI in any session. During performance feedback, the teachers averaged 60%, 66% and 83% integrity. While there is still little to no consensus about what appropriate levels of TI might be, Jones, et al. were one of the few who also measured student behavior in an attempt to gauge the impact of differing levels of TI on student behavior. Student on-task behavior did increase in the performance feedback condition, but only slightly. However, with the relatively low levels of TI present in all conditions, it is difficult to draw any conclusions about the efficacy of the behavior plan or necessary levels of TI. Jones et al. mention the widespread use of the behavioral consultation model even though there is evidence that the traditional model without feedback does not increase TI.

Several years later, Hsieh, Wilder, and Abellon, (2011) followed much the same procedure as the researchers above. Hsieh et. al. were interested in teaching caregivers how to use incidental training at home with three young people with autism and other developmental disabilities. As in the other cases of adherence TI, a checklist was made of each step in the procedure, in this case incidental teaching of specific signs or pictures in order to receive the item. Each step completed correctly was counted and divided by the total number of steps in order to arrive at a percentage of correctly completed steps. Hsieh et al. found that incidental teaching could be taught quickly (under 30 minutes) and providing feedback to caregivers increased TI. Hsieh, et al., also measured student behavior and found that for two of the three students, increased TI did not necessarily lead to increases in using the skill taught through incidental teaching. This result only highlights the importance of tracking student behavior along with TI in order to ensure that only interventions that produce outcomes meaningful to students

are continued. Simply increasing TI is not sufficient if the intervention in question does not produce results that are meaningful to students.

In all of these examples, adherence was measured by observing and counting the number of intervention components implemented correctly. Additionally, adherence was the only measure of TI examined in each of these studies. No mention is made of any of Dane and Schneider's dimensions of TI, however in all of the studies presented here, it appears that adherence is the only dimension examined even though the NIH said "broader examination and measurement of the instructional context is strongly encouraged to document and inform our understanding of fidelity of implementation" (NIH, 2011, in Swanson, Wanzek, Haring, Ciullo, and McCulley, 2011).

Quality. Quality refers to the more elusive aspects of an intervention such as enthusiasm and preparedness of the person implementing the intervention. Dane and Schneider's (1998) inclusion of "leader attitudes toward the program" and "global estimates of session effectiveness" highlight the difficulty of measuring this variable. Perhaps due to its intangible nature, specifics on how to measure quality remain largely undefined. In 1998, Dane and Schneider found that only 26% of studies that measured integrity examined the dimension of quality of implementation. Measures of TI quality are still rare. However, there are a few examples. Swanson et al. (2011) found that in a pool of 76 special and general education academic intervention studies, only five (6%) reported a measure of quality TI. Interestingly, while the researchers reported some measure of TI that seemed to fit into Dane and Schneider's definition of quality, none of the studies reviewed by the Swanson team explicitly stated that they were measuring the quality dimension of TI. Swanson et al. gave the examples of measuring the use of clear language and using concrete examples when modeling a strategy as a measure of

quality of TI. In cases such as these, researchers may embed language about quality of implementation in adherence checklists. This may add to the difficulty in finding information about quality TI, given that in Swanson et al.'s recent review, none of the studies provided the actual TI checklists that were used. Thus there are two main barriers to finding research regarding quality measures of TI. First, researchers are not explicitly using the term *quality,* which impedes searchabilty in databases. Second, quality information that might be embedded in adherence checklists may be overlooked if the checklists are not provided as there is no way for readers to know if quality measures were included.

In an explicit study of quality TI, Power et al. (2005) found that in a school-based prevention model, *adherence* to the intervention was being monitored and focused upon to the determent of the *quality* of intervention implementation. The researchers repeatedly found low adherence TI. Through a series of meetings with school personnel, the researchers found that their strict insistence on adherence to the teaching intervention led to teachers implementing the intervention with low quality. The teachers felt that the lessons were boring to the students and implemented them with low adherence so that they could change the components to be more interesting to students. Unfortunately, the exact nature of the changes the teachers made were not provided. Additionally, the teachers reported that they implemented the intervention with low quality as well. The teachers reported feeling that they had no ability to use the program in a way that made sense in terms of their relationships with students. In turn, this dissatisfaction led to a lack of enthusiasm for the program and thus low quality of implementation. Power et al. posit that a partnership model in which the leadership and implementers work together, rather than a hierarchical model in which the leadership simply tells the implementers what to do, may help to

improve the quality of intervention implementation. This idea has good face validity and merits further investigation.

Helmond, Overbeek and Brugman (2012) were the first in their field (and the only to date) who attempted to measure adherence, exposure, participant responsiveness *and* quality of an intervention for incarcerated youth. Although they cite Dane and Schneider (1998) specifically, there is no mention of why Dane and Schneider's fifth criteria, program differentiation, was not included as a dimension of TI. The researchers evaluated TI of the EQUIP program for incarcerated youth using direct observation and checklists. To measure quality, items such as, "The trainer encourages participation…" (p. 1723) were rated. In their study, they created a composite TI score which included all measured dimensions of TI and found the average quality score to be 61%, with an overall composite TI score to be at 55%. Even with low levels of integrity, students made some gains using the EQUIP program.

In the special issue of the *Journal of Behavioral Education*, DiGennaro Reed and Codding (2014) address the significant overlap that can occur between adherence and quality measures of TI, which they refer to as procedural fidelity. They note that measures of adherence often encompass *content* by counting the number of steps that of an intervention that have been implemented as well as *quality* by measuring how well each step was implemented (see Swanson et al., 2011). Both content and quality will then be reported as measures of adherence TI which serves to highlight difficulty of the study of quality TI.

Exposure. Exposure, sometimes called dosage (Jones, Clarke, & Power, 2008), refers to the duration or number of the prescribed opportunities that the student has to engage in the treatment. Many researchers report student absence or missed sessions of an intervention, but do not do so within the framework of TI. A careful reader can read and understand that student

absence or missed sessions may have an effect on intervention effectiveness. However, examining exposure and dosage as explicit components of TI allows for a more nuanced understanding of this dimension of TI. Power et al. (2005) explicitly measured exposure TI in a study that examined a hierarchical versus partnership models of measuring TI. Power et al. calculated the percentage of times students received the intervention compared to the number of times students were supposed to receive the intervention. They found low exposure TI, which they attributed to the teacher dissatisfaction with the program mentioned earlier.

Helmond et al., (2012) measured exposure by assessing how often the intervention took place, how often sessions were cancelled, and how long interventions sessions lasted. Students were exposed to 66% of the expected dosage of the program. Missed sessions, cancelled sessions, and shorter than required sessions all played a part in the low exposure. However, as stated earlier, students did make some progress even with low TI.

Participant Responsiveness. Participant responsiveness measures the extent to which students respond to the intervention in the desired manner, either through behavior change or skill acquisition. Researchers measure participant responsiveness by comparing baseline levels of the behavior or skill in question to levels during the intervention. Surprisingly, participant responsiveness, while seemingly crucial, is not often reported in the professional literature *as it relates to TI* (Wickstrom, Jones, LaFleur & Witt, 1998). Even in studies where TI is the main focus, how participants respond to the intervention may not be mentioned at all. For example, both Codding et al. (2005) and Codding et al. (2008) completely ignore student response to TI. More recently, McKinney and Vasquez III (2014) investigated the use of Bug In Ear (BIE) feedback to improve the fidelity of Discrete Trial Training (DTT) provided by teachers to students with autism. The study is detailed and describes the BIE intervention well and the effect

it had on TI of DTT instruction. However, it is not reported if the TI of BIE intervention had any impact on student learning. Nelson et al. (2012) put forth a five step model of TI monitoring which includes "linking fidelity measures to outcome measures" as the fifth, and final, step in the process indicating that is an essential part of TI.

Program Differentiation. Dane and Schneider's (1998) fifth and final component is program differentiation, indicating both the difference between the current intervention and others, as well as the clarity of implementation (removal of unimportant steps and steps that are contra-indicated). Dane and Schneider found that only 6% of their total sample of studies in the review included a measure of program differentiation. In Dane and Schneider's study, they do not provide any further information about the 6% of studies that included program differentiation, making it difficult to make inferences about what types of studies measure program differentiation and how this is done. The argument for including program differentiation is that it allows researchers to determine which elements of an intervention are essential to its implementation. Carroll et al. (2007) argue that a better name for this dimension of TI is "component analysis" and that this dimension, while important, is slightly different than TI. Other than theoretical articles describing why this dimension is important, no other mentions of program differentiation were found.

Although Dane and Schneider's framework has persisted and is often cited by those doing TI research (DiGennaro Reed & Codding, 2014), there is little empirical support for these being the only and/or sum total pieces of TI. Much of the empirical research on TI has checklists that are intervention-specific, which makes their utility limited across a variety of contexts and interventions (Schoenwald, Henggeler, Brondino, & Rowland, 2000). For example, in 2009, Sanetti and Kratochwill updated and attempted to gain consensus about the various components

of TI that had been posited. The authors reviewed relevant work in education, clinical psychology, and prevention science in an attempt to find a clear definition of TI. Citing the "complexity and multi-dimensional nature" (p. 446) of TI, Sanetti and Kratochwill created a series of Venn diagrams to highlight the overlap in the works they reviewed. They found that even though many TI researchers say that in order to comprehensively assess TI all of Dane and Schneider's (1998) components should be measured, there is little actual empirical support for all of Dane and Schneider's dimensions. However, there was significant mention of four components of TI in the majority of the works studied by Sanetti and Kratochwill. They called these: content, quality, quantity, and process. Content is the broad term used to refer to the steps of an intervention that are provided. Content seems to closely mirror Dane and Schneider's *adherence* dimension in that it is used to measure the treatment steps completed. Quality measures how well the intervention is provided. Quantity is the general term used to encompass how much of a given intervention was provided. This can include how much time is spent with an intervention, how many sessions are provided of an intervention, or any other relevant measure of the amount of intervention provided. Finally, process refers to the manner in which the intervention is provided. Sanetti & Kratochwill do not define process further. The broad strokes of Dane and Schneider's work can be seen in these four components. However, Sanetti and Kratochwill recognized that empirical support is needed for these larger conceptual ideas to be validated. To that end, they propose the following definition of TI: "…the extent to which essential intervention components are delivered in a comprehensive and consistent manner by an interventionist trained to deliver the intervention." (p. 448).

Measuring TI

Three of the most common ways to measure TI are: (1) self-report, (2) permanent-product data reveiw, and (3) direct observation with performance feedback. Self-report data can be collected with rating scales, interviews, and surveys completed by those implementing the intervention (Gresham. MacMillan, Beebe-Frankenberger, & Bocian, 2000). Self-report data collection is easy to implement, does not require a second person to record data, and is cost effective. Completing rating scales and surveys that contain information about the intervention might also remind teachers of the correct steps and therefore increase TI (Gresham et al., 2000). However, self-report also may result in inflated percentages of TI (Gresham et al., 2000). Wickstrom et al. (1998) found that teachers reported that they were using the intervention between 54% and 67% of the time; however direct observation by the research team found only 4% intervention use.

Permanent-product data review may be more accurate than self-report measures (Noell, Witt, Gibertson, Ranier, & Freeland, 1997; Wilkinson, 2007). Teachers or researchers collect permanent-product TI data by reviewing student work samples, point cards, or behavior tracking sheets to evaluate the implementation of interventions. A benefit of permanent-product data collection is that it does not require a second person to observe or collect data (Lane, Bocian, MacMillan, & Gresham, 2004). Permanent-products may not be appropriate in all situations because not all teaching or behavioral interventions lend themselves to permanent-product recording. For example, some interventions require the teacher to provide verbal instructions or edible reinforcement. These interventions do not leave a permanent product to examine.

A third method for monitoring TI is direct observation with performance feedback (Codding, et al., 2005, Codding et al., 2008). Direct observation with performance feedback is

often used by consultants or researchers in schools. First, the components of an intervention are decided upon and defined; next, the occurrence / nonoccurrence of each component is recorded; last, a percentage is calculated by dividing the total number of occurrence by the total opportunities (Gresham et al., 2000). Using direct observation with performance feedback, Jones et al. (1997) were able to increase TI from a range of 9%-37% to a range of 60%-83%. Witt, Noell, LaFleur, and Mortentson (1997) found that performance feedback was a key component for increasing TI. Including data about both teacher and student performance during performance feedback is also beneficial to increasing TI (Noell et al. 1997). In fact, providing information about how well the teacher implements the intervention may be more powerful than telling the teacher how the student is responding to the intervention (DiGennaro Reed, et al., 2007). Additionally, most teachers do not find observation or feedback procedures problematic or intrusive (Codding et al., 2005; DiGennaro Reed et al., 2007).

Direct observation with performance feedback has drawbacks. In addition to being costly, due to the additional time and personnel resources required, there is the distinct possibility that the teacher may only implement the intervention while the observer is in the room due to the Hawthorne, or observer, effect. Observer effects may lead to labeling an intervention as ineffective without accurate information due to the observer believing that an intervention is being done with good TI (because they see good TI when they observe), when in truth the intervention is not being implemented consistently when the observer is not in the room. This inconsistency in implementation makes is impossible to determine if an intervention is effective overall. However, that concern may be mitigated by the results of the Codding et al. (2008) study mentioned earlier in which teachers were observed without their knowledge with the use of a two-way mirror and found similar levels of TI as during scheduled observations of which the

teachers were aware. Unfortunately, there is some evidence that TI declines once monitoring and feedback stop (Sutherland, Wehby, & Copeland, 2000). Additionally, TI checks do not typically take place over long periods leading to a lack of maintenance data regarding TI. After reasonable levels of TI are achieved, the consultant typically stops coming to the classroom. However, thinning observations to as little as once every two weeks may be enough to help maintain integrity (DiGennaro et al., 2007).

