7-10-2020

Generalizability of Multiple Measures of Treatment Integrity: Response Card Intervention

Elizabeth Kelsey Wilson

GENERALIZABILITY OF MULTIPLE MEASURES OF TREATMENT INTEGRITY: RESPONSE CARD INTERVENTION

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Psychology

by
Elizabeth Kelsey Wilson
B.A., Louisiana State University, 2012
M.S., University of Texas at Dallas, 2014
M.A., Louisiana State University, 2017
August 2020

TABLE OF CONTENTS

Abstract

Treatment integrity is essential for the implementation of interventions in schools as it determines the accuracy or consistency with which different components of a treatment are implemented. There are no current standards regarding the best practices in treatment integrity measurement; however, higher treatment integrity is associated with enhanced student outcomes. At present, there is no database that provides information related to treatment integrity for practitioners, researchers, and policy-makers to reference when choosing an appropriate level of treatment integrity needed for a given intervention. Consequently, there is a need to establish convergent validity among different methods of treatment integrity measurement using multiple evidence-based interventions in order to guide best practices. The current study attempted to replicate and expand the finding that the direct observation method yields the most reliable treatment integrity data when using an evidence-based intervention (Gresham, Dart, & Collins, 2017). For this study, researchers empirically replicated the methods of Gresham and colleagues' work to examine two of the three measures of treatment integrity, direct observation and self-report, for six teachers' implementation of the response card intervention.

Introduction

Treatment integrity measurement is a necessity within implementation science, as it is a method of assessing the accuracy with which an intervention is implemented (Gresham, 1989). Treatment integrity is not an idea that is unique to the field of psychology. The concept of treatment integrity is critical in all fields that provide intervention or treatment to individuals. Different terms are used to explain integrity within the field of psychology. In particular, school psychology refers to this concept as "treatment integrity" or "implementation plan integrity" (Gresham, 1989). Within the consultation framework, the accuracy of implementation of the independent variable, or treatment, in an experimental study is called "treatment integrity." The implementation of a treatment or intervention plan in a natural context is referred to as "intervention plan implementation." Lastly, the extent to which consultation procedures, whether it be the overarching consultation model, or the procedures embedded within it, are implemented as designed is referred to as "consultation procedural integrity" (Noell, 2008). Although there are various names of the concept, they all share a similar goal: to ensure that desired changes in an individual's functioning or behavior are due to the systematic and accurate implementation of treatment and not extraneous variables (Gresham, 2009). With information collected from measures of integrity, scientists can ensure that the changes in the dependent variable are attributable to the intervention and not to outside variables.

Treatment integrity is thought of as multidimensional and researchers often conceptualize the multiple aspects in different ways, sometimes overlapping in their conceptual models (Gresham, 2009). The primary focus of treatment integrity has traditionally been adherence, which is appropriate for simple interventions; however, for more complex interventions other aspects such as implementer competence or participant responsiveness could have an effect on

treatment outcomes (Schulte, Easton, & Parker, 2009). Dane & Schneider (1998) hypothesized four additional dimensions of treatment integrity that should be considered during intervention implementation. These are exposure/dosage, quality/skill, intervention differentiation, and participant responsiveness.

Adherence, or fidelity, is how well the essential intervention components are implemented as prescribed (e.g., percentage of intervention components of the response card intervention actually implemented by the teacher during the intervention period). Exposure, or dosage, is the index of how much the original intervention is delivered (i.e., number, length, and or frequency of sessions). An example of dosage in regard to the response card intervention would be the duration the intervention is implemented, recorded in number of minutes. Quality, or competence of the implementer, is the level of skill with which, or how well various intervention components have been delivered (e.g., expert ratings of the quality of the teacher delivery of the intervention). Intervention differentiation is the extent to which an intervention's theory and practice can be distinguished from other interventions. In other words, when implementing a new intervention, the change should be distinct enough from what is already happening in order to make a difference in the client's behavior (e.g., comparison between tier 1 and tier 2 reading intervention outcomes). Lastly, participant responsiveness is the level of engagement and enthusiasm the client displays during the intervention (e.g., ratings of student engagement or participation) (Gresham, 2009; Sanetti & Kratochwill, 2009).

In practice, a functional relationship is assumed between the delivery of treatment, most often done by third parties such as teachers and parents, and the outcome of the intervention; however, the factors mentioned above can change the intervention and the way it is implemented in ways unknown to the researcher. This can lead to researchers and practitioners making

assumptions that the treatment is working for a given behavioral concern, when in reality there could be extraneous variables that are affecting the measureable outcomes. This can also have the reverse effect, in which a treatment appears to fail to produce the desired outcome, but the failure to evoke the desired change is due to outside factors (Gresham, 2009). Treatment integrity is not synonymous with intervention effectiveness, which is defined as how strong the treatment is in producing the desired outcome. If a treatment is implemented with low integrity and does not produce the desired changes, it could be that it was implemented differently than originally planned, resulting in an intervention that is not representative of what the researchers intended (Perepletchicova, Hilt, Chereji, & Kazdin, 2009).

Treatment integrity has been conceptualized as the degree to which a treatment is implemented in the way that it was intended. That being said, not all treatments are as resistant to outside variables at the same level. Some treatments are stronger than others and are more resistant to poor implementation, therefore these treatments would rely on a high level of treatment integrity to be effective. Whereas weaker treatments can be implemented flawlessly and still have minimal effects on the variable of interest (Gresham, 2009). Thus, different treatments require different levels of treatment integrity in order for the treatment to produce significant changes. Gresham (2009) also raised the issue that certain components or steps of an intervention may be crucial to behavior change, whereas other components of the same intervention may not be as important in driving change.

Treatment Integrity and Policy

In recent years, there has been more interest within the field of school psychology regarding the concept of treatment integrity and its importance in intervention implementation; however, there has been little attention given to the empirical measurement of treatment

integrity. The Institute of Educational Sciences (IES) within the U.S. Department of Education requires that for every intervention, there should be a measure of treatment integrity, but it does not specify or provide any direction as to what levels of treatment integrity are required for specific interventions. IES is moving in the right direction by addressing and requiring treatment integrity, but they do not provide any guidelines as to best practices for measuring treatment integrity. There is no guidance for practitioners regarding which methods of treatment integrity measurement should be used for specific interventions or how many assessments of integrity are needed over what period of time.

One reason for the lack of information regarding these critical issues in treatment integrity is that there is no database that provides information related to treatment integrity for practitioners, researchers, and policy-makers to reference when choosing an appropriate level of treatment integrity needed for a given intervention. There are no current standards regarding the best practices in treatment integrity measurement. There is a shortage of investigation into how much treatment integrity impacts different treatment outcomes, as well as a lack of treatment integrity measurement in practice and research. This has led to a shortage of information, thus no database. As discussed previously, different treatments for different problems may require different levels of treatment integrity to yield the desired change in the dependent variable. Researchers know the "true value," or the accurate measurement of behavior, of an intervention before implementation because, in best practice, they have standardized the individual steps, or treatment components, that interventionists should take when implementing the intervention. They have operationally defined each component of an intervention so that practitioners and researchers will be able to replicate them in a natural setting. However, different methods of treatment integrity measurement may provide researchers with different reported levels or

degrees of treatment integrity, based on how reliable these different measurements are at producing an accurate or "true value" of behavior (Gresham, Dart, & Collins, 2017). There is little evidence to suggest that different components, or steps, of any one intervention should correlate with each other. For example, one component could be accounting for a large percentage of variance of the outcome variable, or target behavior to be changed, whereas another component may not be accounting for very much variance at all (Gresham, 2009).

Barriers to Treatment Implementation and Integrity

In previous research, higher levels of treatment integrity were correlated with better intervention outcomes (Durlak & DuPre, 2008; Fiske, 2008; Sanetti & Kratochwill, 2009). Most of the consultation research has focused on student outcomes, while assuming that treatments were being implemented with integrity, when it is now argued that treatment integrity and changing the behavior of adults is the main focus of the consultation framework that leads to better student outcomes (Noell & Gansle, 2014). School psychologists are having to train and ultimately change the behavior of teachers who have demanding schedules and have the ability directly or indirectly to refuse to participate or implement the intervention correctly (Noell, Gansle, Mevers, Knox, Mintz, & Dahir, 2014).

Accurate and reliable intervention plan implementation leads to accurate data-based decision-making during the Response to Intervention (RTI) process in schools (McIntyre et al., 2007; Roach & Elliot, 2008; Sanetti, Fallon, & Collier-Meek, 2011). High treatment integrity is vital in behavioral interventions implemented in schools because the interventions can lead to the desired behavior change in students only if the intervention is implemented correctly. An array of factors has been identified that can have a negative effect on the accuracy of intervention implementation and the resulting behavior change. These factors, or barriers, are teacher

characteristics, time and resources, complexity of intervention, and number of steps, or degree of complexity, involved in the implementation (McIntyre, Gresham, DiGennaro, & Reed, 2007).

Intervention implementation barriers are defined as factors that hinder the progress of an intervention and lessen the effect the intervention components have on student outcomes. Implementation barriers can be either external variables that obstruct implementation or needed resources that are absent from the intervention (Durlak & Dupre, 2008; Forman et al., 2009; Klingner, Ahwee, Pilonieta, & Menendez, 2003; Long et al., 2016). For example, a barrier for a teacher could be lack of resources in a classroom when trying to implement an intervention such as the response card intervention that requires each student to have his or her own marker. Another example could be a teacher that is trying to implement an individualized reading intervention with a student in a quiet space for thirty minutes per day. If the child is absent for a majority of the days that the intervention was intended to be implemented, this limits the ability to make an informed decision related to the effectiveness of the reading intervention.

