PARAPROFESSIONALS' USE OF DIRECT BEHAVIOR RATING FOR COLLECTING FUNCTIONAL BEHAVIOR ASSESSMENT DATA: AGREEMENT AND ACCEPTABILITY

BY

BRIAN N. SOUTH

A DISSERTATION SUBMITTED TO THE FACULTY OF

ALFRED UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PSYCHOLOGY

IN

SCHOOL PSYCHOLOGY

ALFRED, NEW YORK

DECEMBER 2017

PARAPROFESSIONALS' USE OF DIRECT BEHAVIOR RATING FOR COLLECTING FUCNTIONAL BEHAVIORAL ASSESSMENT DATA: AGREEMENT AND ACCEPTABILITY

BY

BRIAN N. SOUTH

UNIVERSITY OF PITTSBURGH AT BRADFORD, B.S. (1998)

ST. BONAVENTURE UNIVERSITY, MS Ed. (2003)

ALFRED UNIVERSITY, M.A. (2015)

APPROVED BY: Mark Fugate, Ph.D.
Committee Chairperson

Nancy Evangelista, Ph.D.
Committee Member

Nancy Furlong, Ph.D.
Committee Member

ACCEPTED BY Mark Fugate, Ph.D.
Chairperson, Division of School Psychology

ACCEPTED BY Nancy Evangelista, Ph.D.
Associate Provost and Director of Graduate Studies

ACCEPTED BY Rick Stephens, Ph.D.
Provost & Vice President for Academic Affairs

Acknowledgements and Dedication

First and foremost, I would like to thank my family. To my parents, Linda and Bob, for instilling a life-long love of learning, showing me the value of hard work, and teaching me the importance of helping others. Thank you for always being supportive of my ambitions in life. To my brothers, thank you for helping me to never give up no matter the obstacle.

I would like to express my deepest gratitude to Dr. Mark Fugate for his patient guidance while chairing my dissertation committee. Thank you for sharing your expertise in behavior analysis and offering me advice during my journey. Your mentorship these past few years has been invaluable to my development as a psychologist and growth as a behavior analyst.

I would like to thank my committee members, Dr. Nancy Evangelista and Dr. Nancy Furlong, for the friendly guidance, thought-provoking suggestions, and general collegiality they provided to me during this process. I would also like to acknowledge Dr. Chris Lauback, whose positivity and support early on helped frame this project.

In a similar vein, I would like to thank all of my professors at Alfred University; I am especially grateful for each of their contributions that helped me to see my true potential and gain confidence in myself as a psychologist. Dr. O'Connell, thank you for being such an important mentor and for challenging me to excel in my strengths and extend beyond my comfort zone. Dr. Bender, thank you for your time, energy, and supervision while allowing me to pursue my teaching interests.

I also have to express my gratitude to my colleagues at Beacon Light Behavioral Health Systems and The Rehabilitation Center who have been exceptionally supportive and accommodating of my work while pursing my doctorate. I am privileged to work with such dedicated people.

Finally, I offer my most sincere gratefulness to my wife, Jennifer, for her shouldering more than a fair share of the parenting and household duties while I pursued this degree. Thank you for all of the long conversations about this study as I was trying to figure it all out and for keeping me focused when the process became difficult. To my children, Lillian and Lucas, who in no way understand the sacrifices they accepted in order for me to complete my degree, I hope this serves as an inspiration for you to always pursue your educational aspirations to their fullest.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