Importance of TI

Power et al. (2005) and Zvoch (2009) are clear that TI should be used to evaluate not only the implementers of a program (usually teachers), but also the program itself. Using TI data to evaluate a program can lead to the removal of superfluous components. For example, if TI of a behavior intervention plan (BIP) is evaluated, and the data indicate that reinforcement (e.g. stickers) is not offered as prescribed in the BIP, there are two possible routes of action. First, if the student's behavior is not improving, the TI data may be used to remind the teacher of the importance to implementing all BIP components. Second, if the student's behavior is at acceptable levels even without the reinforcement procedure, this component can be assumed to be unnecessary. The reinforcement (sticker component) can then be removed or thinned, streamlining the teacher's workload as well as moving the student toward a less-intensive, less socially-stigmatizing BIP. Using TI data in this way allows for accurate decisions to be made regarding both large (i.e., placement) and small (i.e., day-to-day routine) matters that affect students.

Even when adherence (Dane & Schneider, 1998; Power et al., 2005), is the only aspect of TI being measured, there are clear implications for student outcomes. In many cases, the higher the level of TI, the better the student responds. In an early study, Holcombe, Wolery, and Snyder

(1994) manipulated the TI of constant time delay (CTD) used to teach an academic skill to students with intellectual disabilities. Students learned the skill when CTD was used with both 100% and 50% accuracy. They learned little to nothing when CTD was used with 0% accuracy. However, students learned far more when CTD was used with 100% accuracy. Therefore, not only is the effectiveness of an intervention affected by TI, but the efficiency may be affected as well. Noell, Gresham, and Gansle (2002) found that their students learned both faster and more efficiently when TI was high. Grow et al. (2009) manipulated TI of the well-validated system-of-least-prompts (SLP) to make it resemble the multiple verbal prompts that teachers actually use. Students in that study also learned much more efficiently when SLP was implemented with high TI.

In addition to those mentioned earlier, other researchers have found functional relations between TI and student performance. Sanetti and Kratochwill (2009) reported that students made small, but functionally important gains in math with 100% TI producing the best outcomes. For two participants, the 50% and 0% TI conditions were indistinguishable and produced very poor outcomes, suggesting that even some errors may prove disastrous for some students.

Under-reporting of TI

The variety of fields that study TI as a concept indicates that TI is widely thought to be an important concept. However, numerous reviews across a variety of fields have found that researchers rarely report TI. Dane and Schneider (1998) found a dearth of researchers including TI in the primary and secondary prevention literature. The same under-reporting of TI is found throughout the education and behavior change literature. Over the years many literature reviews have been undertaken to assess the rates at which TI is conducted in various publications or specialties. Generally speaking, TI has been underreported. For example, between the years

1975 and 2000, authors of only 27% of studies involving students with EBD reported any measure of treatment integrity (Smith & Daunic, 2007). Gresham et al.(2000) found that nearly half of the studies in their review of interventions in learning disabilities described TI in some way. However, only18.5% measured TI in a "rigorous" way.

Peterson, Horner, & Wonderlich (1982), found that between 1968 and 1980 only 16% of articles published in the *Journal of Applied Behavior Analysis* (JABA) reported any measure of TI. Gresham et al. (1993) picked up and completed a similar examination of the publications in JABA from 1980-1990 and found the same 16% figure. Wheeler, Baggett, Fox, and Blevins(2006) expanded the search to include other peer-reviewed journals that addressed behavioral interventions for children with autism during the period from 1993-2003 and found strikingly similar results. Only 18% addressed TI, while 3% discussed TI as a concept but provided no data on the topic. McIntyre, Gresham, DiGennaro, and Reed (2007) looked at only school-based studies in JABA and found 30% reported TI data from 1991-2005. Swanson et al. (2011) show that reporting of TI may be increasing with their finding that 67% of the research studies in their sample reported some measure of TI. However it is clear that TI is still not fully reported. Quality measures and student outcomes are too often neglected in favor of simply reporting ways to increase TI.

Researchers outside of education have made advances in the measurement and understanding of TI which can and should be used to inform educational research. The majority of this TI work has been conducted by researchers in psychotherapy, youth violence prevention, and substance abuse treatment (Schulte et al., 2013). Some of the ideas from those outside of education have been incorporated into the development of new tools such as concept-mapping (Green, Fettes, & Aarons, 2012) and measurement systems (Nelson et al., 2012) that can be

applied to both behavioral health and educational research. Many of these innovations can be tied to federally funded studies specifically focused on TI (Shulte et al., 2013).

The new focus on TI research has spurred work applicable to large-scale, group-design studies (Helmond et al., 2012; Nelson et al., 2012; Pas & Bradshaw, 2012). However, the mandate for the inclusion of TI is not limited to group-design research. In the arena of single-case research, Horner, Carr, Halle, McGee, Odom et al. (2005) indicated that measuring TI was "highly desirable" (p. 174) not only as a method to ensure accuracy of intervention delivery but also as a component of social validity when translating research findings into practice by educators and caregivers. Additionally, Kratochwill et al. (2010) listed TI as one of the components that must be included in quality single-case design research in their document for the *What Works Clearinghouse*. Horner et al., in particular, are clear that without measuring TI in the intervention phases of research, it is not possible to draw conclusions about the effect of the independent variable on the dependent variable.

Identifying a functional relation between a dependent and an independent variable is one of the underpinnings of applied behavior analysis (ABA; Fryling et al., 2012). This may be part of the reason that single-case research has a relatively better track record of reporting TI (Swanson et al., 2011). The information gained from the identification of a functional relation is what drives interventions in ABA (Alberto & Troutman, 2017; Kazdin, 2011). Surprisingly, TI (i.e., the monitoring of the independent variable) has not always received the same careful scrutiny that is afforded to the dependent variable (e.g., documenting inter-observer agreement; McIntyre et al., 2007; Wheeler et al., 2009). This may be changing with the Behavior Analysis Certification Board's most recent set of standards stating that behavior analysts have an obligation to measure TI to ensure accurate delivery of services as well as to ensure that

interventions are modified as needed (BACB, 2012). With the recent push to develop quality indicators for single-case research, TI has come to the fore. "Fidelity of implementation is a significant concern within single-subject research because the independent variable is applied over time" (Horner, et al., 2005, p. 168). The What Works Clearinghouse panel headed by Kratochwill (2010) also calls for TI to be designed into studies. As under-reported as TI is, it is unsurprising that there are many nuances left to be explored. The current study attempts to understand some of the nuances.

Acceptable Levels of TI

Researchers have demonstrated that TI is directly related to student outcomes, and so is an important construct to consider when designing and delivering instruction. However, two issues must be considered when evaluating components of TI. The first is the fact that no intervention outside of a clinic will be implemented with 100% TI, 100% of the time (Houchins, Jolivette, Shippen, & Lambert, 2010). Therefore, practitioners must examine the level at which TI promotes positive student outcomes. The second issue is the emerging idea that some level of deviation from prescribed procedures could be beneficial.

Houchins, et al. (2010) acknowledge the complexity of gathering meaningful TI data in non-clinical settings. Clinical research takes place in a unique context in which researchers attempt to hold as many variables as possible constant. However, researchers cannot hold all variables constant. Even if they could, those variables may often revert to the previous (non-constant) conditions once the researchers leave. If the goal is to collect TI data that is meaningful in an applied setting, context must be considered if researchers are to draw any conclusions about the TI of an intervention once they leave the setting. Unfortunately, there is no set level of appropriate or acceptable TI, although 80% adherence is often given as a benchmark (Noell, et

al. 2002; Smith et al., 2007). Subsequently, what level of TI to target vacillates without guidance on how to determine acceptable levels. Gresham (2005) offers the commonsense advice that TI is high enough when the student experiences meaningful change.

There are several factors to consider when determining the necessary level of TI. There are likely many variables at work that affect levels of TI in different situations. For example, the severity of the behavior, idiosyncratic student variables, order effects and type of TI errors committed may all play a role.

Arkoosh et al. (2007) found that in order to reduce inappropriate behaviors to a level that was meaningful for students, more severe behaviors required higher levels of TI than less severe behaviors. Even small deviations from the prescribed interventions impeded behavior change. Likewise, Stage, et al. (2008) found that high TI was required for behavior change strategies.

Idiosyncratic student variables may affect acceptable levels of TI. DiGennaro Reed, Reed, Baez, and Maguire (2011) systematically manipulated the TI of program implementation for students with autism. One hundred percent accuracy produced the best outcomes, while for most participants 50% TI produced middling results. In this case, Gresham's (2005) idea of TI being high enough when it produces meaningful change would indicate that 50% integrity was not high enough for the participants in DiGennaro Reed et al. However, for two participants, 50% and 0% integrity were related to indistinguishably poor outcomes. So, it is possible that there are individual student factors affecting necessary levels of TI, although there is no evidence as to what these factors may be. Other researchers have also found mixed results across participants. Odluyrt, Tekin-Iftar, and Adaliouglu (2012) found that lower TI was associated with lower skill acquisition efficiency for the majority, but not all of their participants. The differences between the participants who required high TI versus low TI are unclear.

Groskreutz, Groskreutz, and Higbee (2011) suggested that different types of errors in implementation may differentially affect student outcomes. They classified errors as those of either commission or omission. Errors of commission involve reinforcement of inappropriate responses or reinforcement at inappropriate times. In contrast, with errors of omission researchers intentionally failed to reinforce appropriate responses (Groskreutz et al., 2011; Grow et al., 2009; Odluyrt et al., 2012). As commission and omission errors are the focus of the dissertation, detailed examples of the few works that focus on this area are provided here.

Errors of Omission and Commission

Vollmer, Roane, Ringdahl, and Marcus (1999) evaluated the effects of errors of omission (neglecting to reinforce when appropriate) and errors of commission (reinforcing when inappropriate) on a direct reinforcement of alternative behavior (DRA) procedure. It is important to note that Vollmer et al. did not use the terms omission and commission. However, a close reading of their work suggests that those were the variables examined. Further, as noted earlier, the terminology used is distinct. They refer to "treatment challenges" rather than TI. This linguistic inconsistency once again highlights the difficulty of comprehensively assessing work in this field.

DRA is a common behavioral procedure wherein a desired behavior is reinforced on a set schedule, while the undesirable behavior is placed on extinction (Vollmer & Iwata, 1992). The participants were two males and one female who had severe to profound intellectual disability and were 4-, 16-, and 17-years-old respectively. All three participants engaged in self-injurious behavior (SIB). The 17-year-old female also engaged in aggression. All were evaluated in their schools. Vollmer et al. (1999) began by running standard functional analysis procedures based on the work of Iwata, Dorsey, Silfer, Bauman, & Richman (1982/1994). Functional analyses

were conducted in order to determine the conditions under which the undesired behaviors (SIB, aggression) were most likely to be elicited. The results from the FA for each student were then used during the baseline condition of the subsequent study.

For example, Rachel's aberrant behavior was maintained by escape from towel folding and utensil sorting tasks. During the baseline sessions, Rachel was presented with these tasks and any appropriate behavior was never reinforced and all inappropriate behavior was reinforced at 100% levels. Vollmer et al. (1999) refer to this as the 0/100 contingency. During full implementation of the DRA procedure, the reverse was true. Rachel's inappropriate behavior was never reinforced and appropriate behavior (30 seconds of compliance) was reinforced with a break 100% of the time. This contingency was called 100/0. The last contingency tested was called partial implementation. In partial implementation, DRA was implemented with different levels of TI. The authors' purpose was to gauge the effects of differing kinds and amounts of TI errors on the students' behavior, after effective treatment had been implemented previously.

In the Vollmer et al. (1999) study, implementing reinforcement incorrectly represented an error of omission (neglecting to reinforce appropriate behavior), while an error in implementing extinction was an error of commission (reinforcing inappropriate behavior). Vollmer et al. (1999) did not always separate errors of omission from errors of commission. For instance, while in baseline conditions, no instances of inappropriate behavior were reinforced (0% errors of omission), but all inappropriate behaviors were reinforced (100% errors of commission). But during other conditions 25/75 (25% correct reinforcement for appropriate behavior- errors of omission, 75% reinforcement of inappropriate behavior –commission errors), 50/50, 75/25, and 100/100 were used. This means that in the same session, errors of omission and commission both occurred.

As would be expected, when differential reinforcement was implemented correctly with no errors of commission or omission, appropriate behavior nearly entirely replaced inappropriate behavior. As TI errors increased, inappropriate behaviors also increased, although there was still a tendency toward appropriate behavior. Across the three participants, they found that although inappropriate behaviors re-emerged when errors of commission were made (inappropriate behaviors were reinforced), appropriate behaviors persisted even when errors of omission were made. It is important to keep in mind that Vollmer et al. (1999) did not isolate errors of omission from errors of commission, thus, their study provided information about the effects of combined errors, but not of one type of error versus the other.