Research has recommended the grouping of important barriers into categories that mirror an ecological framework. The levels included within this ecological framework are the macro, organization, and the individual. Sanetti and Kratochwill (2009) conducted a systematic review of the literature and organized barriers into four broad categories based on the research of Feldstein and Glasgow (2008). The four categories include the external environment, organization, intervention, and implementer (Long et al., 2016). The external environment includes the larger context such as educational policy and legislation, district support, and collaboration between the school and outside organizations, as well as the way these factors influence funding, training, and the resources available (Bosworth, Gingiss, Potthoff, & Roberts-Gray, 1999; Long et al., 2016). The organization category includes administration support,

atmosphere, and resources of the school in which the intervention is implemented (Durlak & DuPre, 2008; Kam, Greenberg, and Walls, 2003; Long et al., 2016). The intervention category centers around complexity of intervention components and compatibility of the intervention with the implementer and the environment in which it is implemented (Fixsen et al., 2005; Gresham, 1989; Long et al., 2016). Lastly, implementer barriers focus on individual characteristics of the implementer and how those relate to the implementer's buy-in and skill level (Bosworth et al., 2005; Long et al., 2016).

Maximizing Treatment Integrity

Intervention implementation occurs in a natural setting, such as a school, where unanticipated events occur and competing contingencies are present. For instance, teachers are regularly required to implement two or more behavior plans at one time (Long et al., 2016) and are often not given the support they need. Therefore, promoting treatment integrity and supporting teachers in their implementation of interventions in schools is best practice and crucial to student outcomes.

One way to maximize treatment integrity is to increase social acceptability, which is defined in the school setting as the teacher's perception of intervention acceptability, effectiveness, feasibility, and amount of time the intervention will take to produce the desired behavior change in the student (Curtis, Pisecco, Hamilton, & Moore, 2006). Kratochwill and Pittman (2002) found that one way to increase teachers' acceptability of classroom behavioral interventions is to use a more collaborative consultation model such as the problem-solving model or the conjoint behavioral consultation model. Kratochwill (2008) emphasizes the importance of the two major goals of the consultation framework, which are to improve client outcomes as well as impart skills to the consultee that address the current concern and similar

future concerns. In order to change the behavior of the consultee it is beneficial to use proactive support strategies such as direct training methods, detailed intervention implementation planning, and giving intervention choice, to address any barriers teachers might come across involving intervention implementation rather than waiting for them to fail.

Implementation planning is a component of the Planning Realistic Implementation and Maintenance by Educators (PRIME) model (Sanetti et al., 2015; Sanetti et al., 2014). Teacher implementation of any school-based intervention can be viewed as a form of adult behavior change as the teachers are applying new practices in the classroom in order to change student behavior – increasing student academic engagement, in the case of the present study. Sanetti et al. (2014) developed the PRIME model based on the Health Action Process Approach (HAPA), a theoretical model of adult health behavior change from health psychology (Schwarzer, 2008).

According to the HAPA model, the process of adopting and maintaining a new behavior occurs through, first, a motivation phase and then a commitment phase. During the motivation phase, a teacher develops a goal to adopt a new behavior or modify a pre-existing one. To develop a goal, the model proposes that a teacher must believe (a) there is a need for behavior change, (b) that the outcomes of the behavior change will be beneficial, and (c) that the behavior change is attainable (i.e., self-efficacy). After the development of a goal, the teacher adopts the new behavior and maintains this behavior change in the commitment phase. The commitment phase begins with action and coping planning, which the HAPA model posits as the fundamental process that moves the teacher from the motivation phase to acquiring successful behavior change. Action planning focuses on identifying and planning the logistics (e.g., when, where, duration, materials) of how the behavior change will be implemented within the target context. Next is coping planning, in which the teacher identifies potential barriers to behavior change and

the corresponding strategies to resolve those barriers. The use of action and coping planning has been shown to result in resilient behavior change for various health behaviors (Schwarzer, 2008) and findings suggest that the use of these activities together is more effective than the application of either one in isolation (Lippke, Wiedemann, Ziegelmann, Reuter, & Schwarzer, 2009).

The PRIME model adapts these HAPA processes for use in education. Unlike health behavior change, where the beneficiary is the person making the change, behavior change by the teachers is beneficial to both the students (e.g., increased student academic engagement and academic performance) and the teachers (e.g., reduced time addressing disruptive behavior).

Under the PRIME model, action and coping planning are combined into one component, called implementation planning, and the implementation support process occurs over a single collaborative consultation meeting between the consultant and teacher. The goal of implementation planning is to proactively promote the treatment integrity of the intervention that the teacher is being trained to implement in the classroom. During the action planning portion of implementation planning, the consultant reviews, modifies if needed, and records the intervention strategy with the teacher. This process starts by first reviewing the standardized intervention procedures and components with the teacher, then the consultant and teacher collaborate to identify any modifications that need to be made to create a better "fit" for the teacher as well as any barriers that he or she can foresee happening in her unique classroom environment. Finally, the logistical steps of the intervention are planned (e.g., when, where, duration, materials needed). After establishing the intervention procedure, the teacher then considers potential barriers during intervention implementation and plans actions to resolve or bypass these barriers. At the end of planning meeting, the consultant provides the teacher with a summary of the session.

Sanetti, Williamson, Long, and Kratochwill (2017) found that implementation planning resulted in increased treatment integrity and quality of individual and class wide behavior intervention plans. Additionally, teachers who received implementation planning rated their consultants as effective, provided higher treatment acceptability ratings, and rated implementation planning as feasible within their schools (Sanetti et al., 2015). The present study will be discussing with each teacher during the collaborative consultation and training meeting, the possible barriers that they foresee happening with the response card intervention in their classrooms. Researchers and teachers will collaboratively action plan to resolve any barriers mentioned.

Another proactive strategy that can promote treatment integrity is goal setting. Previous research on goal setting has shown it to be effective for increasing treatment integrity when paired with performance feedback (Duncan, Dufrene, Sterling, & Tingstrom, 2013; Martens, Hiralall, & Bradley, 1997). The mere act of setting challenging personal goals has been consistently shown to lead to higher levels of task performance when applied as a strategy on its own (Locke & Latham, 2006). The addition of goal setting to implementation planning may further enhance the effectiveness of the procedure as a proactive implementation support strategy.

Fallon et al (2017) showed through meta-analysis of 33 studies that direct training of intervention implementers, specifically instructions, modeling, practice, and feedback, is an effective, evidence-based practice that results in higher levels of treatment integrity. Direct training is usually conducted under the context of a consultative relationship, in which instructions and modeling are provided during the process of skill acquisition, whereas practice and feedback is meant to aid in the implementer gaining fluency in the implementation of a

targeted intervention. In the context of school psychology consultation, either an oral or written description of the intervention is given to the teacher. This description includes specific components, rationale and theory related to the intervention, implementation procedures, and how the intervention addresses specific goals and student outcomes. Demonstrating, or modeling, the intervention to the teacher can occur in a training setting, such as an office or meeting space, or in the classroom with students present or not present. To generalize the skills, modeling should first occur in a training setting and then in the classroom (Fox, Hemmeter, Snyder, Binder, & Clarke, 2011). Research has also shown video modeling to be effective as it allows for standardization of training across participants, efficient training methods, and the opportunity for teachers to review the training video in the absence of the consultant or research team (DiGennaro-Reed, Codding, Catania, & Maguire, 2010). The practice phase provides the teacher with the opportunity to rehearse the intervention in front of the consultant to demonstrate adequate knowledge and ensure accuracy. After the intervention has been implemented, the teacher will receive feedback about his or her implementation from the consultant. This feedback should include positive aspects, such as praise for components of the intervention implemented correctly. Feedback can also be corrective, in which the consultant reviews components that the teacher implemented incorrectly and gives guidance on how to improve implementation in the future (Fallon et al., 2017).

Feedback given to teachers after intervention implementation is a reactive strategy used to promote treatment integrity by providing systematic consultative support commonly known as performance feedback (Long et al., 2016). To date, performance feedback is the only school-based implementation support strategy that has a consistent line of research establishing its efficacy and core components (Fallon, Collier-Meek, Sanetti, Feinberg, & Kratochwill, 2016;

Noell & Gansle, 2014). Performance feedback typically involves one-on-one support within a consultative relationship in which the consultant monitors the consultee's intervention-related behavior and provides specific feedback regarding the accuracy and quality of implementation performance (Noell et al., 2005). Feedback can be provided orally, in writing, or graphically to the teachers regarding their ability to implement the intervention reliably and accurately as a means of improving and maintaining positive changes in their behavior (Noell, 2010). Two factors shown to enhance performance feedback are goal setting with the teacher as well as providing feedback daily, rather than weekly or biweekly. Performance feedback has been shown to be an effective strategy in its own right, increasing treatment integrity in school consultation for academic, behavioral, student-targeted, and class wide interventions (Long et al., 2016).

In addition consequence strategies, such as performance feedback, there are many proactive strategies that have been preliminarily proven to be effective, such as giving a choice of two or more interventions to implement and allowing teachers to select which would fit best in their classroom (Dart, Cook, Collins, Gresham, & Chenier, 2012), using a collaborative consultation model (Kelleher, Riley-Tillman, & Power, 2008), and meeting with teachers to provide and carry out detailed intervention planning that addresses potential barriers that may arise (Sanetti, Kratochwill, & Long, 2013; Sanetti, Collier-Meek, Kim, & Kratochwill, 2014).

Eckert and Hintze (2000) have argued extensively that treatment acceptability, defined as the degree to which interventions are seen as practical, feasible, and appropriate to consumers (i.e., teachers, parents, students, administration) for addressing concerns, should be related to higher levels of treatment integrity. However, research has more recently shown that treatment acceptability may not be significantly related to intervention implementation. Noell and Gansle (2014) found that the reason for the lack of relation between the two could be due to the fact that

treatment acceptability regarding psychological treatments for children, and academic and behavioral interventions for students, is so high across the board that there is no variability in the data. It has been found that teachers will engage in implementation of interventions if enough supports are in place (e.g., accountability to parents or administration, adequate resources and training), even if they have judged the intervention to be especially challenging (Noell, Witt, LaFleur, Mortenson, Ranier, & Levelle, 2000). Although teachers will engage in implementation when they are not fully accepting of the intervention, future research may still be needed to determine if this acceptability rating could have an effect on treatment integrity levels.