This paper reviews existing literature on Functional Behavior Assessment (FBA) and the benefits and challenges associated with conducting the cornerstone of behavioral observation, Systematic Direct Observation (SDO), in schools. It also presents literature on Direct Behavior Rating (DBR) and its application to conducting an FBA. The purpose of this study was to evaluate paraprofessionals' ability to generate believable data using DBR to assess functional relationships relative to SDO and to examine their perceptions regarding DBR's acceptability. In the study, paraprofessionals (n = 7) observed the behavior of 16 students in their classroom setting and then completed DBR ratings of disruptive behavior and four potential consequences (adult attention, peer attention, escape/avoidance, access to tangibles). Concurrently, an expert observer collected SDO data using a combination of a 15-second partial interval recording system and frequency count. Estimates of the true occurrence of disruptive behavior and the percentage of disruptive behavior met with each consequence were calculated based upon the SDO data. Differences in agreement between the DBR and SDO data were examined to obtain an understanding of the degree of association between the two observation methods. The results of this study indicated strong levels of agreement between the DBR and SDO data for overall disruptive behavior. However, despite a visual analysis of the data that suggested similar conclusions would be reached regarding the function of the disruptive behavior, weaker levels of agreement were found regarding the similarity of the data for each of the consequence targets. Results did indicate paraprofessionals perceived DBR to be an acceptable tool for collecting data related to functional contingencies. Guidelines for implementing DBR when conducting an FBA, limitations of the study, and suggested directions for future research are discussed.

Chapter 1: Introduction

Within Response to Intervention and Positive Behavior Intervention and Support frameworks, decisions regarding interventions for interfering behavior are often based on data collected as part of a Functional Behavioral Assessment (Losinski, Maag, Katsiyannis, & Ryan, 2015). Functional Behavioral Assessment (FBA) constitutes a wide range of observation procedures used to assess and analyze behavior that interferes with a student's success in school (Kilgus, Kazmerski, Taylor, & von der Embse, 2016; Steege & Watson, 2009). FBA procedures, rooted in Applied Behavioral Analysis, are used to describe the relationship between environmental contingencies that occasion and maintain interfering behaviors (Cooper, Heron, & Heward, 2007; McIntosh, Brown, & Borgmeier, 2008; Michael, 2007; Olympia, Heathfield, Jenson, & Clark, 2002; Steege & Watson, 2009). In order to understand these contingencies, school personnel observe and record data regarding the magnitude of the behavior, as well as the antecedents and consequences present in the internal and external environments of the student (Carr, 1994). A variety of observation procedures exist for gathering this data including indirect methods (e.g., ratings, interviews), direct methods (e.g., observation), and experimental manipulations (e.g., functional analysis) (Johnston & Pennypacker, 2009). Despite a strong empirical basis supporting the use of FBA procedures for addressing student interfering behavior, many questions remain concerning the practicality of performing FBA in schools (Gable, Park, & Scott, 2014). For instance, commonly identified barriers to broad use of FBA include the time and resource-intensiveness of the assessment's observation procedures, as well as the expertise needed to analyze the data (Kilgus et al., 2016; Gable et al., 2014). Given best practice guidelines and federal mandates, schools must assess the functional relationship between interfering behavior and the environmental contingencies with which it is related (Steege &

Watson, 2009). Therefore, there is an established call for research that recognizes the need for practical and acceptable methods of observation useful for when school personnel are collecting data regarding functional relationships.

Functional Behavioral Assessment

Functional Behavior Assessment (FBA) is an investigative process of gathering information to determine the purpose, or function, of a student's problem behavior (Cooper et al., 2007). With its aim on improving student success in the classroom, FBA procedures are used to generate a hypothesis regarding the nature of the functional relationship between variables in the environment and the interfering behavior of the student (Daley, Witt, Martens, & Dool, 1997; Gable et al., 2014; Lee, Sugai, & Horner, 1999). Observation is used to gather data regarding the antecedents and consequences related to the occurrence of the interfering behavior (Alberto & Troutman, 2003; Cooper et al., 2007; Daley et al., 1997; Filter & Horner, 2009; Lee et al., 1999; Michael 2007; Skinner, 1953). A main assumption of FBA is that interventions based upon the function of the interfering behavior will lead to a greater likelihood of producing the intended effect, which is often a reduction of the behavior of concern (Barton-Arwood, Wehby, Gunter, & Lane, 2003). The FBA process consists of many steps including, but not limited to, target behavior identification, observation and recording of data, visual analysis of the data, and development of function-based interventions (Daley, et al., 1997; O'Neill, Horner, Albin, Storey, and Sprague, 1996). While all the steps are crucial, data collection is among the most critical as it serves as the basis of forming the hypothesis and is often used as a baseline for determining later response to intervention.