St. Peter Pipkin, Vollmer, and Sloman (2010) also studied errors of omission, as well as errors of commission. Their work included two studies. The first study had three components: a large-scale clinical study of combined errors of omission and commission, errors of omission only, and errors of commission only with college students as the participants. The participants were instructed to click on either black or red circles on a computer screen in order to earn points. When exposed to omission errors only, the students did not engage in high rates of inappropriate behavior (clicking black circles), but their rates of appropriate behavior (clicking red circles) did decrease as errors of omission increased. For students who were exposed to errors of commission only, TI had to be at 40% or lower before it had a negative effect on student performance. Students in the mixed group were exposed to both type of errors such that 40% integrity meant that there was a 40% chance that students would receive a point for appropriate behavior and a 60% chance that that they would receive a point for inappropriate behavior. At 20% and 40% students engaged in high rates of inappropriate behavior and low

rates of appropriate behavior. However, at 60% and above, the students exhibited the opposite pattern of behavior and responded appropriately more than inappropriately.

This large-scale clinical study indicated that errors of commission had a more detrimental effect than errors of omission, but only when TI was at very low levels (40% or lower). The trials which included combined errors (similar to Vollmer et al., 1999) had similar results. As with Vollmer et al. (1999), St. Peter Pipkin et al. (2010) found that initially implementing DRA with 100% integrity may provide some protection against later TI errors.

In the second experiment, St. Peter Pipkin et al. (2010) examined the completion of math problems using DRA contingencies. The participant was Helen, a fourth-grade girl with autism. The researchers examined combined errors of omission and commission in a DRA procedure to increase math work completion. Helen responded similarly to the college students in the combined errors condition. She displayed more appropriate behavior than inappropriate behavior when the overall level of TI was 60% or above. At levels below 60%, her pattern of responding reversed, and inappropriate behavior became dominant.

A third and final experiment by St. Peter Pipkin, et al. (2010) was designed to evaluate sequence effects of the two types of integrity failures. Jake was an adolescent described as "trainable mentally handicapped" and was evaluated in his school. After functional analysis, it was determined that Jake engaged in aggression in order to receive attention. DRA was implemented, with the alternative to aggression being Jake engaging in spoken language (saying hi). Baseline, 100% and 50% integrity conditions were evaluated. During the 50% integrity condition, 50% of inappropriate behaviors and 50% of appropriate behaviors resulted in attention from the therapist. When Jake was exposed to 50% integrity following baseline conditions, aggression remained high and appropriate greetings remained low. However, when the 50%

condition followed 100% DRA, aggression was reduced to low levels and appropriate greetings increased to moderate levels. This follows the results from Vollmer et al. (1999) and Experiment 1 from St. Peter Pipkin et al. that initial high integrity may protect against later TI failures.

DiGennaro Reed et al. (2011) extended the study of errors of commission to academic instruction using DTT. They conducted a parametric analysis of errors of commission during discrete-trial teaching. They used DTT to teach receptive non-sense shapes to three, 8-year-old boys with autism. DiGennaro Reed et al. implemented DTT with 100%, 50%, and 0% levels of TI. The only errors committed were errors of commission whereby they reinforced incorrect shape identification at pre-programed levels. They found that 100% integrity produced the best skill acquisition and that 50% integrity and 0% were equally detrimental to skill acquisition for two of their three participants. The third participant's acquisition level matched the TI level. For this student, in the 50% contingency he averaged 45% accuracy, and in the 0% TI condition he averaged 10% accuracy. DiGennaro Reed et al. used a multiple baseline across participants design and did not examine the possible effects of the order of the levels of TI. To date, Vollmer et al. (1999), St. Peter Pipkin, et al. (2010) and DiGennaro Reed et al. (2011) are the only three studies which have systematically manipulated errors of commission; therefore, implications for practice are unclear.

Conclusion

The importance of documenting TI in education is clear. Both the reauthorization of NCLB and IDEIA mandate that professionals must use evidence-based practices to make educational decisions (Smith, et al. 2007). Educators cannot make informed decisions unless it is clear that those evidence-based practices have been implemented accurately; therefore, TI is an essential component in the application of interventions in both clinical and applied settings.

Measuring the fidelity with which an intervention is implemented allows for a more thorough and accurate understanding of which components of an intervention are effective, necessary, and feasible. Failing to examine TI can lead to what Vollmer, Sloman, and St Peter Pipkin (2008) called false positives and false negatives. False positives occur when a practitioner or a clinician erroneously assumes that improvements in student behavior are based upon an intervention, when in fact that intervention has been provided with low TI. False negatives occur when a treatment is abandoned due to poor student outcomes, even though the treatment was not implemented with TI at a high enough level to determine its efficacy. A broader understanding of the concept of TI, as well as better methods for measuring all dimensions of TI, will lead to more accurate understanding of which treatments are beneficial for students.

Overall reporting of TI in the educational and behavioral literature is low (Smith et al., 2007). However, even in studies in which TI is explicitly examined, the essential component of how TI may have affected student performance is not reported consistently. Gresham's (2005) idea that TI is sufficient when students receive benefit is not adequate to determine what acceptable levels of TI are. However, unless student performance is reported as a function of TI, it is impossible to know what levels of TI are required. Given that it is unreasonable to expect 100% TI in applied settings, appropriately training teachers requires a thorough understanding of the types of TI errors that may be committed, as well as the levels of TI necessary for student success when errors inevitably occur. It is essential that TI be fully understood because of the real-world implications for students, including movement into special education and special education placement changes which are based upon students' responses to interventions. It is not possible to make educational decisions on intervention effectiveness unless it is clear that

evidence-based practices have been implemented accurately. If TI is low, there is no way to determine which services and interventions a student requires.

Future research must include empirical testing of the theoretical constructs used to describe TI. This will bring much needed clarity to what is clearly an important concept. Operationally defining the components of TI, as well defining whole construct as a whole, is essential to the growth and cohesiveness of the field. In addition to allowing for more precision in determining which factors comprise TI and which factors need to be measured to adequately assess and report TI, researchers must agree on a common language to use when describing it. This will facilitate the sharing of ideas within fields as well as across related fields.

Currently, most TI research focuses on adherence monitoring (Durlak & Dupree, 2008). Adherence is the easiest component to measure, but there is no evidence to show adherence monitoring is sufficient measure of TI (Swanson et al., 2011). Quality, exposure, participant responsiveness, and program differentiation all require a substantial amount of research to determine if they are required components of TI.

In addition to determining which components of TI should be measured, more research is required into how to measure TI. Direct observation may be more accurate than self-report, and appropriate across more settings than permanent product measures of TI (Gresham et al., 2000; Wickstrom et al., 1998; Wilkinson et al., 2007). However, direct observation is costlier because it requires a second observer to be present to collect and then share the data. The accuracy of direct observation may also be impeded by the observer effect although that is unclear (Codding et al, 2008). There is some evidence that booster sessions may be sufficient to keep TI from eroding over time. Further research about acceptable levels of TI is also necessary. Currently,

most research uses the 80% benchmark while offering no basis for that number (Arkoosh et al., 2007; Wilder et al., 2007)

Researchers have begun to study the types of errors made during treatment implementation. Errors of commission and omission have received initial study. This area of study is ripe for further research. While some initial research indicates that errors of omission may be more less detrimental to student success than errors of commission (St. Peter Pipkin et al., 2010, Vollmer et al., 1999), more data are needed before making sweeping statements about the two types of errors. Research in this area is in its infancy and more data will provide valuable information about which errors are negligible and which errors are most important to avoid in order to ensure student success.

REFERENCES

Alberto P.A.& Troutman, A. C. (2017). *Applied behavior analysis for teachers* (Interactive 9th ed.). Boston, MA Pearson.

Alexander, J.L., Ayers, K.M, & Smith, K. A. (2015). Training teachers in evidence-based practice for individuals with autism spectrum disorder: A review of the literature. *Teacher Education and Special Education, 38,* 13 –27.

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

Barnes., A. C. & Harlacher, J. E. (2008). Clearing the confusion: Response-to-intervention as a set of principles. *Education and treatment of Children, 31,* 417-431.

Barnett, D., Hawkins, R., McCoy, D., Wahl, E., Shier, A., Denune, H., & Kimener, L. (2014). Methods used to document procedural fidelity in school-based intervention research. *Journal of Behavioral Education, 23,*898-107.

Behavior Analyst Certification Board. (2012). *Guidelines: Health Plan Coverage of Applied Behavior Analysis Services for Autism Spectrum Disorders*. Retrieved from http://www.bacb.com/Downloadfiles/ABA_Guidelines_for_ASD.pdf.

Boardman, A. G., Argüelles, M. E., & Vaughn, S. (2012). Special education teachers' views of research-based practices. *The Journal of Special Education, 39,* 168-180.

Borders, C. M, Bock, S. J., & Szymanski, C. (2015). Teacher ratings of evidence-based practices from the field of autism. *Deaf Studies and Deaf Education, 20,*91-100.

Browder, D. M. & Cooper-Duffy, K. (2003). Evidence-based practices for students with severe disabilities and the requirement for accountability in "No Child Left Behind". *The Journal of Special Education, 37,* 157-163.

Burns, M. K. & Ysseldyke, J. E. (2009). Reported prevalence of evidence-based instructional practices in special education. *The Journal of Special Education, 43,* 3-11.

Carnine, D. (1997). Bridging the research-to-practice gap. *Exceptional Children, 63,* 513-521.

Caroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2,* page numbers not available. doi: 10.1186/1748-5908-2-40

Century, J. Rudnick, M. & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation,* 3, 199-218.

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,* 205-219.

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

Cook, B. G., Tankersly, L., Cook, M., & Landrum, T. J. (2008). Evidence-based practices in special education: Some practical considerations. *Intervention in School and Clinic, 44,* 69-75.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18,* 423-445.

DeAngelis, T. (2010). Closing the gap between practice and research. *Monitor on Psychology, 41*, 42.

DiGennaro Reed, F. D., Reed, D. D., Baez, C. N., & Maguire, H. (2011). A parametric analysis of errors of commission during discrete-trial training. *Journal of Applied Behavior Analysis, 44*, 611-615.

DiGennaro Reed, F.D. & Codding, R.S. (2014). Advancements in procedural fidelity assessment and intervention: Introduction to the special issue. *Journal of Behavioral Education, 23,*1-18.

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*, 447-461.

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*, 220-231.

Doll, J. J., Eslami, Z., & Walters, L. (2013). Understanding why students drop out of high school, according to their own reports. *SAGE Open*: 1-15

Durlak, J. A, Dupree, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*:327–350.

Erchul. W. (2011). School consultation and response to intervention: A tale of two literatures. *Journal of Educational and Psychological Counseling, 21,* 191-208.

Every Student Succeeds Act of 2015, Pub. L. No. 114-95 § 114 Stat. 1177 (2015-2016).

Fryling, M.J., Wallace, M. D., & Yassine, J. N. (2012). Impact of treatment integrity on intervention effectiveness. *Journal of Applied Behavior Analysis, 45,* 449-453.

Fuchs, D. & Fuchs, L. S. (2005). Responsiveness-to-intervention: A blueprint for practitioners, policymakers, and parents. *Exceptional Children, 38,* 57-61.

Glennie, E., Bonneau, K., Vandellen, M., & Dodge, K.A. (2012). Addition by subtraction: The relation between dropout rates and school-level academic achievement. *Teachers College Record, 114,* 1-26.

Green, A. E., Fettes, D. L., & Aarons, G. A. (2012). A concept mapping approach to guide and understand dissemination and implementation. *Journal of Behavioral Health Services & Research, 39,* 362-373.

Greenwood, C. R. (2009). Treatment integrity: Revisiting some big ideas. *School Psychology Review, 38,* 547-553.

Gresham, F. M. (2005). Treatment integrity and therapeutic change: Commentary on Perepletchikova and Kazdin. *Clinical Psychology: Science and Practice, 12,* 391-394.

Gresham, F. M. (2009). Evolution of the treatment integrity concept: Current status and future directions. *School Psychology Review, 38,* 533-540.

Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenlum, S. (1993). Treatment integrity of school-based behavior intervention studies: 1980-1990. *School Psychology, Review, 22,* 254-272.

Gresham, F. M., MacMillan, D. L., 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 & Practice, 15,* 198-205.

Groskreutz, N.C., Groskreutz, M.P., & Higbee, T.S. (2011). Effects of varied levels of treatment integrity on appropriate toy manipulation of children with autism. *Research in Autism Spectrum Disorders, 5,* 1358-1369.

Gross, T., Duhon, G., & Doerksen-Klopp, B. (2014). Enhancing treatment integrity maintenance through fading with indiscriminable contingencies. *Journal of Behavioral Education, 23*,108-131.

Grow, L. L., Carr, J. E., Gunby, K. V., Charania, S. M., Gonsalves, L., Ktaech, I. A., & Kisamore, A.N. (2009). Deviations from prescribed prompting procedures: Implications for treatment integrity. *Journal of Behavioral Education, 18,* 142-156.

Hallinger, P., Heck, R. H., & Murphy, J. (2014). Teacher evaluation and school improvement: An analysis of the evidence. *Educational assessment evaluation & accountability, 26*, 5–28.

Helmond, P., Overbeek, G., & Brugman, D. (2012). Program integrity and effectiveness of a cognitive behavioral intervention for incarcerated youth on cognitive distortions, social skills, and moral development. *Children and Youth Services Review, 34,* 1720-1728.

Holcombe, A., Wolery, M., & Snyder, E. (1994). Effects of two levels of procedural fidelity with constant time delay on children's learning. *Journal of Behavioral Education, 4*, 49-73.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71,* 165-179.