Measurement of Treatment Integrity

When deciding how to measure treatment integrity, researchers must first think about the intervention selected and the reason it is being used (e.g., high stakes decisions, behavioral modification at a universal level, increasing academic performance for a tier 3 evaluation). The specificity in terms of intensity of intervention and types of decisions being made based on the outcome of that intervention should guide the selection of a treatment integrity measurement method that matches the situation. For example, a tier 1, universal intervention for a classroom of students would not have the same specificity as a tier 3, targeted intervention for one student (Noell, 2008).

Treatment integrity can be captured through various direct (e.g., systematic observation using an observer in the classroom, videotaping, audiotaping, and computer software) and indirect (e.g., permanent products, self-report, rating scales, interviews, checklists, and lesson plan reviews) approaches. The researchers will be examining two of the most common methods of collecting treatment integrity, which are self-report and direct observation (Fiske, 2008). Both self-report and direct observation methods can be constructed so that treatment integrity is

measured by specifying intervention components in standard, observable, and absolute terms, creating a components checklist, and computing the percentage components that the teachers accurately complete (Gresham, 2009).

Self-report methods rely on the agents of change to report on their completion of the components of the intervention. Self-report methods can measure multiple dimensions of treatment integrity, are resource-efficient, and can be completed during or immediately after an intervention has been completed (Noell et al., 2014). Just as direct training has been shown to increase levels of treatment integrity, direct training on how to use and complete self-report forms should be utilized to increase accuracy and reliability (Fallon et al., 2017).

Direct observation is the most commonly used method, in which, typically, a consultant will observe the teacher during the intervention utilizing a components checklist. After the observation, the consultant will calculate the percentage of components completed (Noell, Witt, Slider, & Connell, 2005). Direct observations can be utilized for most interventions used in schools and can readily gather multiple dimensions of treatment integrity. It is the most direct assessment of treatment integrity, requiring the least amount of inference, of all of the methods, and most strongly related to teacher and student outcomes (Noell et al., 2014).

Even though these methods are frequently relied upon to reflect accurate measures of treatment integrity, studies have shown that these assessments do not possess strong psychometric qualities (Sanetti & Kratochwill, 2009). For example, direct observations have been shown to cause reactivity in intervention agents (Sheridan et al., 2009) and may not reflect accurate treatment integrity over time due to observer drift, which is an unintended change in the accuracy of the observer's performance (Gresham et al., 2009). Direct observation also is high in resource demand in that it is time intensive and may require researchers to recruit extra research

assistants to help with data collection. It may not be representative of actual teacher behavior if the observer is not able to observer the entire intervention or at least an adequate sampling of the implementation (Noell et al., 2014).

In addition, self-reports have been shown to display a severe upward bias, in which teachers will report themselves as completing the intervention or parts of the intervention correctly when in reality they did not. This could be due to social desirability and memory issues, leading teachers to overestimate their implementation levels, as well as ignorance or lack of understanding of intervention components. Another problem with self-report measures is that intervention implementers do not always consistently complete the measures (Noell et al., 2014).

Whether treatment integrity is being measured using direct observation or self-report, a treatment integrity components checklist is often utilized so that implementation data can be easily collected. Components checklists list the operationally defined steps that are judged to be critical components of the intervention, so that treatment integrity can be measured in the same manner across observations and measurement methods (Hineline, 2003; Kelleher et al., 2008). Researchers know the accuracy with which the steps of an intervention should be implemented, as they have standardized these individual steps by operationally defining each component so that intervention is able to be replicated in a natural setting.

Lambert et al. (2006) standardized the components of the response card intervention. Those components are listed as follows: (a) Teacher presents questions to the class, (b) teacher provides adequate wait time for students to use response cards, (c) teacher requests students to present their cards, (d) if more than one fourth (i.e., four or five students) of the class makes an incorrect response, the teacher instructs students to fix their answers to the problem and provides

rationale for the answer; if fewer than one fourth of the students make an incorrect response, the teacher reveals answer to class, (e) teacher provides praise for correct responses.

The steps have been modified slightly for clarity and simplicity, as well as to fit the present study. The seven components of the present study are as follows: (a) teacher states to the class that it is time to respond using response cards, (b) teacher reviews the rules of the intervention with the class (e.g., ten seconds to answer the question, do not raise your cards until directed, do not talk to your peers, answer the question by yourself, markers only write on response cards, caps go on the back of markers, etc.), (c) teacher presents a question to the class, (d) teacher provides adequate time for students to write their answers on the response cards, (e) teacher requests that students present their cards, (f) if more than five students provide incorrect answers, the teacher instructs the class to fix their answers and provide a rationale for the answer; if not, the teacher reveals the correct answer to the class, (g) the teacher provides verbal praise for correct responses.

Response Card Intervention

Treatment integrity is needed to assess the accuracy with which all types of interventions are implemented, including academic interventions in general education classrooms. The present study will be assessing different methods of treatment integrity measurement when general education teachers implement the response card intervention. Response cards have been shown to be effective for a wide range of ages. Although it has generally been used at the primary and secondary education levels, this intervention has been used in as young as preschool age students to teach basic subjects such as colors and calendar math (Godfrey et al., 2003; Christle & Schuster, 2003), as well as in university settings to increase academic engagement in courses such as introduction to psychology and research methods (Shabani & Carr, 2004). The response

card intervention has also been shown to increase academic engagement, student participation in whole class activities, on-task behavior, and academic achievement when compared to usual strategies used by teachers as an attempt to increase these factors (i.e., hand raising, cold-calling, etc.) (Christle & Schuster, 2003).

Research has shown that students who make more responses during a lesson learn more than students who make fewer responses (Heward, 2000; Christle & Schuster, 2003), yet active learning methods are not widely used in general education classrooms. Typically, the high-achieving students will raise their hands and answer questions posed by the teacher to the whole class, while lower achieving students do not raise their hands as often, consequently leaving them with less opportunities to respond and actively participate during the lesson (Heward, 2000; Christle & Schuster, 2003). Active student participation in instruction is a result of a deliberate attempt by the teacher to elicit students' overt participation (Heward, 2000). These behaviors contrast to passive responses, such as the teacher simply giving the answer to the group or one student answering, usually the high-achieving students as previously mentioned, while the other students listen.

Christle and Schuster (2003) found that using the response card intervention increased students' academic performance on weekly quizzes as well as increased on-task time compared to the hand-raising condition. This is beneficial as this can help researchers recruit teachers for the present study, since they will be trained on a class wide intervention that will increase academic performance in their classrooms.

The response card intervention has also been shown to reduce disruptive behavior of students (Billingsley, 2001; Billingsley & Tomchin, 1992; Darling- Hammond, 2003; Fimian & Santoro, 1983; Lambert, Cartledge, Heward, & Lo, 2006). Teachers spend a significant amount

of time redirecting inappropriate behavior in the classroom, which has a negative impact on teacher satisfaction as well as the disruptive students' academic performance. Engelmann and Colvin (1983) found that through careful planning and intervention implementation, instead of using time redirecting and reacting to this inappropriate behavior, teachers can eliminate as much as 90% of this disruptive behavior by employing effective, academically engaging instruction.

Response cards are appealing to low achieving students, who are unlikely to be motivated to participate by the academic content alone. Response cards involve the entire class in rapid, and energetic responding, in which the teacher can quickly scan each students' answer for accuracy. This allows the hesitant learners to participate without feeling embarrassed or singled out (Lambert et al., 2006). Lambert et al. (2006) examined the relationship between the use of response cards and the decrease in disruptive behavior and the increase in academic responding of students in fourth grade classrooms. They found positive results in that all students using response cards showed an increase in on task behavior, correct responses, less disruptive behavior, and increase in participation. When students are forced to attend to instruction by using response cards, they are less distracted by external stimuli and have less time to disrupt the learning of others. Further, the response card intervention allows teachers to obtain observable responses from students, thus allowing them to provide immediate feedback on student performance which enhances student learning (Lambert et al., 2006).

Academic Engagement Assessment

Observers in the present study rated the level of academic engagement of students during a baseline period as well as during the intervention periods when teachers implemented the response card intervention, in order to monitor progress and ultimately show improvement in academic engagement from baseline to response card intervention implementation. The

assessment of student behavior in the context of RTI is challenging because traditional behavior rating scales were not designed for the frequent use that is needed to monitor the progress of an intervention or to monitor the behavior of large groups of students at once, such as academic engagement in a classroom of students. Throughout the literature much attention has been given to the psychometric properties of two alternate assessment options for measuring academic engagement, Direct Behavior Ratings (DBR) and Systematic Direct Observations (SDO).

Though the feasibility, repeatability, and familiarity of DBRs suggest high potential for these types of measures (Riley-Tillman et al., 2008), there is concern regarding the influence of rater error. Riley-Tillman, et al. (2009) uses the term *rater error* to describe cases where ratings tend to either over- or underestimate the true score of the target behaviors being observed via a DBR measure. Briesch et al.'s (2010) results suggested DBR rating variances were mainly accounted for by rater-related effects. This is concerning since behavioral assessments via observation should contain reliable scores by the raters. Even more so, The Individuals with Disabilities Education Improvement Act of 2004 (IDEIA; 2004) requires information regarding student behavior be gathered via tools that are technically sound (i.e., reliable and valid), since high stakes decisions regarding children's education placements are often partially determined by assessment results.

SDOs are distinguished by the five following characteristics: (a) the goal of observation is to measure specific behaviors, (b) the behaviors being observed have previously been precisely and operationally defined, (c) observations are conducted under standardized procedures and are highly objective, (d) the logistics (e.g., time, place, people) of the observation are specific and carefully chosen, (e) and the scoring and reporting of data are standardized and do not differ from one observer to another. SDOs have been shown to provide a very direct estimate of

behavior; however, they come with practical limitations as they are very time and resource intensive. Findings from Briesch et al. (2010) supported the idea that a large proportion of the rating variance in the SDO used comes from changes in student behavior, whereas the greatest proportion of DBR rating variance is due to rater-related effects. Therefore, in the present study, academic engagement was assessed during baseline and intervention periods using momentary time sampling recording via a systematic direct observation method of data collection.

Briesch et al. (2010) found that the SDO method yielded more dependable estimates of student engagement than DBR when using a 15-second momentary time sampling recording procedure. The present study used the same SDO methods as in the Briesch et al (2010) study, as well as utilized a 20-minute observation duration for assessing student academic engagement during baseline and intervention periods. Findings from Ferguson et al. (2012) have shown that over the course of one day, a 15-minute observation was found adequate to yield reliable data for making low stakes decisions.

Although SDOs have lower variance due to rater-related effects than DBRs, observers still need to be sufficiently trained in order to conduct an SDO successfully. This study employed similar training techniques to that of Briesch et al. (2010). Graduate research assistants participated in a training session in which they will be shown a video of recorded classroom situations and asked to record instances of student academic engagement using a momentary time sampling recording method. They then participated in a full school week of in vivo training.

Generalizability Theory

The approach of classical test theory (CTT) is typically used for analyzing the degree of accuracy with which measurements are conducted, as well as the overall strength or weakness of a measure (Briesch, et al., 2010). CTT does not yield any valuable information regarding

methods for reducing error or strengthening the measurement (Briesch, et al., 2010). However, the Generalizability Theory (G Theory) not only makes it possible to note that measurement error exists, indicating the overall strength and weakness of the measure, but it provides information regarding how to improve a measurement in the future by pinpointing exactly which facets are contributing to the variance in the outcome (Briesch et al., 2010). G Theory is a statistical framework that is used to evaluate the dependability or reliability of behavioral measurements (Cronbach, Gleser, Nanda, & Rajaratnam, 1972). G Theory has been supported as an alternate and more comprehensive approach for analyzing the psychometric properties of assessments, specifically treatment integrity measurement methods (Gresham, 2009). Unlike CTT, G theory allows for multiple sources of error variance concurrently analyzed by partitioning them out. The separation of each source of error variance allows researchers to identify specific sources of measurement error that are of concern, and then evaluate the relative effect each of these sources has on the dependent variable (Hintze & Matthews, 2004). Although G Theory focuses on sources of variability that contribute to error, it also produces a generalizability coefficient, similar to CTT's reliability coefficient, that shows how accurate the "generalization score" is from what is observed in the sample of behavi.or to what is the "true score."

Just as an ANOVA partitions an individual's score into the effects for independent variables, their interactions, and error, G theory partitions that same score into an effect for the object of measurement (i.e., the variable the researcher predicts will account for most variance in the model), an effect for each facet or source of error, and an effect for each of their interactions. After performing a G study, a researcher will then perform a Decision (D) study, in which information is used from the G study to design or improve a measurement that minimizes error

with a specific purpose in mind, often based on one of the facets used in the study (Shavelson & Webb, 1991).

Current Literature

Convergent validity is defined as when two measures that are supposed to be measuring the same construct are shown to be related (Gresham, 2009). Convergent validity among different measurement methods for treatment integrity needs to be established across multiple evidence-based interventions in order to build a reliable database. Gresham (2009) proposes the idea of establishing a "treatment integrity effect norms" database, in which multiple researchers across multiple sites work to quantify "what levels of treatment integrity, measured by what methods, with what intervention procedures, produces what level of treatment integrity outcomes" (pp. 5). Gresham et al. (2017) is the first study to examine the dependability of different methods of treatment integrity of a widely known evidence-based intervention, the Good Behavior Game, using Generalizability Theory methods. The researchers sought to begin the first of a series of studies that would assess multiple methods of treatment integrity assessment across multiple evidence-based interventions.

Gresham and colleagues assessed treatment integrity twice per week and used a .80 reliability criterion. They found that although all three methods were reliable, direct observation provided the most reliable measure of treatment integrity (i.e., 4 observations yielded reliable data), followed by permanent products (i.e., 5 observations provided a reliable measure of treatment integrity), and then self-reports (i.e., 7 observations yielded reliable data). The researchers also found low correlations between the three measures, with significantly lower correlations between self-report and the other two methods.

The Present Study

The previous study is the first to apply Generalizability Theory to the assessment of treatment integrity and enhances the existing research that the direct observation method is the standard in treatment integrity assessment when examining behavior in schools. The results are critical to the field as they guide practitioners as to which method to use when making important decisions in schools, and they are the first of many studies that will add to a database of treatment integrity effectiveness norms. The goal of the present study is to add to this database by replicating the findings of Gresham et al. (2017), by extending the methods to another well-established, academic, evidence-based intervention, response cards (Christle & Schuster, 2003). The response card intervention was selected because, as previously stated, it has been shown to increase academic engagement and student participation (Christle & Schuster, 2003). Researchers offered to train teachers on an empirically supported academic intervention that has been shown increase academic engagement in their classrooms in an effort to increase teacher buy-in and establish a positive consultative relationship throughout the study (Eckert, Russo, & Hier, 2014).

Gresham et al. (2017) examined the Good Behavior Game, which is a class-wide, behavioral intervention. Researchers were interested in the possibility of the positive results found with the Good Behavior Game, a behavioral intervention, generalizing to the response card intervention, an academic intervention with related behavioral benefits. Researchers examined two methods of treatment integrity measurement, direct observation and self-report, when used to assess teachers' implementation of response cards.

Researchers used Generalizability Theory to conduct a decision (D) study in order to determine how many observations each method of integrity requires so that it reaches a point at

which it provides a reliable measure of treatment integrity. In other words, researchers investigated how efficient each measure of treatment integrity is when general education teachers implement the response card intervention. Researchers then analyzed the correlation between the two measures of treatment integrity. Additionally, average student academic engagement was measured at baseline and during the intervention in order to examine the effectiveness of response cards on student academic engagement.

The present study investigated the following research questions:

1. Are direct observation and self-report reliable methods of treatment integrity measurement when general education teachers implement the response card intervention in elementary classrooms?
2. How efficient is each method (e.g., direct observation and self-report) of treatment integrity measurement? In other words, how many observations does each method of treatment integrity measurement require before it reaches a point at which it provides a reliable measure of treatment integrity?
3. Are these direct observation and self-report methods significantly correlated?
4. Does the use of response cards in the classroom significantly improve student academic engagement in third, fourth, and fifth grade classrooms?

It is hypothesized that the results of this study will align with the Gresham et al. (2017) study, with direct observation yielding the most reliable measure of teacher implementation followed by self-report. Additionally, it is hypothesized that correlations will be significant but low between the two methods, as it is expected that teachers will be more likely to rate themselves as completing the components (Wickstrom, Jones, LaFleur, & Witt, 1998). It is also

hypothesized that student academic engagement will be significantly higher when teachers implement the response card intervention (Christle & Schuster, 2003).

Method

Participants and Setting

The primary participants in the study were six elementary school teachers from a low-income, high needs public school in East Baton Rouge Parish. To be included in the study, teachers were required to primarily teach classes from third through fifth grade. Researchers sought out teachers who operate with set blocks of direct instruction built in to their daily schedules. To participate in this study teachers had to be willing to be responsible for the implementation of the evidence-based, class-wide intervention (i.e., response cards) once every afternoon for ten consecutive school days to increase academic engagement in their classrooms. The afternoon block was chosen as a way of keeping time of day consistent across teachers. Teachers decided during which afternoon block, or academic subject, to implement the intervention depending on when they perceive academic engagement to be lowest. All of the study procedures took place within each teacher's classroom.

Graduate students in the school psychology doctoral program at Louisiana State University who had previous training in systematic direct observation, consultation with teachers, and classroom management strategies participated in the study. The graduate students were responsible for collecting baseline academic engagement data as well as ongoing academic engagement and treatment integrity data throughout the duration of the study.

Procedure

Before participants are recruited, the Institutional Review Board at Louisiana State University reviewed and approved the study methods and procedures. Administrative consent was obtained, which allows the researchers to recruit in their respective schools. Teachers were recruited to participate in this study if they had reached out for consultation services regarding

classroom management strategies in the past. Research assistants from Louisiana State University currently provide consultation services at the public school, so it is common for these services to be requested by teachers, especially in the elementary grades. Research assistants explained the procedures and timeline of the study and obtain informed consent from the first six teachers to agree to participation.

Once a teacher provided consent to participate in the study, a collaborative consultation meeting was scheduled with the teacher, in which the primary researcher explained the response card intervention and helped establish the intervention in the classroom. The primary researcher and the teacher engaged in implementation planning together based on the processes outlined in the PRIME model (Sanetti et al., 2015; Sanetti et al., 2014). This meeting involved a discussion about the intervention and its components, as well as any barriers to implementation that the teachers might anticipate. Teachers were asked to select an academic subject block in the afternoon (e.g., math, ELA, science, social studies) that they perceived to be problematic in terms of participation and academic engagement, so that the intervention would be implemented during that time period. The teachers were instructed to implement the response card intervention for 15 to 25 minutes each day during the selected academic block.

Researchers trained teachers to implement the response card intervention, based on the direct training research of Fallon et al. (2017) to promote treatment integrity using proactive strategies such as supplying teachers with oral and written descriptions of instructions, providing demonstration and modeling of the intervention, time for teachers to practice, as well as directed feedback. Teachers were given all of the materials necessary to implement the intervention as well as an opportunity to go through the materials and ask any questions during the first training

meeting. Performance feedback sessions were provided as often as needed throughout the 10 days of implementation process to improve treatment integrity.

Additionally, during the initial training session, teachers were trained on using the self-report form, also based on the direct training research of Fallon et al. (2017). They were given a typed protocol (Appendix A), which included in depth instructions of what would happen on observation days as well as outlined steps of the intervention in greater detail than the self-report forms that they would use to fill out during the observation days. These protocols were meant for the teachers to keep and refer back to if they have any questions.