Houchins, D. E., Jolivette, K., Shippen, M.E., & Lambert, R. (2010). The advancement of high quality literacy research in juvenile justice: Methodological and practical considerations. Behavioral Disorders, 36(1), 61-69

Hsieh, H., Wilder, D. A., & Abellon, O. A., (2011). The effects of training on caregiver implementation of incidental teaching. *Journal of Applied Behavvior Analysis, 44,* 199-203.

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004)

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1982/1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*, 197–209.

Johnson, E., Mellard, D.F., Fuchs, D., & McKnight, M.A. (2006*). Responsiveness to intervention (RTI): How to do it.* Lawrence, KS: National Research Center on Learning Disabilities.

Johnson, L. D., Wehby, J. H., Symons, F. J., Moore, T. C., Maggin, D. M., & Sutherland, K. S. (2014). An analysis of preference relative to teacher implementation of intervention. *The Journal of Special Education, 48*¸214-224.

Jones, H. A., Clarke, A.T., & Power, T. J. (2008). Expanding the concept of intervention integrity: A multidimensional model of participant engagement. *In Balance, 23*, 4-5.

Jones, K. M., Wickstrom, K. F., & Friman, P. C. (1997). The effects of observational feedback on treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 12*, 316-326.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings*. New York, NY: Oxford University Press.

Keller-Margulis, M. A. (2012). Fidelity of implementation framework: A critical need for response to intervention models. *Psychology in the Schools*, 1-11.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity: An essential--but often forgotten--component of school-based interventions. *Preventing School Failure, 48,* 36-43.

Leon, Y., Wilder, D., 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,*19-33.

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,* 659-672.

McKenna, S. A., Rosenfield, S., & Gravois, T. (2009). Measuring the behavioral indicators of instructional consultation: A preliminary validity study. *School Psychology Review, 38,* 496-509.

McKinney, T & Vazquez III, E. (2014). There's a bug in your ear!: Using technology to increase the accuracy of DTT implementation *Education and Training in Autism and Developmental Disabilities,49,* 594-600.

McLeod, B. D., Southam-Gerow, M. A., & Weisz, J. R. (2009). Conceptual and methodological issues in treatment integrity measurement. *School Psychology Review, 38,* 541-546.

Missett, T. C. & Foster, L. H. (2015). Searching for evidence based practice: A survey of empirical studies on curricular interventions measuring and reporting fidelity of

implementation published during 2004-2013. *Journal of Advanced Academics, 26,* 96-111.

Nelson, M. C. Cordray, D. S., Hulleman, C. S., Darrow, C. L. &Somme, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research, 39,* 374-396.

No Child Left Behind Act of 2001, P.L. 107-110, 20 U.S.C. § 6319 (2002).

Noell, G. (2014). Implementation of treatments in research and practice: Commentary on the special issue. *Journal of Behavioral Education, 23,*192-200.

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

Noell, G., Gansle, K., Mevers, J., Knox, R., Mintz, J. & Dahir, A. (2014). Improving treatment plan implementation in schools: A meta-analysis of single subject design studies. *Journal of Behavioral Education, 23,*168-191.

Noell, G., 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,* 77-88.

Odluyrt, S., Tekin-Iftar, E., & Adalioglu, I. (2012). Does treatment integrity matter in promoting learning of children with developmental disabilities? *Topics in Early Childhood Special Education, 32,* 143-150.

Odom, S. L., Brantlinger, E., Gerstein, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 7,* 137-148.

Pas, E. T., & Bradshaw, C. P. (2012). Examining the association between implementation and outcomes: Statewide scale-up of school-wide positive behavior interventions and supports. *Journal of Behavioral Health Services & Research, 39,* 417-433.

Pence, S., St. Peter, C. Giles, A. (2014). Teacher acquisition of functional analysis methods using pyramidal training. *Journal of Behavioral Education, 23,*132-149.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice, 12,* 365.

Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15,* 477-492.

Power, T. J., Blom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kelleher, C., & Manz, P. (2005). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools, 42,* 495-507.

Ransford, C. R., Greenbert, M. T., Domitrovich, C. E., Small, M., & Jacobson, L., (2009). The role of teachers' psychological experiences and perceptions of curriculum supports on the implementiona of a social and emotional learning curriculum. *School Psychology Review, 38,* 510-532.

Reinke, W., Stormont, M., & Herman, K., (2014). Using coaching to support teacher implementation of classroom-based interventions. *Journal of Behavioral Education, 23,*150-167.

Rosenheck, R. A. (2001). Organizational process: A missing link between research and practice. *Psychiatric Services, 52 (12),* 1607-1612.

Sanetti, L. H. & Collier-Meek, M. (2014). Increasing the rigor of procedural fidelity assessment: An empirical comparison of direct observation and permanent product review methods. *Journal of Behavioral Education, 23,*60-88.

Sanetti, L. H. & Fallon, L. M. (2011). Treatment integrity assessment: How estimates of adherence, quality, and exposure influence interpretation of implementation. *Journal of Educational and Psychological Consultation, 21,* 209–232.

Sanetti, L. M. & Kratochwill T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38,* 445-459.

Schoenwald, S. K, Henggeler, S. W., Brondino, M., J, & Rowland, M. D. (2000). Mutisystemic therapy: Monitoring treatment fidelity. *Family Process, 39,* 83-103. doi: 10.1111/j.1545-5300.2000.39109.x

Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review, 38,* 460-475.

Sheridan, S. M., Swanger-Gagne, M., Welch, G. W., Kwon, K, & Garbacz, S. A. (2009). Fidelity measurement in consultation: Psychometric issues and preliminary examination. *School Psychology Review, 38,* 476-495.

Simpson, R. L. (2005). Evidence-based practices and students with autism spectrum disorder. *Focus on Autism and Other Developmental Disorders, 20,* 140-149.

Slocum, T. A., Dietrich, R., Wilczynski, S. M., Spencer, T. D., Lewis, T., & Wol, K. (2014). The evidence-based practice of applied behavior analysis. *The Behavior Analyst, 37,* 41-56.

Smith, S. W., Daunic, A. P., & Taylor, G. G. (2007). Treatment fidelity in applied educational research: Expanding the adoption and application of measures to ensure evidence-based practice. *Education and Treatment of Children, 30*, 121-134.

Sprague, J. R., Scheuermann, B., Wang, E., Nelson, C. M., Jolivette, K. & Vincent, C. (2013). Adopting and adapting PBIS for secure juvenile justice settings: Lessons learned. *Education and Treatment of Children, 36,* 121-134.

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

Stage, S. A., Jackson, H. G., Jensen, M. J., Moscovitz, K. K., Bush, J. W., … Pious, C. (2008). A validity study of functionally-based behavioral consultation with students with emotional/behavioral disabilities. *School Psychology Quarterly, 23*, 327-353.

Stahmer, A. C., Reed, S., Lee, E., Reisinger, E. M, Connell, J. E., & Mandell, D. S. (2015). Training teachers to use evidence-based practices for autism: Examining procedural implementation fidelity, *Psychology in the Schools, 52,* 181-195.

Suess, A., Romani, P, Wacker, D., Dyson, S., Kuhle, J., Lee, J., …Waldron, D. (2014). Evaluating the treatment fidelity of parents who conduct in-home functional communication training with coaching via Teleheath. *Journal of Behavioral Education, 23*,34-59.

Sutherland, K. S., Wehby, J. H., & Copeland, S. R. (2000). Promoting positive interactions between teachers and students with emotional/behavioral disorders. *Preventing School Failure, 44*, 110-115.

Swanson, E., Wanzek, J., Harn, C., Ciullo, S., & McCulley, L. (2011). Intervention fidelity in special and general education research journals. *The Journal of Special Education, 47* 3-13.

Vartuli, S. & Rohs, J. (2009). Assurance of outcome evaluation: Curriculum fidelity. *Journal of Research in Childhood Education, 23,* 502-512.

Vollmer, T. R., & Iwata, B. A. (1992). Differential reinforcement as treatment for behavior disorders: Procedural and functional variations. *Research in Developmental Disabilities, 13*, 393–417.

Vollmer, T., Roane, H., Ringdahl, J, & Marcus, B. (1999). Evaluating treatment challenges with differential reinforcement of alternative behaviors. *Journal of Applied Behavior Analysis, 32*, 9-23.

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

Watson, T. S., Ray, K. P., Sterling-Turner, H., & Logan, P. (1999). Teacher-implemented functional analysis and treatment: A method for linking assessment to intervention. *School Psychology Review, 28,* 292-302.

Wheeler, J., J, Baggett, B. A., Fox, J., & Blevins, L. (2006). Treatment integrity: A review of intervention studies conducted with children with autism. *Focus on Autism and Other Developmental Disabilities, 21* 45-54.

Wickstrom, K. F., Jones, K. M., LaFleur, L. H., & Witt, J. C. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 13,* 141-154.

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

Wilkinson, L. A. (2007). Assessing treatment integrity in behavioral consultation. *International Journal of Behavioral Consultation & Therapy, 3*, 420-432.

Witt, J. C., Noell, G. H., LaFleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30,* 693-696.

Zvoch K. (2009). Treatment fidelity in multisite evaluation: A multilevel longitudinal examination of provider adherence status and change. *American Journal of Evaluation.30,* 44–61.

Chapter 2

EFFECTS OF ERRORS OF COMMISSION ON STUDENT PERFORMANCE DURING DISCRETE TRIAL TASKS

Statement of the Problem

Treatment integrity (TI) is often defined as the extent to which an intervention is implemented as intended (Codding, Livanas, & Pace, 2007; Dane & Schneider, 1998; Gresham, 2005;Gresham, Gansle, Noell, Cohen, & Rosenblum, 1993; Perepletchikova & Kazdin, 2005). The terms *procedural fidelity, treatment fidelity,* and *intervention fidelity* are sometimes used to refer to the same concept. The importance of documenting TI in education is clear. Both the reauthorization of NCLB and IDEIA mandate that professionals must use evidence-based practices to make educational decisions (Smith, Daunic, & Taylor 2007). However, it is not possible to make decisions on evidence-based practices unless it is clear that those evidence-based practices have been implemented accurately. For example, the Response to Intervention (RTI) model is designed to ensure that a student receives the least-intrusive practices necessary to make educational progress. Once an intervention has been tried, and the student has not profited from the intervention, she may then be moved into the next tier of RTI where more intensive interventions and services are offered. Eventually, if the student has not made adequate progress, special education services will be required. However, if the instruction has not been implemented with appropriate levels of TI, it is impossible to know if the student needs more intensive services, or simply needs the current instruction to implemented accurately and consistently.

The stakes are high for students because educators regularly use information from RTI and behavior intervention plans (BIPs) to determine educational placement. Placement decisions are supposed to be based upon the interventions that a student requires. However, if TI is low, there is no way to determine which services and interventions have benefited a student or which interventions are necessary.

Power et al. (2005) and Zvoch (2009) are clear that TI should be used to evaluate not only the implementers of a program (usually teachers), but also the program itself. Using TI data to evaluate a program can lead to the removal of superfluous components. For example, if TI of a BIP is evaluated, and the data indicate that reinforcement in the form of stickers are never offered as prescribed in the BIP, there are two possible routes of action. First, if the student's behavior is not improving, the TI data may be used to remind the teacher of the importance to implementing all BIP components. Second, if the student's behavior is at acceptable levels even without the sticker reinforcement, this component can be assumed to be unnecessary. The sticker component can then be removed, streamlining the teacher's workload as well as moving the student toward a less-intensive BIP. Using TI data in this way allows for accurate decisions to be made regarding both large (placement) and small (day-to-day routine) matters that affect student functioning.

The brief definition of TI (an intervention being used as intended; Codding, et al., 2007; Gresham, 2005) is useful as a way to explain a complex concept, but due to its brevity, the complexity of TI can be lost if the brief definition is the only one examined. Dane and Schneider (1998) identify five dimensions of TI that should be examined. These five dimensions have served as an initial basis to analyze group-design studies of TI (see for example Nelson, Cordray, Hulleman, Darrow & Sommer, 2012 and Pas & Bradshaw, 2012). Dane and Schneider's five

dimensions are: adherence, quality, exposure, participant responsiveness, and program differentiation. Briefly, adherence means the extent to which the written procedure is followed exactly as intended. Adherence is often measured through checklists that count the number of components that are completed accurately. Quality refers to the more intangible aspects of an intervention such as enthusiasm of the implementer, but specifics on how to measure quality are unclear. Exposure often refers to the duration or number of the prescribed opportunities that the student receives. This is measured by collecting the appropriate frequency or duration data. Participant responsiveness measures the extent to which students respond to the intervention in the desired manner, either through behavior change or skill acquisition. Researchers measure participant responsiveness by comparing baseline levels of the behavior or skill in question to levels during the intervention. Surprisingly, participant responsiveness, while seemingly crucial, is not often reported in the professional literature (Wickstrom, Jones, LaFleur & Witt, 1998). Instead, levels of TI and ways to improve TI are reported, but the effects of differing levels of TI on students are not reported. The fifth and final component is program differentiation, indicating both the difference between the current intervention and others, as well as the clarity of implementation (removal of unimportant steps and steps that are contra-indicated). Perhaps predictably, the component that is measured most often is adherence, primarily because of the ease of data collection.

In educational research, even with a more nuanced definition, TI is rarely measured (Wheeler et al., 2009). Between the years 1975 and 2000, authors of only 27% of studies involving students with EBD reported any measure of treatment integrity (Smith et al., 2007). Behavioral researchers report TI at similarly low rates. (McIntyre, Gresham, DiGennaro, & Reed, 2007; Wheeler, Baggett, Fox, & Blevins, 2006; Wilkinson, 2007).