Observation schedule. Each teacher was required to implement the response card intervention for 10 consecutive school days, or two full school weeks (i.e., once per afternoon during the designated block). Each intervention was observed by a researcher or graduate research assistant, resulting in 10 total observations per teacher – 60 observations overall. According to Ferguson et al. (2012), over the course of one day, a 15-minute observation was found adequate to yield reliable data for making low stakes decisions. Observations for the current study lasted anywhere from 15 to 20 minutes each, depending on how long it took the teacher to get started with the intervention and if there were any disruptions during the intervention. Each teacher presented 5 questions to the class during each intervention period, since the intervention periods were intended to last 15-20 minutes, and researchers initially assumed each question would take about 3 to 5 minutes.

During the observation, the primary researcher directly observed the intervention implementation for treatment integrity and completed the appropriate form to acquire an objective measure. Immediately after each component, the teacher answered either “yes” or “no” to the corresponding item on the self-report integrity form; consequently, each observation

resulted in one assessment of treatment integrity per each of the different sources. For the purpose of necessary interobserver reliability, there were two graduate students present for 25% of each teacher's treatment integrity observations. Graduate student research assistants attended a training specific to this study led by the main researcher to ensure adequate training on systematic direct observation as well as the direct observation of the response card intervention implementation, as treatment integrity and student academic engagement were observed simultaneously by the observers. All graduate research assistants were given a typed protocol (Appendix B), which outlines the observation schedule, procedure, materials needed, and general guidelines regarding observer duties for the direct observation period as well as instructions on how to conduct the SDO specific to the present study.

Although all graduate research assistants have had training and practicum experience using SDOs to rate on task, inattentive, and disruptive student behavior, the graduate research assistants were trained to collect SDO data specific to the present study. This study employed similar training techniques to that of Briesch et al. (2010). Graduate research assistants participated in two weeks of live training. During the first week, they were brought into the teachers' classrooms with the main researcher and asked to record instances of student academic engagement using a momentary time sampling recording method via systematic direct observation, just as they would during the actual study. This training activity was repeated every day for one week until the research assistants reached 90 percent interobserver agreement with the primary researcher on two consecutive observation trials. This data was used as baseline student academic engagement data for the teachers; therefore, each teacher was observed everyday either by only the primary researcher or the primary researcher and a graduate assistant. SDO training also included a meeting to review operational definitions and data

collection procedures followed by additional practice observations of recorded classrooms if needed. The following week, teachers began implementing the response card intervention as a practice week; however, they did not complete self-report forms. The graduate research assistants were brought into the same teachers' classrooms just as they were the week before; however, this time, in addition to recording student academic engagement on the SDO form, they concurrently completed the direct observation treatment integrity form. This training activity was repeated every day for one week until the research assistants reached 90 percent interobserver agreement with the primary researcher on two consecutive observation trials. On the third week, the study began, and data was collected. IOA was calculated by dividing the number of agreements by the number of agreements plus disagreements and then multiplying by 100. A recording is considered an agreement when two observers record the same behavior in the same observation interval. Additionally, reliability checks were conducted during study data collection. The academic engagement data was presented to individual teachers at the end of 10-day implementation period to show teachers the positive effect that response cards had in their classrooms.

Measures

Two different measures of treatment integrity were analyzed – direct observation and self-report. These specific measures were chosen because they are found to be the most commonly used methods to measure treatment integrity in practice (Lane, Boccian, MacMillan, & Gresham, 2004). Each component of the response card intervention was operationalized so that these components could be standardized across both measures (Lambert et al., 2006).

Direct observation. In order to assess the percentage of components completed by teachers during a given intervention period, an observation form (see Appendix C) was created

in which the seven items directly match the seven components mentioned in the introduction. During each intervention period, each teacher asked the class five questions that the students answered using their response cards. These questions were reflected on the observation form as item numbers. Observers circled either "yes" or "no" depending on whether they observed the teacher complete the component for that specific item number. All of the components are discrete, explicit events that can be easily identified by the observer. Treatment integrity was calculated by dividing the number of "yes" components by the total number of components listed and multiplying by 100 to yield a percentage of treatment integrity.

Self-report form. The self-report form (see Appendix D) is identical to the observation form but was completed by the teacher, rather than an observer, immediately after each component is completed. The teachers circled either "yes" or "no" depending on whether they completed that specific component on that specific item number. Treatment integrity was calculated in the same way as direct observation, by dividing the number of "yes" components by the total number of components listed and multiplying by 100 to yield a percentage of treatment integrity.

Student Academic Engagement. Student academic engagement is defined as time when the student is actively attending to teacher's instruction during the intervention period. This includes activities such as writing on a response card, attending to the teacher (e.g., eyes on the teacher, body listening to the lecture/instructions), looking at one's own response card, responding to a question by raising response card or answering aloud if called on, or listening to a peer respond to a question. Observers collected academic engagement data for 5 days prior to the intervention starting in order for each teacher to establish a baseline level of student academic engagement. Data was then collected during every intervention period throughout the

10-day implementation for each teacher. Academic engagement was recorded using a momentary time sampling SDO method (see Appendix E for SDO form) with 15-second intervals for twenty minutes, resulting in 80 intervals. Observers watched one student for the first three seconds of each 15-second time interval and if the student was academically engaged during those three seconds, the observer placed an "A" in the box to indicate academic engagement. If the student was not engaged during those three seconds, the observer placed an "X" in the box to indicate no academic engagement. The observer cycled through each student in the classroom and repeated the process until the end of the observation period.

Intervention materials. The response cards were created from half pages of card stock paper that were then laminated so that they would be durable. Each teacher was provided with enough response cards so that every student in the classroom had one. The same number of markers and socks, used as erasers, were provided to the teachers as well. The teachers were also given a self-report form by the observer each day that the teachers were observed.

Results

The researchers conducted a G study to evaluate the generalizability of each method of treatment integrity. A fully crossed analytic design with two facets was created in which treatment integrity method (*m*: i.e., self-report and direct observation) is fully crossed with person, the object of measurement (*p*: i.e., teachers). Person is also fully crossed with occasion (*o:* i.e., observation), and occasion is fully crossed with method (see Figure 1 for conceptualization).

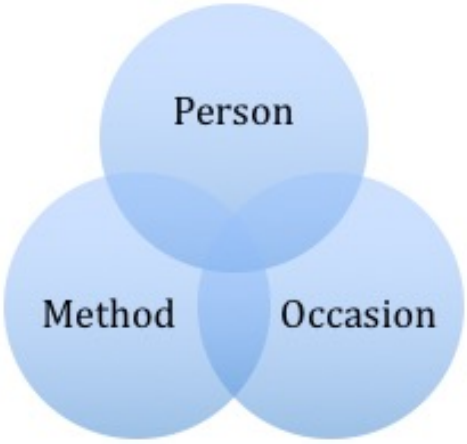


Figure 1. 2-facet, fully crossed analytic design: *m x p x o*. This figure demonstrates the analytic design of the current study. Teachers' implementation of the intervention will be assessed using both of the treatment integrity measurement methods during every observation.

In other words, teachers' implementation of the intervention was assessed using both of the treatment integrity measurement methods during each observation. The observation schedule was created so that each teacher's implementation is observed for ten consecutive school days. Furthermore, researchers planned to conduct a follow-up D study to establish the G and Phi coefficients associated with each treatment integrity method when examining different numbers of observations. The D study was intended to determine the amount of observations needed for

each method to achieve reliable treatment integrity scores when manipulating the most malleable facet (occasion) using a cutoff criterion of .80 (Briesch, Swaminathan, Welsh, & Chafouleas, 2014). In other words, researchers hoped to determine how many days each method must be used before producing a reliable treatment integrity score (> .80). All analyses were conducted using syntax specifically written for generalizability theory analyses in SPSS (Mushquash & O'Conner, 2006). Finally, a Pearson's correlation (*r*) matrix was constructed for each method and correlations between the results of each method were tested.

Researchers hypothesized that results would show that the method of direct observation would produce the most reliable measure of treatment integrity followed by the self-report measure. Finally, researchers expected that correlations would be low between the two methods, as teachers are expected to be more likely to rate themselves as completing the components (Wickstrom et al., 1998).

The results of the G study are presented in Table 1. Estimations of variance in treatment integrity scores associated with each facet and interactions between facets were compared to total variance of scores to examine the percentage of variance accounted for by each. Once variance components were computed, it was possible to calculate a G coefficient and Phi coefficient signifying relative and absolute dependability of scores respectively, which determined if the methods produced sufficiently reliable estimates of treatment integrity when measured once per day for 10 consecutive days. Overall, the interaction between the Person (*p*: teacher) and Occasion (*o*: observation day) facets was the single largest contributor to variance in treatment integrity rating scores accounting for 60.5% of the variance, meaning the scores varied by 60.5% depending on individual teacher ratings on certain observation days (e.g., teacher 4 on observation day 6). Since Method (*m*: type of treatment integrity method: self-report vs. direct

observations) is a fixed facet (i.e., analogous to a fixed factor in ANOVA) the variance for each type of method could not be parsed, however, the Method facet accounted for 5.2% of the overall variance. In other words, the ratings scores varied by 5.2% depending on which method of treatment integrity was used to obtain those scores. The Occasion facet accounted for 2.9% of the variance, meaning that scores varied by 2.9% depending on which observation (e.g., observation 1, observation 2, etc.) was examined. The Person facet accounted for 0% of the variance, meaning that scores did not vary depending solely on individual teacher ratings. The interaction between the Person and Method facets accounted for 4.8% of the variance, meaning the scores varied by 4.8% depending on which method was used to rate individual teachers. The interaction between the Method and Occasion facets accounted for 1% of the variance, meaning that scores varied by 1% depending on which method was used on which day. Finally, error variance accounted for 6.5% of the variance.

Table 1. Proportion of Variance for Each Facet

Facet	Proportion of Variance
Person (*p*)	0
Method (*m*)	5.2
Occasion (*o*)	2.9
Person x Method	4.8
Person x Day	60.5
Method x Day	1
Error	26.5

Once the variance components were computed it was possible to calculate a G coefficient and Phi coefficient, which represent the relative and absolute dependability of the scores generated by each assessment method, respectively. The G coefficient was .000 for both the direct observation method and self-report. The Phi coefficient was also .000 for both direct observation method and self-report. When looking at these results with no context, these

coefficients suggest that neither method produces sufficiently reliable estimates of treatment integrity when measured once per day for ten consecutive days, regardless of the type of decision being made (i.e., relative or absolute).