Researchers outside of education have made advances in the measurement and understanding of TI which can and should be used to inform educational research. The majority of this TI work has been conducted by researchers in psychotherapy, youth violence prevention, and substance abuse treatment (Schulte, Easton, & Parker, 2013). Some of the ideas from those outside of education have been incorporated into the development of new tools such as concept-mapping (Green, Fettes, & Aarons, 2012) and measurement systems (Nelson et al., 2012) that can be applied to both behavioral health and educational research. Many of these innovations can be tied to federally funded studies specifically focused on TI (Shulte et al., 2013).

The new focus on TI research has spurred work applicable to large-scale, group-design studies (Helmond, Overbeek, & Brugman, 2012; Nelson et al., 2013; Pas & Bradshaw, 2012). However, the mandate for the inclusion of TI is not limited to group-design research. Horner et al. (2005) indicated that measuring TI was "highly desirable" (p. 174) not only as a method to ensure accuracy of intervention delivery but also as a component of social validity when translating research findings into practice by educators and caregivers. Additionally, Kratochwill et al. (2010) listed TI as one of the components that must be included in quality single-case design research in their document for the *What Works Clearinghouse*. Horner et al. in particular are clear that without measuring TI in the intervention phases of research, it is not possible to draw conclusions about the effect of the independent variable on the dependent variable.

Identifying a functional relation between a dependent and an independent variable is one of the underpinnings of applied behavior analysis (ABA; Fryling, Wallace, & Yassine, 2012). The information gained from the identification of a functional relation is what drives interventions in ABA (Alberto & Troutman, 2017; Cooper, Heron, & Heward, 2007; Kazdin, 2011). As mentioned earlier, TI (i.e., the monitoring of the independent variable) has not always

received the same careful scrutiny that is afforded to the dependent variable (e.g., documenting interobserver agreement; McIntyre et al., 2007; Wheeler et al., 2009). This may be changing with the Behavior Analysis Certification Board's most recent set of standards stating that behavior analysts have an obligation to measure TI to ensure accurate delivery of services as well as to ensure that interventions are modified as needed (BACB, 2012).

Review of ABA Literature

TI Related to Student Outcomes

Dane and Schneider (1998) parsed TI into five components, or dimensions: adherence, quality, exposure, participant responsiveness, and program differentiation. Other researchers have used these dimensions to describe TI as well (Power et al. 2005). However adherence, or ensuring that each part of an intervention is completed as intended is the component most frequently measured in the TI research (Durlak & Dupree, 2008; Sanetti & Kratochwill, 2009; Frying et al., 20212). Even when adherence is the only aspect of TI being measured, there are clear implications for student outcomes as has been shown in applied behavior analytic single-case research. Often, the higher the level of TI, the better the student response. In an early study, Holcombe, Wolery, and Snyder (1994) manipulated the TI of constant time delay (CTD) used to teach an academic skill to students with intellectual disability. Students learned the skill when CTD was used with both 100% and 50% accuracy. They learned little to nothing when CTD was used with 0% accuracy. However, students learned far more quickly when CTD was used with 100% accuracy than with 50% accuracy. Therefore, not only is the effectiveness of an intervention affected by TI, but the efficiency may be affected as well. Noell, Gresham, and Gansle (2002) also found that their students learned both faster and more efficiently when TI was high. Grow et al. (2009) manipulated TI of the well-validated system of least prompts (SLP) to

make it resemble the multiple verbal prompts that teachers actually use. Students in that study also learned much more efficiently when SLP was implemented with high TI.

Other researchers have found functional relations between TI and student performance. Sanetti and Kratochwill (2009) reported that students made small, but important gains in math with 100% TI producing the best outcomes. For two participants, the 50% and 0% TI conditions were indistinguishable and produced very poor outcomes, suggesting that even some errors may prove disastrous for some students. Groskreutz, Groskreutz, and Higbee (2011) suggested that different types of errors in implementation may differentially affect student outcomes. They classified errors as those of either commission or omission.

Errors of commission involve reinforcement of inappropriate responses or reinforcement at inappropriate times. In contrast, errors of omission involve a failure to reinforce appropriate responses (Groskreutz et al., 2011; Grow et al., 2009; Odluyrt, Tekin-Iftar, & Adalioglu, 2012). St. Peter Pipkin, Vollmer, and Sloman (2010) conducted a systematic examination of different types of errors. They manipulated errors of omission, as well as errors of commission. They found that when errors of omission were committed, the target behavior did not improve, and lower rates of alternative behavior were seen. Interestingly, errors of commission did not affect student behavior until TI was at 40%. This finding lends initial support for the theory that errors of commission may not be as detrimental as errors of omission. DiGennaro Reed, Reed, Baez, and Maguire, (2011) conducted a parametric analysis of errors of commission during discrete-trial teaching. They implemented discrete trial teaching (DTT) with 100%, 50%, and 0% levels of TI. They found that 100% integrity produced the best skill acquisition and that 50% integrity and 0% were equally detrimental to skill acquisition for two of their three participants. To date,

these are the only two studies which have systematically manipulated errors of commission; therefore, implications for practice are not yet clear.

Treatment integrity is an essential component in the application of interventions in both clinical and applied settings. Measuring the fidelity with which an intervention is applied allows for a more thorough and accurate understanding of which components of an intervention are effective, necessary, and feasible. Failing to examine TI can lead to what Vollmer, Sloman, and St. Peter Pipkin (2008) called false positives and false negatives. False positives occur when a practitioner or a clinician erroneously assumes that improvements in student behavior are based upon an intervention, when in fact that intervention has been provided with low TI. False negatives occur when a treatment is abandoned due to poor student outcomes, even though the treatment was not implemented with TI at a high enough level to determine its efficacy. A broader understanding of the concept of treatment integrity, as well as better methods for measuring all dimensions of treatment integrity will lead to more accurate understanding of which treatments are beneficial for students.

Given that it is unreasonable to expect 100% TI in applied settings, appropriately training teachers requires a thorough understanding of the types of TI errors that may be committed, as well as the levels of TI necessary for student success when errors inevitably occur. It is essential that TI be fully understood because of the real-world implications for students, including movement into special education and special education placement changes which are based upon students' responses to interventions. It is not possible to make educational decisions on intervention effectiveness unless it is clear that evidence-based practices have been implemented accurately. If TI is low, there is no way to determine which services and interventions a student requires.

The purpose of this study is to gain a more nuanced understanding of TI failures in the form of errors of commission committed by DTT instructors and the role commission errors have on participant responsiveness (Power et al., 2005). This study will extend the results of DiGennaro Reed et al. (2011) by including an intermediate (80%) level of TI which may more accurately represent an attainable level of TI in applied settings. The hypothesis is that errors of commission will negatively affect student performance, and that higher levels of TI will result in increased skill acquisition. The research question is as follows: How do errors of commission in discrete trial teaching affect the skill acquisition of elementary-aged students with autism?

Method

Participants

The researcher chose elementary-age students with autism because of the documented utility of DTT for this group (Eikeseth, Smith, Jahr, & Eldevik, 2002; Howard, Sparkman, Cohen, Green, & Stanislaw, 2005; Smith, Groen, & Wynn, 2000). Initially, the researcher planned to enroll eight students because that number is sufficient to show any treatment effects using the design described below while allowing for possible attrition of participants (Kazdin, 2011).

Six eligible students returned permission slips. All six students were enrolled in a self-contained public school that served students with behavioral problems. The students were served in three different classrooms and had three different teachers. Additionally, all students had special education eligibilities of autism, with measured IQ <70 (as reported by the students' teachers based on testing done in the school district). The students were between the ages of 6 and 9 years old at the time consent forms were signed. All students had IEP objectives for which DTT was appropriate according to the criteria in Smith (2001). All students had been taught

using DTT during the current school year. All students were able to see the materials and hear verbal instructions and none had reports of visual or hearing impairments in their school files. To be eligible to participate, students had a record of regular attendance (≤1 absence per week). Students who did not meet these criteria were excluded. Four students completed the study. One student was moved to a private behavioral treatment facility before completion of the study and one student moved out of the district. See Table 1 for descriptive information about each participant who completed the study as well as the DTT objectives targeted.

Table 1.

Basic Participant Information

Student	Age	IEP Objective (100, 80, 50% TI)	Reinforcers Identified by RAISD
Gray	8.4	Receptive Color ID Red, Blue, Green	Candy, Oreos
Duquon	8.3	Receptive Letter ID O, N, I	Starburst, M&Ms, small horse toys
Cormac	6.2	Receptive Letter ID I, E, N	Small manipulatives like counting beads, tickles, spinning tops
Thomas	8.7	Receptive Sound ID T, N, K	Birthday-themed items, Dora the Explorer, candy

Gray. Gray was 8.4 years old. Gray was not very physically active and sat still during all sessions, although he was prone to episodes of prolonged, loud crying that his teacher indicated could last from 1 to 40 minutes. Gray's family spoke Spanish and English at home. Both Gray's parents and school staff indicated that he understood English and English was the only language

used in the classroom. Gray communicated by bringing objects to adults or leading adults by the hand to desired objects.

Duquon. Duquon was 8.3 years old. His teacher reported that he was an agile boy who was quick to smile, and equally quick to scratch himself and others if he became agitated. Duquon did not have any consistent means of communication, although he would occasionally tap on or bring pictures of desired items to adults in his environment. On more than one occasion, the researcher witnessed Duquon tap on icons without gaining anyone's attention first. Sometimes, he would then appear agitated that no one had honored his requests, other times, he would simply move on to another activity.

Cormac. Cormac was 6.2 years old. His teacher reported that he was an active little boy who often engaged in aggressive behavior in the form of hair pulling and hitting others. He had very little verbal communication (mama, go, eat) and primarily communicated by pointing to pictures and/or handing Mayer-Johnson Boardmaker symbols to adults in his environment. His teacher reported that Cormac used the picture symbols to request items, but not to describe his environment or communicate his feelings.

Thomas. Thomas was 8.7 years old. Thomas communicated verbally. He used communication to indicate his wants and needs only, rather than to engage in conversation. His speech was often echolalic, but the communicative intent was often clear. For example, if the teacher said, "What do you want to drink, Thomas?"; he would reply, "What do you want to drink, Thomas? I want juice." Thomas's primary maladaptive behavior was what the teacher called "meltdowns" which were comprised of hitting and scratching others, screaming, throwing furniture and attempts at elopement.

The four objectives chosen by the teachers are all appropriate for DTT instruction according to the criteria referenced in Smith (2001). This meant that the skills to be taught were discrete or could be broken down into discrete components, and were either new forms of behavior, or in the case of the students in this study, were new discriminations of current skills such as learning to identify new targets for a previously acquired skill (Smith, 2001). For all of the students, the objectives were receptive identification skills. Receptive identification skills are commonly taught to children with autism as a way to increase vocabulary (Maurice, 1996; Goldstein, 2002).

Setting

The study took place in special education classrooms in a public school for students who have autism and behavioral problems in the Southeastern United States. For all but Duquon, sessions took place in the students' regular classrooms. Each classroom was approximately 25 by 25 feet. It was standard procedure at the school for students to receive DTT instruction in their classrooms while other students received group or individualized instruction. All DTT instruction was provided by the researcher in the same location within the classroom as when it was provided by the teacher. For Cormac and Thomas, DTT took place in the "DTT station" for their classrooms. In both of these classrooms, this area had a kidney-shaped table in a corner of the room. The researcher sat in the indentation, and the student sat on the outside of the table. Gray received DTT instruction in his desk in his classroom. Duquon's teacher found him to be so distracted in the large classroom setting that little to no progress was being made on DTT objectives. For this reason, Duquon received his DTT instruction in an empty classroom across the hall. This change was made by the teacher for her DTT sessions during the regular school day so research sessions were conducted in the same way. Duquon sat at a rectangular table

across from the researcher during his sessions. DTT constituted the typical form of instruction in these settings, so participation in the study did not disrupt students' educational programming.

Teachers chose each student's DTT objectives from his IEP and provided the researcher with information about which targets had not yet been introduced. All materials were either provided by the classroom teacher or made by the researcher according the size, shape, and color of materials used by the teacher when teaching previously introduced targets.

Materials

Discrete trial data sheets for recording student progress (Appendix A), and a combined inter-observer agreement and TI data sheet (Appendix B) are attached. The Reinforcer Assessment for Individuals with Severe Disabilities (RAISD; Fisher, Piazza, Bowman, & Amari, 1996; Appendix C) was used to choose potential tangible reinforcers. Additionally, materials appropriate for teaching the relevant IEP tasks (flash cards, objects) were used. A small camcorder was used to tape all sessions for the purposes of monitoring TI and collecting data for inter-observer agreement (IOA).

A social validity checklist was used (Appendix D). The survey was administered online using the SurveyMonkey website. The 3 teachers received an email with a link to the survey at the end of the study. The classroom staff did not have a significant role in the research project, other than making the students available to the researcher at scheduled times. However, it is hoped that the results from the study benefitted the classroom staff and informed their DTT practices.

Dependent Variable

The dependent variable was the percentage of discrete trials with correct student responses. The researcher conducted all sessions and simultaneouusly collected discrete trial data

in real time using the data sheet in Appendix A. Each session consisted of approximately10 trials of the same skill. The number of correct trials was converted to percentage accuracy by dividing the total number of correct responses by the total number of trials and multiplying by 100%. The researcher calculated interobserver agreement (IOA) and verified TI levels by having a trained observer independently collect data for 30% of sessions in all phases. The independent data were collected by having the trained observer watch videotaped sessions.