Decision Studies

Because the G study suggested that the original measurement model did not produce sufficiently dependable assessment data across the two methods, a decision study was conducted to determine the point at which each method would produce a reliable assessment when manipulating the most malleable facet, Occasion (i.e., number of observations). Using .80 as a cutoff criterion for adequate dependability (e.g., Briesch, Swaminathan, Welsh, & Chafouleas, 2014), the relative dependability of each measure of treatment integrity does not increase from day one of observations to day 10. At day 10, neither the direct observation nor self-report method would produce reliable treatment integrity assessments, with G coefficients remaining at .000 for both methods of measurement.

Correlation

Pearson's correlation (r = .689) was significant between direct observation and self-report methods, which was not the expected result.

Student Academic Engagement

A repeated measures t-test was conducted to compare average student academic engagement during baseline and post-intervention. There was not a significant difference in the scores for baseline (M = 76.2, SD = 13.26) and post-intervention (M = 79.0, SD = 6.03); $t(5) = -0.822$, $p = 0.224$. These results are also not as researchers expected. See Table 2 for further explanation of results.

Table 2. Average Student Academic Engagement

Teacher	Baseline (%)	Post-Intervention (%)
1	88	81
2	69	81
3	89	86
4	70	75
5	56	69
6	85	82

Discussion

The current study is an empirical replication of the Gresham et al. (2017) study and results were expected to align accordingly. The purpose of the study was to determine the dependability of two different measures of treatment integrity when general education teachers implement the response card intervention. Repeated measurement involving direct observation and self-report allowed for comparison of two of the most common measures of treatment integrity in schools. Results indicated that neither method produced a reliable measure of treatment integrity, when teachers were observed once per day for ten consecutive days. The issue that is driving these poor G study results is the lack of overall variance in scores. G studies are used to assess the degree to which variability in scores are attributable to different facets of measurement. In the present study's data, there is really not much variability. Eighty-eight of the 120 total treatment integrity scores are equal to 100, and the mean is 97.75 (SD = 5.98). Therefore, there is not much variance for the model to explain, which is not necessarily a negative outcome when looking at treatment integrity methods of measurement. The assessment method was examined as a fixed facet in the model, as the same methods were used each day by the same person (i.e., self-report form used by the teacher and direct observation form used by the observer). The variance component for the object of measurement, Persons, or Teachers in this case, was -1.902, which should be interpreted as zero variance explained.

When examining these results at face value and without context, the results are unimpressive. These results are the opposite of what the researchers expected; however, the implications regarding treatment integrity and student academic engagement when the response card intervention is implemented by well-trained teachers is noteworthy. Due to the extensive training given to the teachers in the form of detailed protocols, direct training methods,

collaborative consultation and intervention implementation planning meetings between teachers and researchers, as well as providing performance feedback as often as needed, the teachers in the study were well-prepared for accurate implementation.

Similar to the G and D study results, the non-significant correlation between the two methods was not originally expected; however, in the context of the small amount of variability found, these results make sense. Treatment integrity was so high that both methods yielded perfect integrity on about 73% of measurements.

There was not a statistically significant increase in average student academic engagement after teachers implemented the response card intervention. Overall, each teacher's class, no matter how they performed at baseline, was closer to a student academic engagement average of 80% at the post-intervention measurement. Teachers whose classes demonstrated lower baseline averages showed an increase when measured after intervention. Conversely, the teachers whose classes demonstrated higher averages at baseline showed a small decrease when measured after intervention. These results may suggest that the response card intervention is an effective tool to increase academic engagement for teachers struggling with classroom management. Those teachers who already have classroom management tools in place may not benefit from adding something new or changing up their strategies when their average academic engagement is already at the 80% mark or higher.

Limitations and Future Directions

These results should be interpreted in the context of a number of limitations. First, the largest factor contributing to the low coefficients was the combination of person and day effects. Although the sampling plan was ideal (i.e., once per day for 10 consecutive days), it may not have been the most advantageous time of the year. The study was conducted in the last week of

October and first week of November. During this time, the school was conducting science fairs as well as having Halloween parties. Because of this, two days had to be skipped and added on at the end. The students were understandably distracted during the days leading up to these events, which could have had an effect on the overall academic engagement scores for those days. In the future, picking a two-week period when the researchers can ensure with administration beforehand that there will be no extracurricular activities happening within the school that would be distracting to teachers or student would be ideal, as this would limit any extraneous variables between observations that could affect implementation or teacher motivation.

Additionally, training research assistants and teachers on study procedures took a substantial amount of time. Research assistants attended a meeting with the primary researcher to review operational definitions and data collection procedures. They were first trained in vivo for one week on how to rate students' academic engagement, then in vivo for the second week on how to concurrently rate students' academic engagement while completing the direct observation treatment integrity measurement form. During the second week of the research assistants' training, teachers implemented the response card intervention so that research assistants could be trained as well as so they could practice and receive performance feedback. Teachers also attended a collaborative consultation meeting with the primary researcher and engaged in implementation planning. Teachers were able to request meetings with the primary researcher and research assistants at any time during the training and data collection. Overall, the training for research assistants and teachers was time intensive and may not be feasible in practical application. Service providers in schools should consider the cost in time and resources used for training and implementation of this tier 1 classroom management intervention and decide whether the benefits outweigh the cost.

Another limiting factor is that all data collected in this study came from one high needs public school with limited resources. Future research should examine the generalizability of these two treatment integrity measurement methods in samples with different characteristics (e.g., high socioeconomic status, rural areas, private and parochial schools, etc.).

Conclusions

Although the results of this study were not significant, the results coupled with the study's design and limitations add to the research literature, which utilizes Generalizability Theory to assess the reliability of behavioral assessment in schools (e.g., Chafouleas, Christ, Riley-Tillman, Briesch, & Chanese, 2007; Hintze & Matthews, 2004; Volpe & Briesch, 2014) as well as the new concept started by Gresham et al. (2017), which encourages researchers to add to the treatment integrity database. The results should be interpreted with caution and more research is needed involving the response card intervention as well as different behavioral and academic interventions to allow for more solid conclusions to be made. As stated previously, certain components of an intervention may be crucial to behavior change, whereas other components of the same intervention may not be as important in driving change, and more research is needed regarding which components of the response card intervention are responsible for most change in student academic engagement. This study is an important step forward in creating the treatment integrity database that will benefit practitioners, researchers, and policy-makers when they are faced with the decision of which treatment integrity measurement to use for various treatments for various issues, as well as how many assessments are needed in order to yield the most reliable results most efficiently.

Appendix A: Teacher Training Protocol

Study begins:

Materials: You have been given enough white boards, socks, and markers so that each student will receive one of each. These are yours to keep if you wish.

Intervention period: _____

Beginning _____, you will implement the Response Card Intervention in the instructional block that you chose (indicated above) for around 20 minutes, asking eight questions during each intervention session. You will conduct class as you normally would during this period. The only difference is that for whole group instruction, instead of the students responding orally, they will write their answers on their cards (see steps below). You will do this for 10 consecutive school days, ending _____. All 10 intervention sessions will be observed by trained graduate students from LSU. When the graduate student observes, they will fill out an observation form. They will hand you a self-report form, which is identical to the observation form that they will be filling out, that you are to fill out while implementing the intervention. This self-report form lists the steps of the intervention. Whenever you complete a step, you will circle "Y" indicating that you completed the step. If you fail to complete a step, you will circle "N." You will hand this completed form to the observer when they leave.

Steps of Intervention
** It is important to note that steps 3 through 7 will be repeated eight times on the self-report form, reflecting the eight questions that will be posed to the class during the intervention session. These eight questions will be reflected in intervention steps 3a through 10e (see Self-Report Form).

1. Teacher states to the class that it is time to respond using response cards.
    a. Students will take out their white boards, sock (used as eraser), and marker.
2. The teacher reviews the rules of the intervention to the class.
    a. E.g., ten seconds to answer the question, do not raise your cards until I say to do so, do not talk to your peers, answer the question by yourself, markers only write on response cards, caps go on the back of markers, etc.
3. The teacher presents a question to the class.
    a. This does not have to be a formal process. These are questions that would come naturally with the lesson and that you would normally ask the class. Instead of choosing a particular student to answer or having the whole class respond at once, each student will write their answer on the response card.
4. The teacher provides adequate time for students to write answers on their response cards.
5. The teacher requests that students present their cards.
6. If less than 75% of students provide correct answers, the teacher instructs the class to fix their answers and provide a rationale for the answer; if 75% or more of the class answers correctly, the teacher reveals the correct answer to the class.

    a. This can be a quick scan of the classroom; it does not have to be exact. If you think that less than ¾ of the class answered correctly, you will have them try again. If the majority answered correctly, you will reveal the correct answer.
7. The teacher provides verbal praise for correct responses.

Appendix B: Graduate Research Assistant Protocol

Schedule: The study will run for two weeks (10 consecutive school days). Ten observations of implementation of the response card intervention as well as student academic engagement must be completed for each teacher (6). Therefore, 60 observations must be conducted over the course of the project. Furthermore, 30% of those implementations will require interobserver agreement (IOA). That means a second observer will be present for 3 observations for each teacher. You will find the observation schedule in the dropbox folder. This schedule outlines the times that each teacher will be implementing the intervention, observation days, and observers for each time slot. Each observation will last for at least 20 minutes, in the case of the student academic engagement assessment, and could possibly last 25 minutes for intervention implementation. In the case that the teacher and students are absent from school (e.g., fieldtrips, holidays), the observation schedule will be pushed back the same number of days as the absence. We can adjust the end of the 10-day period as we come across absences.

Observation Procedure: To determine the observation time for the teacher you will observe, check the schedule that is located in the dropbox folder. If there are any complications, email the group by as soon as possible.