Table 2

Interobserver Agreement between Investigator and Second Observer

Student	Total # of Sessions	% of Sessions Evaluated	IOA (range) (mean)
Gray	12	33%	97-100% 99.7%
Duquon	15	36%	95-100% 99.3%
Cormac	14	57%	94-100% 99.5%
Thomas	16	33%	97-100% 99.5%

A Panasonic HDC-TM90 digital video recorder with internal memory was used to record sessions. The camera was positioned to capture both student responses to DTT stimuli and the researcher's reinforcement of student responses. Due to technical difficulties and student behaviors (throwing the camera, room too noisy to hear student responses on the tape) the percentage of sessions taped successfully varied for each student. The percentage of sessions successfully recorded was as follow: Gray (33%), Duquon (36%), Cormac (57%), and Thomas (33%). Because of the low number of sessions that were successfully recorded, all taped sessions

were viewed by the second observer. Interobserver agreement was calculated using point-by-point agreement between the researcher and the second observer, taking the total number of agreements and dividing by the total number of agreements plus disagreements and multiplying by 100%. IOA ranged from 97.8% to 99.6% across all phases of the study. IOA data for individual participants are reported in Table 2.

Independent Variable

The independent variable was the level of treatment integrity with which instruction was provided. TI was provided at 100%, 80% and 50% levels. Both St. Peter Pipkin et al. (2010) and DiGennaro Reed et al. (2011) found 0% TI to be extremely detrimental to learning. Given the applied setting, and the fact that 0% has not been shown to be beneficial to students, this study did not include a 0% condition. However, the 80% condition provided information about higher levels of TI and may give insight about what level of accuracy is necessary and sufficient for effective DTT.

The researcher reinforced and corrected responses according to the procedures outlined in DiGennaro Reed et al. (2011). For unprompted correct responses, specific social praise ("Good touching *A.*") and a tangible reinforcer (based upon the classroom reinforcement system and the results of the RAISD) were provided. Least-to-most prompting (gestural, then physical) was used for error correction. Prompted correct responses were followed by a neutral statement ("that is *A*"). The researcher demonstrated errors of commission by presenting reinforcement for incorrect responses at predetermined levels to examine effects on student performance. An example of a programmed error of commission followed this pattern: Student instructed to touch *B,* student touches *A*, the researcher provides verbal praise as for a correct response ("Good touching *B.*") and a tangible reinforcer.

Each level of TI (100%, 80%, and 50% accurate reinforcement) was associated with a different stimulus from the same IEP objective. For example, for a letter identification IEP objective, identifying the letter *A* received programmed consequences at the 100% TI level, the letter *B* at the 80% level, and the letter *C* at the 50% level. During the 100% TI condition (e.g., *A*), no errors of commission were made and students were reinforced only for correct responses. During 80% TI (e.g., *B*), one out of every five errors was reinforced by committing the error of commission on the first and 6th errors. During the 50% errors of commission condition (e.g., *C*), every other incorrect response was reinforced using the same verbal praise and tangible reinforcer provided during a correct trial. Each session consisted of approximately 10 interspersed trials of each target, reinforced at the pre-determined level. The order of conditions was counterbalanced across sessions. The number of trials was approximate because student performance dictated how many trials were necessary to get to the required percentage of TI. See Table 1 for information about individual student's targets and associated levels of TI.

Treatment integrity data were collected to ensure that the researcher provided error correction and consequence manipulation correctly. Please see the second page of Appendix B. for an example of a completed 80% errors of commission TI/IOA data sheet. In order to determine TI, the second observer watched the video tapes described earlier to see if error correction, and consequence manipulation were done correctly. Agreements were divided by agreements plus disagreements. The quotient was multiplied by 100%. See Table 3 for TI results.

Table 3.

Treatment Integrity of Independent Variable

Student	Total # of Sessions	% of Sessions Evaluated	TI (range) (mean)
Gray	12	33%	98-100% 99.7%
Duquon	15	36%	95-100% 99.2%
Cormac	14	57%	97-100% 99.8%
Thomas	16	33%	98-100% 99.8%

Design

A combined multi-element and multiple-baseline across participants design was used as in DiGennaro Reed et al. (2011) with the slight modification to include probe data. Using a multiple-baseline across participants design allowed for replication of the TI manipulation (Kazdin, 2011). The probe data were collected just before moving the students in tier one and two from baseline into intervention in order to ensure that the student still needed to learn the target skill and rule out history or maturation effects. The multi-element component of the design was chosen because it allowed for the manipulation of multiple interventions in the same phase (Kazdin, 2011). In this study, each level of treatment integrity was associated with a distinct stimulus (chosen from the student's IEP objectives). The multiple-baseline component of the design was chosen because once one of the IEP skills has been introduced, there was no feasible

way to withdraw the skills already acquired by the students (Kazdin, 2011). The replication from each subsequent tier demonstrates the functional relation in this design. The six participants were divided into two groups of three students each. Within groups, the students began intervention in the order in which they returned their permission forms. The reason for student grouping was so that no student had to be in baseline for an extended period. Another safeguard against extended baselines is the modification to include probe data rather than consecutive baseline sessions for the second, and third participants in each group. The first student in each group remained in intervention until a clear differentiation was seen among the conditions (no overlapping data points for nine sessions—three per TI level), or one condition had an ascending trend (three consecutive ascending non-overlapping data points), or the student has participated in five sessions at each TI level (15 total). At this point, the second student from the corresponding group moved from baseline to intervention.

Unfortunately, the students with whom Gray was initially paired dropped out of the study before enough data were collected to demonstrate replication of the intervention. Therefore, Gray's data are presented as the top tier of the graph as another visual depiction of the intervention. It is important to note that all of Gray's sessions took place in a different classroom, in a another wing of the building from the other participants and it is unlikely that their performance influenced him.

Procedures

Reinforcer assessment. Prior to entering baseline, teachers and classroom aides completed the *Reinforcement Assessment for Individuals with Severe Disabilities* (RAISD; Fisher, Piazza, Bowman, & Amarai, 1996). The RAISD is a ten question short-answer survey that takes approximately 15-20 minutes to complete. The researcher verbally asked the teacher

and aides the RAISD questions and recorded their responses. The results were used to identify potential reinforcers for each student. The top tangible reinforcers that were able be delivered during a discrete trial session were used for each student in conjunction with the classroom reinforcement system. Using the RAISD allowed the researchers to identify current reinforcers across a variety of categories (edible, sensory, etc.) for each student. Additionally, using the RAISD helped to control the salience of reinforcers across skills and students. The results from the RAISD are described are in Table 1.

Baseline. The researcher presented stimuli appropriate to the IEP skill chosen in conjunction with the teacher using the basic discrete trial format outlined by Smith (2001) using a 5s delay to allow for student responding. The researcher did not provide any consequences contingent on performance. The intervention phase began when the student performance data were stable (at least three data points collected for each skill, and no data points 50% above or below the mean; Alberto & Troutman, 2017).

Intervention. Intervention took place in the same setting as baseline, using the same materials. The only differences between baseline and intervention phases were the errors of commission committed by the researcher as described and the reinforcement schedule. Reinforcement for correct responses was provided on an FR1 schedule. Incorrect responses received either an error of commission response (reinforcement) according to the predetermined schedule, or least-to-most prompting (Alberto & Troutman, 2017) followed by a neutral statement. One session of each TI condition (100%, 80%, 50%) counted as one "cycle" of sessions. There was a maximum of four cycles per child, per day. There was a minimum of five minutes between each cycle. A predetermined cap of 60 sessions for any one student was also in place.

Social validity. Teachers completed the social validity questionnaire anonymously at the end of the study. A link to the survey was e-mailed to teachers at the end of the study. The survey was designed to take approximately 5 minutes to complete. The questions were on a four-point Likert-type scale. The teachers did not find the study to be intrusive and generally found there data presented by the researcher to be beneficial. Results of the social validity questionnaire are presented in Table 4.

Table 4.

Social Validity Survey Results

Teacher	Study useful to me	Study useful to my students	Having researcher in room was intrusive	Study took unacceptable amount of student time	Feedback received at end of study was helpful	Feedback received will affect teaching
Ruth	1	1	4	4	1	1
Maggie	2	2	4	4	2	2
Laura	1	1	4	4	1	1
Average	1.3	1.3	4	4	1.3	1.3

*1 = strongly agree, 2=agree, 3=disagree, 4=strongly disagree

In addition to the information collected using the social validity checklist in Appendix D, teachers were informed about the level of TI that worked best for each of their students. This information was provided for teachers so that they could customize instruction. While it was not recommended that teachers intentionally make errors of commission, it is important to tell teachers just how crucial it is for each student to receive evidence-based practices exactly as written. For some students, 100% integrity may be required, while for others, 80% may be sufficient to demonstrate progress in DTT. This type of information can inform staffing decisions (putting the most experienced staff with students who need the most rigorous level of TI), and staff training.

Results

The results of the varying levels of TI in the form of errors of commission on student skill acquisition during DTT are presented in Figure 1. It was expected that students would learn the most when exposed to DTT with 100% TI (no errors of commission). It was also expected that decreasing levels of TI would be associated with decreased learning, with 80% TI being superior to 50%. This would mirror the outcomes from studies of TI errors of omission (Groskreutz et al., 2011; Grow et al., 2009; Noell et al., 2002; Sanetti & Kratochwill, 2009). However, the results from this study are mixed and more closely align with those of the work of DiGennaro Reed, et al. (2011) who found errors of commission may have differing effects across students.

Figure 1 shows the percentage accuracy across all sessions. Baseline performance was low for all students across all targets. The researcher used visual analysis as this is the "primary means" of evaluation in single-case research (Kazdin, 2011; p.285). Changes in means across phases, level changes across phases, and percent non-overlapping data were all used to evaluate the data (Alberto & Troutman, 2017; Kazdin, 2011). Each student's data will be discussed according to these parameters. For all data, the target associated with 100% TI is listed first, followed by 80%, and then 50% TI, with ranges presented in parentheses. Gray's data will be discussed last as there is no replication of his data.

Duquon averaged low accuracy for all conditions in baseline. Five baseline sessions were conducted and data were stable (no points 50% above or below the mean) for all of the targets. There was a clear change in the means across phases, with the highest change occurring in the 100% TI phase (identifying letter O) with a mean of 65% (20%-90%) versus 18% (10%-20%) in baseline. There was not an immediate level change in the 100% condition, rather a positive level

change began in the second session of intervention. This first data point was the only point of overlapping data for a total of 10% overlapping data. There was also a change in means for the 80% condition (identifying letter N) with a mean of 43% (20%-70%) versus 12% (10%-20%) in baseline. There was an immediate positive level change in the 80% TI condition, and there was 0% overlapping data. The smallest change was when identifying letter *I* in the 50% TI condition with a mean of 23% (10%-40%) during intervention versus 14% (10%-22%) in baseline. There was a level change from 0% to 30%, however that change did not sustain. The 50% TI condition had the highest percentage of overlapping data with 50%. Duquon showed clear differentiation among the conditions, with the 100% condition associated with highest accuracy, followed by 80%. For Duquon higher TI was associated with better skill acquisition. See Figure 1.

As seen in Figure 1, Cormac averaged low accuracy for all conditions in baseline. Four baseline sessions were conducted, with the last one being a probe session to ensure that the baseline had remained stable over time (Kazdin, 2011). Letter M was taught in the 100% TI condition. There was a change in the mean from 24% (10%-50%) in baseline to 80% (30%-100%) in intervention. There was an immediate positive level change from 25% to 80% between these phases as well. There was one overlapping data point in the 100% TI condition for a total of 10% overlapping data. The letter E was introduced in the 80% condition. In baseline, Cormac averaged 20% accuracy (0%-30%). The mean increased to 85% (60%-100%) when 80% TI was introduced. There was an immediate positive level change in this condition from 20% to 80%, and there were no overlapping data points. Finally, letter M was taught with 50% TI. There was a mean change in this phase as well, although it was slightly less than the other two conditions, from 20% (10%-30%) to 70% (30%-100%). There was 10% overlapping data in this condition. Cormac never showed clear differentiation among the conditions, with all levels of TI associated

with higher accuracy than in baseline. By the sixth intervention session, all conditions were associated with 80% or better accuracy. For Cormac, it did not appear that errors of commission at the levels presented were detrimental to skill acquisition.

As with the other participants, Thomas averaged low accuracy for all conditions in baseline. Thomas had four baseline sessions with the last being a probe to ensure maturation and history effects had not occurred. When working on the letter sound T in the 100% TI condition, there was a change in the mean from 13% (10%-20%) to 59% (40%-75%). There was an immediate positive level change from 10% to 50%. There were no overlapping data points in this condition. While Thomas only averaged 59% accuacy in this condition, it is important to note that Thomas's last four sessions were at 70% or better. While this does not meet his IEP goal criteria of 80% or better, it is approaching that level. It would have been preferable to run more sessions to see if an improving trend emerged or if data leveled off. The letter sound N was taught with 80% TI. There was a level change in the means from 15% (10%-30%) during baseline to 31% (10%-57%) during intervention. There was a modest positive level change from 10% to 30%. There were 64% overlapping data in this phase. The sound K was taught with 50% TI. The change in means was from 38% (30%-40%) in baseline to 37%(20%-50%) in intervention. There was a slight positive level change from 30% to 40% in intervention. There was 64% overlapping in this condition as well. While the letter sound K had a higher baseline than the letters assigned to the other two conditions, that sound had been predetermined (by the order the letters appeared in the student's name) to be assigned to the 50% condition. Any "advantage" that may K may have had initially should have been balanced out by being placed in the 50% TI condition. Additionally, none of the conditions showed great variability in baseline, strengthening the idea that change in the intervention phase may be due to the intervention

(Kazdin, 2011). By the eighth intervention session, Thomas showed clear differentiation among the conditions, with the 100% TI condition showing the best results. For Thomas higher TI was associated with better skill acquisition and the two lower TI conditions were not differentiated. See Figure 1.

Gray's data must stand alone as the two students with which he was paired left school before the study could be completed, therefore there is no replication of any effects. As shown in Figure 1, baseline data were stable after only three sessions in each condition. Gray's baseline mean for color identification for the color red taught with 100% TI was 10% (10%-10%) with a mean change to 70% (20%-100%). There was an immediate positive level change from 10% to 70%, and 0% overlapping data between baseline and intervention. In the 80% condition for the color blue, Gray averaged 13% accuracy (10%-20%), and had a mean change to 93% (70%-100%). There was an immediate, positive level change from 10%-70%. Again, there were no overlapping data points. Finally, when being taught the color green in the 50% TI condition, Gray averaged 10% (10%-10%) accuracy and had a mean change to 82% (40%-100%) during intervention. There was an immediate, positive level change from 10%-40%. There were no overlapping data points in this condition. For Gray, the 80% TI condition was associated with a highest level of accuracy. Both the 80% and the 50% conditions were associated with higher levels of accuracy than the 100% condition, as well as faster acquisition of the target skill. It is important to note that the 100% condition was technically "stable" in that no single data point was 50% above or below the mean of all intervention data points. This supports the idea that the DTT, even at low levels of TI may be associated with an increase in skill acquisition. However, the reason for the range in Gray's data, particularly in the 100% condition is unclear. There were factors outside of the researcher's control such as classroom noise and the behavior of other

students in the room that may have had an impact on Gray's performance. It is not clear why those factors would have had a stronger effect on the 100% TI condition than the others. Across all three conditions, there was an increase in the means from baseline to intervention, with Gray's highest average performance in the 80% TI condition. This was also the condition with the lowest range and least variability. The 50% TI condition, while lower than the 80% condition still produced better outcomes than the 100% condition. Finally, the percentage of non-overlapping data was examined as a means for assessing the effectiveness of each level of TI. There was a 0% overlap in all three conditions, suggesting that all three levels of TI were associated with increased receptive color ID for Gray.

Additionally, there were never 3 consecutive data points at 80% accuracy or higher (the IEP objective's target goal) in the 100% condition. It is possible that additional sessions may have improved stability in the 100% condition. However, both the 80% and 50% conditions were stable and had 3 or more consecutive data points above 80% and the teacher requested that the researcher stop working on red so that Gray could have a "break" from it as the school year was ending she wanted him to end the year on a "fun" note.

In summary, for Thomas and Duquon the 100% TI condition was associated with the highest level of acurracy the target DTT skills. For one student, Duquon, the 80% TI condition was associated with higher skill acquisition than the 50% condition. However, no other students demonstrated a difference between the 80% and 50% conditions. In Gray's case, 100% integrity was generally associated with *lower* acquisition of color identification than either the 80% or 50% conditions. For the remaining student, Cormac, all conditions were associated with skill acquisition of 80% or better, with no differentiation among the conditions.

Discussion

The effect of different levels of errors of commission on the skill acquisition of elementary-aged students with autism on DTT tasks was measured. The results from this study are mixed. Relatively higher rates of TI yielded better skill acquisition for two participants. For both Duquon and Thomas, 100% TI was associated with better skill acquisition, although it took longer for those effects to manifest for Thomas. However, for the other two participants, higher levels of TI were not associated with higher accuracy in skill acquisition. In fact, for Gray, higher TI was actually associated with lower accuracy. This study adds to the minimal body of research on the effect that errors of commission have on student learning. Although the effects were not consistent across participants, student performance did change when DTT instruction began. This suggests that DTT instruction had an effect on student behavior (Kazdin, 2011) across all levels of errors of commission. Visual analysis tends to show only strong effects (Kazdin), so it is possible that for the students who did not show clear differentiation amongst the conditions, there were weaker effects that could not be detected by this method. This study makes a small contribution to the body of literature on errors of commission. This study was the first to attempt to measure the way that different types of TI errors affect learning in DTT with students who have significant behavior problems and autism.

While there is a significant body of research demonstrating that students learn best when treatments are implemented with high integrity, research on which types of errors affect student learning, and in which ways, is scarce. Sanetti and Kratochwill (2009) found that students learned math best when TI was high. Groskreutz et al. (2011) delineated TI errors into either errors of omission (failing to provide reinforcement at appropriate times) and errors of

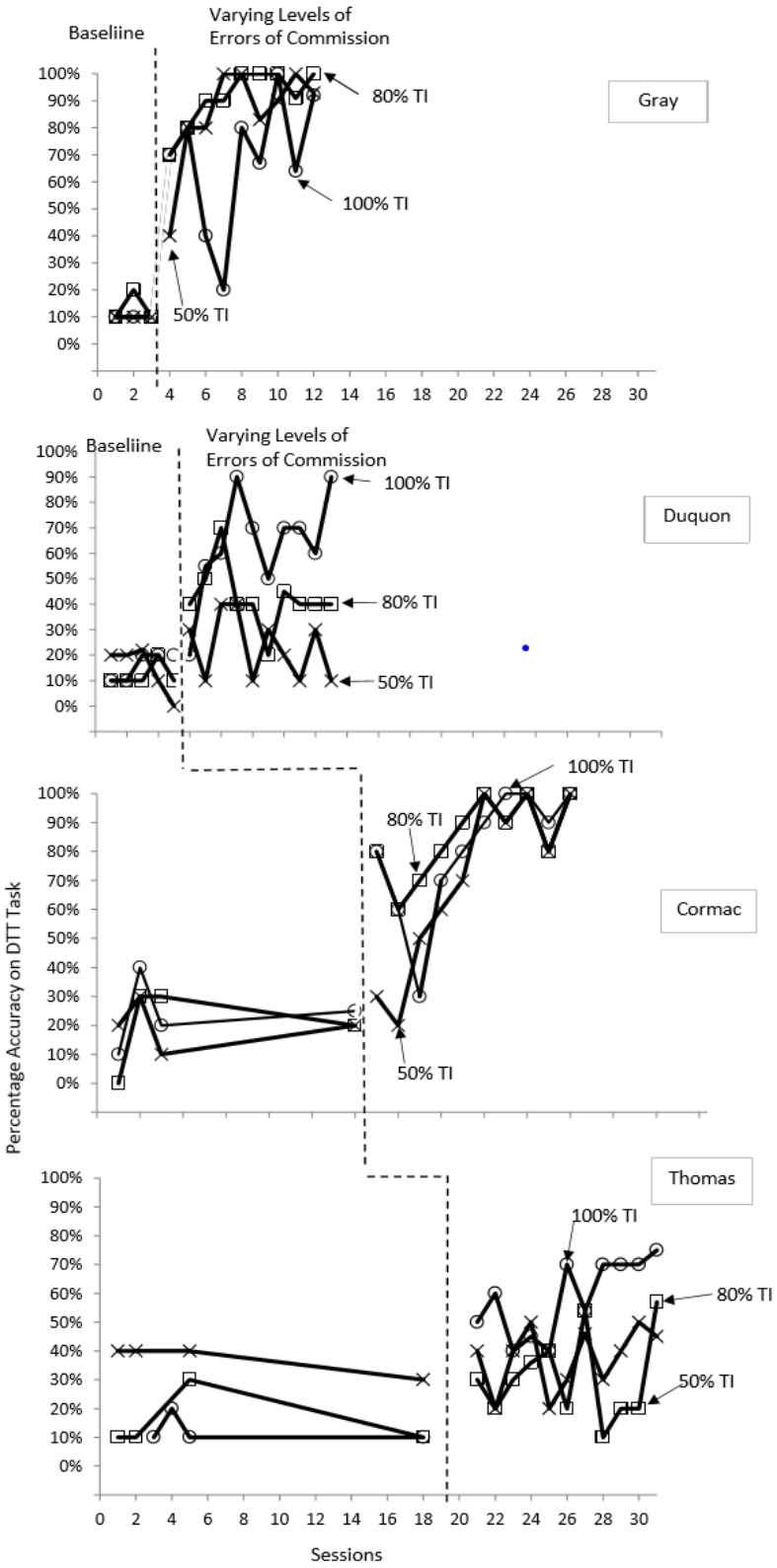

Figure 1. Varying levels of errors of commission on DTT accuracy

commission (providing reinforcement at an inappropriate time). Groskreutz et al.'s work was extended into a systematic analysis of types of errors by St. Peter Pipkin et al., (2010). St. Peter Pipkin et al. found that errors of commission were detrimental to student behavior once TI was at 40% or lower. It is possible that the 50% level chosen here was not low enough to show effects for some students. However, DiGennaro Reed et al. (2011) found that for some participants 50% TI was as detrimental as 0% TI. This study was intended to replicate DiGennaro Reed et al. (2011) with the modification that the 0% TI condition be dropped and an 80% TI condition added. The 0% condition was dropped because of the data showing that 0% integrity is not beneficial to students (Leon, Wilder, Majdalany, Myers, & Saini, 2014; DiGennaro Reed et al, 2011; Vollmer, Roane, Ringdahl, & Marcus, 1999). Given that this study took place in a public school, it did not seem ethical to use student learning time to engage in a practice shown to have no beneficial results. The 80% condition was added to test how a level of TI between 50% and 100% (the other conditions tested in DiGennaro Reed et al.) might affect student learning.

The results of this study partially mirror the work of DiGennaro Reed et al. (2011) and others. Two of the participants in this study, Duquon and Thomas had results that replicate previous research (Noell et al., 2002; DiGennaro Reed et al., 2011; Wilder, Atwell, & Wine 2006). Both of these students learned targets with greater accuracy when taught with higher levels of TI. However, results from the other two other participants differ from the published literature in that higher TI did not seem to have an impact on their learning. Gray consistently performed worse in the 100% TI condition. Neither the teacher nor researcher had a clear explanation for this pattern of behavior. Although this did not affect baseline performance, there may have been factors outside of the researcher's control such as the level of distraction in the room, or a unknown student characteristic such a preference for one color that may have affected

Gray's performance during intervention. It is important to note that while Gray performed most poorly in the 100% TI condition, he averaged 83% accuracy over the last five sessions. It would have been preferable to run more sessions to see if his data would stabilize, but the school year was coming to a close and this was not possible.

Cormac did not appear to be affected by errors of commission positively or negatively. The first two sessions at 50% TI were lower than the first two sessions at 80% and 100% TI. After that, all TI conditions appeared equal. Cormac learned all of the required letters with good accuracy. It would have been beneficial to collect both maintenance and generalization data to see if the skills taught with differing levels of TI maintained at different rates, as well as introducing a new set of targets to see if the pattern of responding held.

For all students in this study, their inappropriate behaviors (hitting and scratching self and others, leaving the area, sreaming, noncompliance), may have had an effect on student progress. Due to these inappropriate behaviors, students were often distracted and may have performed differently if they were more focused on the task at hand. Additionally, this school exclusively served students with behavior problems. At any given time, the classrooms in which sessions were run might be loud, or other students might be engaging in distracting or disruptive behaviors. This may have been a factor for all students except Duquon who worked on DTT sessions in a room with no other students, although noise from the hallway and neighboring classrooms could be heard at times. It is hard to determine exactly what effect these behaviors may have had on student learning, but it seems safe to assume most students may work better in a room free of distractions other than his own in appropriate behaviors.

Claims about maintenance and generalization cannot be made as the school year ended before these data could be collected. It was harder than anticipated to get students to return

permission slips, so the study began later than was intended. Additionally, collecting data took longer than anticipated due to the study being conducted in the students' natural environment. As stated earlier, the behavior of students in the study, and other students in the building sometimes lead to the cancellation of sessions.

This study has several limitations. First, all students in this study have severe behavior which delayed, interrupted, and cancelled sessions at various times. It is possible that these behaviors and their effects interfered with student performance to a degree that is impossible to parse. Running a study of errors of commission with an 80% TI condition with students who have autism but not severe behavior might yield clearer results. For example, DiGennaro Reed et al. (2011) describe their participants as having autism, but do not provide any information about IQ or behavioral concerns. St. Peter Pipkin et al. (2010) examined errors of commission with typically functioning undergraduate students and one student with moderate intellectual disability and some disruptive behavior. Second, although all students in this study had previous exposure to DTT, the extent that each student has benefitted from DTT in the past, or if other teaching strategies have been more beneficial, is unknown.

Although teachers and others who work in non-clinical settings are unable to teach with 100% TI, the frequency with which teachers make different types of errors in different settings and with different teaching techniques is not known. Given this lack of information, it is unlikely that the predetermined levels of TI chosen for this study, and that fact that only errors of commission were made, mirror what happens naturally in schools. In order to improve the ecological validity of TI studies in natural environments, more descriptive research is required to measure what types of errors are made in schools. Further, it would be beneficial to know if different teaching strategies are susceptible to different types of errors.