Materials: Observer Report Form
Teacher Self-Report Form
Student Academic Engagement Observation Form
Writing utensil
Timer (can be on phone)
Interval recording (located in dropbox)

Arrive to the classroom with the necessary materials at least 5 minutes before the response card intervention is scheduled to be implemented. It is critical to the study's success that the observer be present for the entire observation period in which the response card intervention is scheduled. This will require adjusting your travel time for traffic and reasonably predictable events.

When you enter the classroom, do so with minimal interruption, hand the self-report form to the teacher, and position yourself in an area that will not cause any disruption throughout the entire observation period. You may sit or stand, whichever you prefer, as long as you are comfortable, have an unobstructed view of the classroom, and can easily record teacher and student behavior.

Your 25-minute observation begins at the time indicated on the observation schedule. Your first responsibility will be to begin the systematic direct observation of student academic engagement that you have been trained on (see Student Academic Engagement Observation Form for instructions). Next, you will look for components 1 and 2 on Observer Report Form, as these should all be completed by the teacher before the intervention begins. Once these steps have been completed (or omitted) you should focus primarily on recording whether or not the teacher completes components 3a through 10e. The teacher should ask eight questions that the students will be responding to using response cards. Continue this process until the observation period ends. At this time the teacher should have completed components 1 through 10e on the self-

report form you handed her before the intervention began. **DO NOT AT ANY TIME DURING THE INTERVENTION REMIND THE TEACHER TO COMPLETE ANY INTERVENTION STEPS.**

When the 25-minute observation period is over, even if the teacher has not completed all the steps, approach the teacher discreetly to collect the self-report form. If the teacher would like performance feedback immediately, let her know how she rated her self-report form compared to yours as well as which steps of the intervention she incorrectly or failed to implement. If she would like to schedule a performance feedback meeting, schedule one at your earliest convenience or give refer her to the study email address.

Do not leave the classroom without the self-report form. It is critical that this data is collected immediately after the end of each response card intervention session that you observe.

Ensure you have written all observation information on the forms, including the date and time of the observation, your initials, and IOA information, and the teacher's name. This is especially important to note on the Teacher Self-Report Form because the teacher does not record this information.

Appendix C: Observer Report Form

The teacher states to the class that it is time to respond using response cards.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

The teacher reviews the rules of the intervention with the class.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

The teacher presents a question to the class.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

The teacher provides adequate time for students to write their answers on the response cards.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

The teacher requests that students present their cards.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

If more than five students provide incorrect answers, the teacher instructs the class to fix their answers and provide a rationale for the answer; if not, the teacher reveals the correct answer to the class.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

The teacher provides verbal praise for correct responses.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

Appendix D: Self Report Form

The teacher states to the class that it is time to respond using response cards.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

The teacher reviews the rules of the intervention with the class.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

The teacher presents a question to the class.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

The teacher provides adequate time for students to write their answers on the response cards.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

The teacher requests that students present their cards.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

If more than five students provide incorrect answers, the teacher instructs the class to fix their answers and provide a rationale for the answer; if not, the teacher reveals the correct answer to the class.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

The teacher provides verbal praise for correct responses.

1. Y N
2. Y N
3. Y N
4. Y N
5. Y N

Appendix E: Student Academic Engagement Form

Observer _____ **Teacher** _____
Date_____

IOA? _____ **Time** _____

The observation is 20 minutes.
Place an **A** in the box if the student was academically engaged for the first three seconds of the interval.

Academic engagement is defined time when the student is actively attending to teacher's instruction during the intervention period. This includes activities such as writing on a response card, attending to the teacher (e.g., eyes on the teacher, body oriented toward the teacher, listening to the lecture/instructions), looking at one's own response card, responding to a question by raising response card or answering aloud if called on, or listening to a peer respond to a question.

Place an **X** in the box if the student engaged was inattentive, or not academically engaged for the first three seconds of the interval.

1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.
25.	26.	27.	28.	29.	30.	31.	32.	33.	34.	35.	36.
37.	38.	39.	40.	41.	42.	43.	44.	45.	46.	47.	48.
49.	50.	51.	52.	53.	54.	55.	56.	57.	58.	59.	60.
61.	62.	63.	64.	65.	66.	67.	68.	69.	70.	71.	72.
73.	74.	75.	76.	77.	78.	79.	80.				

Comments:

Appendix F: IRB Approval

ACTION ON EXEMPTION APPROVAL REQUEST

Institutional Review Board
Dr. Dennis Landin, Chair
130 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.5983
irb@lsu.edu
lsu.edu/research

TO: Elizabeth Wilson
Psychology

FROM: Dennis Landin
Chair, Institutional Review Board

DATE: September 7, 2018

RE: **IRB#** E11165

TITLE: Generalizability of Multiple Measures of Treatment Integrity: Response Card Intervention

New Protocol/Modification/Continuation: New Protocol

Review Date: 8/31/2018

Approved X **Disapproved**_____

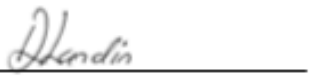
Approval Date: 9/7/2018 **Approval Expiration Date:** 9/6/2021

Exemption Category/Paragraph: 1; 2b

Signed Consent Waived?: No

Re-review frequency: (three years unless otherwise stated)

LSU Proposal Number (if applicable):

By: Dennis Landin, Chairman 

PRINCIPAL INVESTIGATOR: PLEASE READ THE FOLLOWING –
Continuing approval is CONDITIONAL on:

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU's Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submittal of a termination report), prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins); notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.
5. Continuing attention to the physical and psychological well-being and informed consent of the individual participants, including notification of new information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
7. Notification of the IRB of a serious compliance failure.
8. **SPECIAL NOTE: When emailing more than one recipient, make sure you use bcc. Approvals will automatically be closed by the IRB on the expiration date unless the PI requests a continuation.**

* *All investigators and support staff have access to copies of the Belmont Report, LSU's Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at http://www.lsu.edu/irb*

References

Billingsley, B. S. (2001). Beginning special educators: Characteristics, qualifications, and experiences. SPeNSE Summary Sheet. (ERIC Document Reproduction Service No. ED467269).

Billingsley, B. S., & Tomchin, E. M. (1992). Four beginning LD teachers: What their experiences suggest for trainers and employers. *Learning Disabilities Research & Practice, 7,* 104–112.

Bosworth, K., Gingiss, P. M., Potthoff, S., & Roberts-Gray, C. (1999). A Bayesian model to predict the success of the implementation of health and education innovations in school-centered programs. *Evaluation and Program Planning, 22,* 1–11.

Briesch, A.M., Chafouleas, S.M., Riley-Tillman, T.C. (2016). *Direct behavior rating: Linking assessment, communication, and intervention.* New York: The Guilford Press.

Briesch, A.M., Swaminathan, H., Welsh, M., & Chafouleas, S.M. (2014). Generalizability theory: A practical guide to study design, implementation, and interpretation. *Journal of School Psychology, 52*, 13-35.

Briesch A.M., Chafouleas, S.M., & Riley-Tillman, T. C. (2010). Generalizability and dependability of behavior assessment methods to estimate academic engagement: A comparison of systematic direct observation and Direct Behavior Rating. *School Psychology Review,* 39, 408-421.

Chafouleas, S. M., Christ, T. J., Riley-Tillman, T. C., Briesch, A. M., & Chanese, J. A. M. (2007). Generalizability and dependability of direct behavior ratings to assess social behavior of preschoolers. *School Psychology Review, 36(1),* 63–79.

Christle, C.A. & Schuster, J.W. (2003). The effects of using response cards on student participation, academic achievement, and on-task behavior during whole-class, math instruction. *Journal of Behavioral Education, 12(3),* 147-165.

Chronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: John Wiley & Sons, Inc.

Curtis, D. F., Pisecco, S., Hamilton, R. J., & Moore, D. W. (2006). Teacher perceptions of classroom intentions for children with ADHD: A cross cultural comparison of teachers in the United States and New Zealand. *School Psychology Quarterly, 21(2),* 171-196.

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,* 23–45.

Darling-Hammond, L. (2003). Keeping good teachers. *Educational Leadership, 60(8),* 6–77.

Dart, E. H., Cook, C. R., Collins, T. A., Gresham, F. M., & Chenier, J. S. (2012). Test driving interventions to increase treatment integrity and student outcomes. *School Psychology Review, 41*, 467–481.

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

Duncan, N. G., Dufrene, B. A., Sterling, H. E., & Tingstrom, D. H. (2013). Promoting teachers' generalization of intervention use through goal setting and performance feedback. *Journal of Behavioral Education, 22*(4), 325-347.

Durlak, J. A., & DuPre, 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(3-4),* 327-350.

Eckert, T. & Hintze, J. (2001). Behavioral conceptions and applications of acceptability: Issues related to service delivery and research methodology. *School Psychology Quarterly, 16,* (1), 122.

Eckert, T. L., Russo, N., & Hier, B. O. (2014). Best practices in school psychologists' promotion of effective collaboration and communication between school professionals. In P. L. Harrison & A. Thomas (Eds.), *Best Practices in School Psychology: Data-Based and Collaborative Decision Making* (pp. 541-552)

Engelmann, S., & Colvin, G. (1983). Generalized compliance training: A direct- instruction program for managing severe behavior problems. Austin, TX: PRO-ED.

Fallon, L. M., Kurtz, K. D., & Mueller, M. R. (2017, June 15). Direct Training to Improve Educators' Treatment Integrity: A Systematic Review of Single-Case Design Studies. School Psychology Quarterly. Advance online publication. http://dx.doi.org/10.1037/spq0000210

Fallon, L. M., Collier-Meek, M. A., Sanetti, L. M., Feinberg, A. B., & Kratochwill, T. R. (2016). Implementation Planning to Promote Parents' Treatment Integrity of Behavioral Interventions for Children with Autism. *Journal of Educational and Psychological Consultation, 26*(1), 87-109.

Feldstein, A. C., & Glasgow, R. E. (2008). A practical, robust implementation and sustainability model (PRISM). *The Joint Commission Journal on Quality and Patient Safety, 34,* 228–243.