There are numerous areas for future research. First, descriptive studies that simply measure how often errors of omission and commission occur during DTT in natural settings are essential to understand what current teacher practices look like. Once an understanding of how often teachers make each kind of error is reached, each type of error should be studied in several ways. Studying each type of error in isolation is important so that a basic concept of how each error affects students is clear. Then, looking at the different combinations of errors that occur (omission first, commission first, omission heavy, commission heavy, etc.) should provide deeper insight into how errors may harm student learning. This type of study will also further the preliminary work of St. Peter Pipkin et al. (2010) that indicates that initial high TI may be protective if TI errors of commission occur later. It is unknown if the same is true of errors of omission or a combination of errors. This area is ripe for study. It hoped that a shared language, empirical work regarding the dimensions of TI, as well as types of errors that take place in natural settings will lead to a better understanding of how to train teachers and help students learn.

REFERENCES

Alberto P.A.& Troutman, A. C. (2017). *Applied behavior analysis for teachers* (Interactive 9th ed.). Boston, MA Pearson.

Behavior Analyst Certification Board. (2012). *Guidelines: Health Plan Coverage of Applied Behavior Analysis Services for Autism Spectrum Disorders*. Retrieved from http://www.bacb.com/Downloadfiles/ABA_Guidelines_for_ASD.pdf.

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

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis*, 2nd ed. Upper Saddle River, N.J.: Pearson Prentice Hall.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18,* 423-445.

DiGennaro Reed, F. D., Reed, D. D., Baez, C. N., & Maguire, H. (2011). A parametric analysis of errors of commission during discrete-trial training. *Journal of Applied Behavior Analysis, 44*, 611-615.

Durlak, J. A, Dupree, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*:327–350.

Eikeseth, S. Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive behavioral treatment at school for 4- to 7-year-old children with autism: A 1-year comparison controlled study. *Behavior Modification, 26*, 49-68.

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

Fryling, M.J., Wallace, M. D., & Yassine, J. N. (2012). Impact of treatment integrity on intervention effectiveness. *Journal of Applied Behavior Analysis, 45,* 449-453.

Goldstein, H. (2002). Communication intervention for children with autism: A review of treatment efficacy. *Journal of Autism and Developmental Disorders, 32,* 373-396.

Green, A. E., Fettes, D. L., & Aarons, G. A. (2012). A concept mapping approach to guide and understand dissemination and implementation. *Journal of Behavioral Health Services & Research, 39,* 362-373.

Gresham, F. M. (2005). Treatment integrity and therapeutic change: Commentary on Perepletchikova and Kazdin. *Clinical Psychology: Science and Practice, 12*, 391-394.

Gresham, F. M., Gansle, K. A., Noell, G. H., Cohen, S., & Rosenlum, S. (1993). Treatment integrity of school-based behavior intervention studies: 1980-1990. *School Psychology, Review, 22,* 254-272.

Groskreutz, N.C., Groskreutz, M.P., & Higbee, T.S. (2011). Effects of varied levels of treatment integrity on appropriate toy manipulation of children with autism. *Research in Autism Spectrum Disorders, 5,* 1358-1369.

Grow, L. L., Carr, J. E., Gunby, K. V., Charania, S. M., Gonsalves, L., … Kisamore, A.N. (2009). Deviations from prescribed prompting procedures: Implications for treatment integrity. *Journal of Behavioral Education, 18,* 142-156.

Helmond, P., Overbeek, G., & Brugman, D. (2012). Program integrity and effectiveness of a cognitive behavioral intervention for incarcerated youth on cognitive distortions, social skills, and moral development. *Children and Youth Services Review, 34*, 1720-1728.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165-179.

Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities, 26*, 359-383.

Holcombe, A., Wolery, M., & Snyder, E. (1994). Effects of two levels of procedural fidelity with constant time delay on children's learning. *Journal of Behavioral Education, 4*, 49-73.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings*. New York, NY: Oxford University Press.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Leon, Y., Wilder, D., 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*,19-33.

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

McIntyre, Gresham, DiGennaro, & Reed. (2007). Treatment integrity of school-based interventions with children in the *Journal of Applied Behavior Analysis* 1991-2005. *Journal of Applied Behavior Analysis, 40,* 659-672.

Nelson, C. M., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *Journal of Behavioral Health Services & Research, 39*, 374-396.

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

Odluyrt, S., Tekin-Iftar, E., & Adalioglu, I. (2012). Does treatment integrity matter in promoting learning of children with developmental disabilities? *Topics in Early Childhood Special Education, 32,* 143-150.

Pas, E. T., & Bradshaw, C. P. (2012). Examining the association between implementation and outcomes: Statewide scale-up of school-wide positive behavior interventions and supports. *Journal of Behavioral Health Services & Research, 39*, 417-433.

Perepletchikova, F., & Kazdin, A. E. (2005). Treatment integrity and therapeutic change: Issues and research recommendations. *Clinical Psychology: Science and Practice, 12*, 365.

Power, T. J., Blom-Hoffman, J., Clarke, A. T., Riley-Tillman, T. C., Kelleher, C., & Manz, P. (2005). Reconceptualizing intervention integrity: A partnership-based framework for linking research with practice. *Psychology in the Schools, 42,* 495-507.

Sanetti, L. M. H., & Kratochwill, T. R. (2008). Treatment integrity assessment in the schools: An evaluation of the treatment integrity planning protocol. *School Psychology Quarterly, 24,* 24-35.

Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review, 38*, 460-475.

Smith, S. W., Daunic, A. P., & Taylor, G. G. (2007). Treatment fidelity in applied educational research: Expanding the adoption and application of measures to ensure evidence-based practice. *Education and Treatment of Children, 30*, 121-134.

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

Smith, T., Groen, A., & Wynn, J. W. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal on Mental Retardation, 105*, 269-285.

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

Wheeler, J. J., Baggett, B. A., Fox, J., & Blevins, L. (2006). Treatment integrity: A review of intervention studies conducted with children with autism. *Focus on Autism and Other Developmental Disabilities, 21*, 45-54.

Wheeler, J. J., Mayton, M. R., Carter, S. L., Chitiyo, M., Menendez, A. L., & Huang, A. (2009). An assessment of treatment integrity in behavior intervention studies conducted with

persons with mental retardation. *Education and Training in Developmental Disabilities, 44*, 187–195.

Wickstrom, K. F., Jones, K. M., LaFleur, L. H., & Witt, J. C. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 13*, 141-154.

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

Wilkinson, L. A. (2007). Assessing treatment integrity in behavioral consultation. *International Journal of Behavioral Consultation & Therapy, 3*, 420-432.

Vollmer, T., Roane, H., Ringdahl, J, & Marcus, B. (1999). Evaluating treatment challenges with differential reinforcement of alternative behaviors. *Journal of Applied Behavior Analysis, 32*, 9-23.

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

Zvoch, K. (2009). Treatment fidelity in multisite evaluation: A multilevel longitudinal examination of provider adherence status and change. *American Journal of Evaluation, 30*, 44-61.

APPENDICES

Appendix A

Discrete Trial Teaching

Student:_____ Trainer:_____

Objective:
Condition:
Sd:
Reinforcement Schedule: 1:1 80% EC 50% EC

Date		Date		Date		Date	
	Prompt		Prompt		Prompt		Prompt
10		10		10		10	
9		9		9		9	
8		8		8		8	
7		7		7		7	
6		6		6		6	
5		5		5		5	
4		4		4		4	
3		3		3		3	
2		2		2		2	
1		1		1		1	
%		%		%		%	
>*f* prompt: %		>*f* prompt: %		>*f* prompt: %		>*f* prompt: %	

Prompts: G = gestural

P = physical

Marking: ---- incorrect

❍ correct

☐ total correct for session

★If a prompt is used, the trial is marked incorrect.

Comments:

Appendix B TI/IOA Data Collection Sheet

Discrete Trial Teaching IOA/TI

Student:_____ Trainer:_____ 2nd Observer: _____

Objective:
Condition:
Sd:
Reinforcement Schedule: 1:1 80% EC 50% EC

Date			Date			Date			Date		
	Prompt	Correct R		Prompt	Correct R		Prompt	Correct R		Prompt	Correct R
10			10			10			10		
9			9			9			9		
8			8			8			8		
7			7			7			7		
6			6			6			6		
5			5			5			5		
4			4			4			4		
3			3			3			3		
2			2			2			2		
1			1			1			1		
	%				%			%			%

Prompts: G = gestural

P = physical

Marking: ---- incorrect

❍ correct

☐ total correct for session

★If a prompt is used, the trial is marked incorrect.

Comments:

Appendix B.1 Discrete Trial Teaching IOA/TI Example 80% Commission

Student: Hazel Trainer: Carina 2nd Observer: Jill B.

Objective: H. will identify the letters of the alphabet
Condition: Intervention
Sd: "Hazel, touch ***B."***
Reinforcement Schedule: 1:1 80% EC 50% EC

Date	8/22/13		Date			Date			Date		
	Prompt	Correct R		Prompt	Correct R		Prompt	Correct R		Prompt	Correct R
(10)		+	10			10			10		
(9)		+	9			9			9		
~~8~~	P	EC +	8			8			8		
(7)		+	7			7			7		
(6)		+	6			6			6		
~~5~~	G	+	5			5			5		
(4)		+	4			4			4		
~~3~~	G	EC+	3			3			3		
(2)		+	2			2			2		
(1)		+	1			1			1		
	70%	100%			%			%			%
IOA 100% TI 100%											

Prompts: G = gestural

P = physical

Comments: Trials 3 and 8 labeled *EC* + in the *Correct R* column, indicate the two errors of commission made by the researcher. The second observer recorded that the errors of commission were administered correctly.

Marking: ---- incorrect

❍ correct

☐ total correct for session

★If a prompt is used, the trial is marked incorrect.

Appendix C

Reinforcement Assessment for Individuals with Severe Disabilities (RAISD)

Student's Name: _____

Date: _____

Recorder: _____

The purpose of this structured interview is to get as much specific information as possible from the informants (e.g., teacher, parent, caregiver) as to what they believe would be useful reinforcers for the student. Therefore, this survey asks about categories of stimuli (e.g., visual, auditory, etc.). After the informant has generated a list of preferred stimuli, ask additional probe questions to get more specific information on the student's preferences and the stimulus conditions under which the object or activity is most preferred (e.g., What specific TV shows are his favorite? What does she do when she plays with a mirror? Does she prefer to do this alone or with another person?)

We would like to get some information on _____'s preferences for different items and activities.

1. Some children really enjoy looking at things such as a mirror, bright lights, shiny objects, spinning objects, TV, etc. What are the things you think _____ most likes to watch?

   _____

   Response(s) to probe questions:

   _____

2. Some children really enjoy different sounds such as listening to music, car sounds, whistles, beeps, sirens, clapping, people singing, etc. What are the things you think _____ most likes to listen to?

   _____

   Response(s) to probe questions:

   _____

3. Some children really enjoy different smells such as perfume, flowers, coffee, pine trees, etc. What are the things you think _____ most likes to smell?

Response(s) to probe questions:

4. Some children really enjoy certain food or snacks such as ice cream, pizza, juice, graham crackers, McDonald's hamburgers, etc. What are the things you think _____ most likes to eat?

Response(s) to probe questions:

5. Some children really enjoy physical play or movement such as being tickled, wrestling, running, dancing, swinging, being pulled on a scooter board, etc. What activities like this do you think _____ most enjoys?

Response(s) to probe questions:

6. Some children really enjoy touching things of different temperatures, cold things like snow or an ice pack, or warm things like a hand warmer or a cup containing hot tea or coffee. What activities like this do you think _____ most enjoys?

Response(s) to probe questions:

7. Some children really enjoy feeling different sensations such as splashing water in a sink, a vibrator against the skin, or the feel of air blown on the face from a fan. What activities like this do you think _____ most enjoys?

Response(s) to probe questions:

8. Some children really enjoy it when others give them attention such as a hug, a pat on the back, clapping, saying "Good job", etc. What forms of attention do you think _____ most enjoys?

Response(s) to probe questions:

9. Some children really enjoy certain toys or objects such as puzzles, toy cars, balloons, comic books, flashlight, bubbles, etc. What are _____'s favorite toys or objects?

Response(s) to probe questions:

10. What are some other items or activities that _____ really enjoys?

Response(s) to probe questions:

After completion of the survey, select all the stimuli which could be presented or withdrawn contingent on target behaviors during a session or classroom activity (e.g., a toy could be presented or withdrawn, a walk in the park could not). Write down all of the specific information about each selected stimulus on a 3" x 5" index card (e.g., likes a female adult to read him the 'Three Little Pigs' story.) Then have the informant(s) select the 16 stimuli and rank order them using the cards. Finally, list the ranked stimuli below.

1. _____
2. _____
3. _____
4. _____
5. _____
9. _____
10. _____
11. _____
12. _____
13. _____

6. _____
7. _____
8. _____
14. _____
15. _____
16. _____

Notes:

Appendix D
Social Validity Form

This is an example of the survey that teachers will see online.

For each of the following questions, please indicate your level of agreement by clicking on a number.

1= strongly disagree
2=disagree
3=agree
4= strongly agree

1. Allowing my students to take part in this study was useful to ME.

1 2 3 4

2. Taking part in this study was useful to MY STUDENTS.

1 2 3 4

3. Having researcher(s) in the room was intrusive.

1 2 3 4

4. This study took an unacceptable amount of MY STUDENTS' time.

1 2 3 4

5. The feedback I received at the end of the study was useful.

1 2 3 4

6. The feedback I received will affect my teaching.

1 2 3 4

Please feel free to add any other comments or questions you have.