Ferguson, T. D., Briesch, A. M., Volpe, R. J., & Daniels, B. (2012). The influence of observation length on the dependability of data. *School Psychology Quarterly, 27(4),* 187-197.

Fimian, M. J., & Santoro, T. M. (1983). Sources and manifestations of occupational stress as reported by full-time special education teachers. *Exceptional Children, 49,* 540–543.

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

Fixsen, D. L., Naoom, S. F., Blase′, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida. Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231) Retrieved from http://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-MonographFull-01-2005.pdf.

Forman, S. G., Olin, S. S., Hoagwood, K. E., Crowe, M., & Saka, N. (2009). Evidence-based interventions in schools: Developers' views of implementation barriers and facilitators. *School Mental Health, 1,* 26–36. http://dx.doi.org/10.1007/s12310-008-9002-5.

Fox, L., Hemmeter, M., Snyder, P., Binder, D. P., & Clarke, S. (2011). Coaching early childhood special educators to implement a comprehensive model for promoting young children's social competence. *Topics in Early Childhood Special Education, 31,* 178-192.

Godfrey, S. A., Grisham-Brown, J., Hemmeter, M. L., & Schuster, J. W. (2003). The effects of three techniques on student participation with preschool children with attending problems. Education and Treatment of Children.

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

Gresham, Frank M. (2009). Evolution of the treatment integrity concept: Current status and future directions. *School Psychology Review, 38(4),* 533-540.

Gresham, F.M., Dart, E., & Collins, T. (2017). Generalizability of multiple measures of treatment integrity: Comparisons among direct observations, permanent products, and self-report. *School Psychology Review, 46(1),* 108-121.

Heward, W. L. (2000). Exceptional children: An introduction to special education (6th ed.). New Jersey: Prentice Hall.

Hineline, P. N. (2003). The aesthetics of behavioral interventions. Invited address presented at the Pennsylvania Association for Behavior Analysis Conference, Harrisburg, PA.

Hintze, J.M., & Matthews, W.J. (2004). The generalizability of systematic direct observations across time and setting: A preliminary investigation of the psychometrics of behavioral observation. *School Psychology Review, 33*(2), 258-270.

Individuals with Disabilities Education Improvement Act of 2004, 20 U.S.C. 1400, Pub. L No. 108–446.

Institute of Education Sciences.

Kam, C., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science, 4,* 55–63. http://dx.doi.org/10.1023/A:1021786811186.

Kelleher, C., Riley-Tillman, T. C., & Power, T. J. (2008). An initial comparison of collaborative and expert-driven consultation on treatment integrity. *Journal of Educational and Psychological Consultation, 18,* 294–324.http://dx.doi.org/10.1080/10474410802491040.

Kratochwill, T. R. & Pittman, P. H. (2002). Expanding problem-solving consultation training: Prospects and framework. *Journal of Educational & Psychological Consultation, 13(1-2),*69-95.

Lambert, M.C., Cartledge, G., Heward, W.L., & Lo, Y. (2006). Effects of response cards on disruptive behavior and academic responding during math lessons by fourth-grade urban students. *Journal of Positive Behavior Interventions, 8(2),* 88-99.

Lane, K. L., Boccian, 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(3*), 36–43.

Lippke, S., Wiedemann, A. U., Ziegelmann, J. P., Reuter, T., & Schwarzer, R. (2009). Self-efficacy moderates the mediation of intentions into behavior via plans. *American journal of health behavior, 33*(5), 521-529.

Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current directions in psychological science, 15*(5), 265-268.

Long, A. C., Sanetti, L. M. H., Collier-Meek, M. A., Gallucci, J., Altschaefl, M., & Kratochwill, T. R. (2016). An exploratory investigation of teachers' intervention planning and perceived implementation barriers. *Journal of school psychology, 55*, 1-26.

Martens, B. K., Hiralall, A. S., & Bradley, T. A. (1997). A note to teacher: Improving student behavior through goal setting and feedback. *School Psychology Quarterly, 12*(1), 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.

Mushquash, C., & O'Connor, B. P. (2006). SPSS, SAS, and MATLAB programs for generalizability theory analyses. *Behavior Research Methods, 38*, 542-547.

Noell, G. H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W. P. Erchul, & S. M. Sheridan (Eds.), *Handbook of research*

in school consultation: Empirical foundations for the field (pp. 315–334). Mahwah, NJ: Erlbaum.

Noell, G. H. (2010). Empirical and pragmatic issues in assessing and supporting intervention implementation in schools. In G. G. Peacock, R. A. Ervin, E. J. Daly, & K. W. Merrell (Eds.), *Practical handbook in school psychology: Effective practices for the 21st century* (pp. 513–530). New York, NY: Guilford Publications.

Noell, G. H., Witt, J. C., LaFleur, L. H., Mortenson, B. P., Ranier, D. D., & LeVelle, J. (2000). Increasing intervention implementation in general education following consultation: A comparison of two follow-up strategies. *Journal of Applied Behavior Analysis, 33,* 271–284.

Noell, G. H., Witt, J. C., Slider, N. J., & Connell, J. E. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review*, *34*(1), 87.

Noell, G. H., & Gansle, K. A. (2014). The use of performance feedback to improve intervention implementation in schools. In L. M. H. Sanetti & T. R. Kratochwill (Eds.), *Treatment integrity: A foundation for evidence-based practice in applied psychology* (pp. 161-183). Washington, DC: American Psychological Association.

Noell, G.H., Gansle, K.A., Mevers, J.L., Knox, R.M., Mintz, J.C., & Dahir, A. (2014) Improving treatment plan implementation in schools: A meta-analysis of single subject design studies. *Journal of Behavioral Education, 23*, 168-191. DOI 10.1007/s10864-013-9177-1

Perepletchikova, F., Treat, T., & Kazdin, A. (2007). Treatment integrity in psychotherapy research: Analysis of studies and examination of associated factors. *Journal of Consulting and Clinical Psychology, 75,* 829–841.

Riley-Tillman, T. C., Chafouleas, S. M., Briesch, A. M., & Eckert, T. L. (2008). Daily behavior report cards and systematic direct observation: An investigation of the acceptability, reported training and use, and decision reliability among school psychologists. Journal of Behavioral Education, 17(4), 313–327.

Riley-Tillman, T. C., Chafouleas, S. M., Christ, T. J., Briesch, A. M., & LeBel, T. J. (2009). The impact of item wording and behavioral specificity on the accuracy of direct behavior ratings (DBRs). *School Psychology Quarterly, 24*, 1-12.

Riley-Tillman, T. C., Christ, T.J., Chafouleas, S. M., Boice-Mallach, C. H., Briesch, A. (2011). The impact of observation duration on the accuracy of data obtained from direct behavior rating (DBR). *Journal of Positive Behavior Interventions*, 13(2), 119-128)

Roach, A. T., & Elliot, S. N. (2008). Best practices in facilitating and evaluating intervention integrity. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology V*, Vol. 2. (pp. 195–208). Bethesda, MD: National Association of School Psychologists.

Sanetti, L. M. H., Collier-Meek, M. A., Long, A. C., Byron, J., & Kratochwill, T. R. (2015). Increasing teacher treatment integrity of behavior support plans through consultation and implementation planning. *Journal of school psychology, 53*(3), 209-229.

Sanetti, L. M. H., Collier-Meek, M. A., Long, A. C., Kim, J., & Kratochwill, T. R. (2014). Using Implementation Planning to Increase Teachers' Adherence and Quality to Behavior Support Plans. *Psychology in the Schools, 51*(8), 879-895.

Sanetti, L. M. H., Fallon, L. M., & Collier-Meek, M. A. (2011). Treatment integrity assessment and intervention by school-based personnel: Practical applications based on preliminary study. *School Psychology Forum, 5*(3), 87–102.

Sanetti, L.M.H., & Kratochwill, T.R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. *School Psychology Review, 38*(4), 445-459.

Sanetti, L. M. H., Kratochwill, T. R., & Long (2013). Applying adult behavior change theory to support mediator-based intervention implementation. *School Psychology Quarterly, 28,* 47–62. http://dx.doi.org/10.1037/spq0000007.

Sanetti, L. M. H., Williamson, K. M., Long, A. C. J., & Kratochwill, T. R. (2017). Increasing In-Service Teacher Implementation of Classroom Management Practices Through Consultation, Implementation Planning, and Participant Modeling. *Journal of Positive Behavior Interventions*. doi:10.1177/1098300717722357

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(4),* 460-475.

Schwarzer, R. (2008). Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology, 57*(1), 1-29.

Shabani, D. B., & Carr, J. E. (2004). An evaluation of response cards as an adjunct to standard instruction in university classrooms: A systematic replication and extension. *North American Journal of Psychology, 6(1),* 2004. pp. 85-100.

Shavelson, Richard & Webb, Noreen. (1991). *Generalizability Theory: A Primer*. Thousand Oaks, CA: Sage Publications, Inc.

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*(4), 476-495.

Volpe, R. J., & Briesch, A. M. (2012). Generalizability and dependability of single-item and multiple-item direct behavior rating scales for engagement and disruptive behavior. *School Psychology Review, 41(3),* 246–261.

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

Vita

Elizabeth Kelsey Wilson was born in Baton Rouge, Louisiana and attended Louisiana State University in Baton Rouge from August 2008 to May 2012 where she earned a Bachelor of Arts in psychology. She then attended the University of Texas at Dallas in Richardson, Texas, from August 2012 to May 2014 where she earned a Master of Science in psychological sciences with a focus in developmental psychology from the School of Brain and Behavioral Sciences. Elizabeth is a fifth-year graduate student in school psychology at Louisiana State University under the supervision of Dr. Frank M. Gresham. She is completing her internship at Casa Pacifica Center for Children and Families in Camarillo, California, where she is working at the non-public school and short-term therapeutic residential center, providing services to students dealing with issues such as complex trauma, depression, anxiety, mood disorder, and attachment issues. Elizabeth's clinical and research interests include behavioral interventions for students with emotional and behavioral disorders, mental health awareness, as well as the influence of trauma and violence exposure on students. Elizabeth will earn her doctorate in Psychology in August of 2020.