As noted by Johnston and Pennypacker (2009), a variety of procedures exist for observing and recording data regarding the behavioral contingencies related to the student's

interfering behavior. Commonly used observation procedures include indirect methods of behavioral assessment [i.e., Functional Assessment Interview (FAI)] in which data is gathered through interviews, rating scales, and/or anecdotal reports from those who have witnessed the behaviors (Johnston & Pennypacker, 2009; O'Neil et al., 1996, pg. 18). Benefits of indirect assessment include aiding in defining target behaviors, gauging the intensity of the behavior, and identifying key observation opportunities (O'Neil et al., 1997). Indirect assessments are traditionally less time consuming (than more direct methods), require limited expertise, and consume minimal resources (Gable et al., 2014). Given this level of efficiency, school personnel often resort to indirect methods of assessment (Gable et al., 2014). However, criticisms of indirect methods include the potential for error related to reporter bias as well as failure to capture contextual variables.

Other forms of behavioral assessment include direct methods of observation (i.e., Systematic Direct Observation) in which data is gathered through direct observation of the interfering behavior within the context it typically occurs (Johnston & Pennypacker, 2009; O'Neil et al., 1996). Benefits of direct forms of behavior assessment include their repeatability, resistance to error, and strong empirical support within the literature base (Shapiro & Kratchowill, 2000). However, the accuracy of direct methods is often a product of the duration of the observation interval, the training of the observer, and the frequency of the behavior (Olympia et al., 2002). As such, traditional methods of direct observation carry intense time and resource requirements to perform accurately. Notwithstanding these challenges, direct observation of behavior has long been considered the foundation of behavioral assessment (O'Neil et al., 1996; Steege and Watson, 2009).

Overview of systematic, direct observation procedures

As noted by Shapiro and Kratchowill (2000), direct observation of student behavior is an essential component of data collection methodology and problem-solving strategies in schools. Systematic Direct Observation (SDO) involves standardized procedures used for continuously observing data relevant to an identified, operationally defined behavior, under naturally occurring conditions (Cooper et al., 2007; Chafouleas, 2011; Hinze & Matthews, 2004). SDO procedures include continuous (i.e., frequency, duration, latency) and discontinuous (i.e., partial-interval, whole- interval) observation methods (Cooper et al., 2007; Johnston & Pennypacker, 2009). Whereas continuous methods of observation target and record all occurrences of the behavior within a given time frame, discontinuous methods observe and record a sample of behavior performed in a given interval (Cooper et al., 2007; Fisk & Delmolino, 2012; Johnston & Pennypacker, 2009). Among their many strengths, both continuous and discontinuous methods are widely used by school personnel, do not require disruption to an individual's normal routines, demonstrate strong inter-rater agreement, sensitivity to behavior change, and allow for precision and repeatability (Briesch, Chafouleas, & Riley-Tilman, 2010; Hinze & Matthews, 2004; Miller, Chafouleas, Riley-Tilman, & Fabiano, 2014). Within the process of a FBA, direct, continuous assessment of interfering behavior, including its frequency as well as the antecedents and consequences with which it is connected, is considered essential to discovering functional relationships (O'Neil et al., 1996). Nevertheless, criticisms of traditional SDO procedures include the level of training needed, failure to capture the complexity of variables related to interfering behavior, and the time needed to perform them (Lewis, Scott, Wehby, & Willis, 2014).

Considerations when selecting an observation procedure

Given the empirical support for the use of indirect and direct behavioral assessment methods when completing FBA across various grades and disabilities, there are noted challenges to performing them in schools (Couvillion, Bullock, & Gable, 2009; Gable et al., 2014). Within the literature, there are noted semantic differences among definitions regarding components, techniques, and terms used to describe FBA data collection procedures (Losinski, Maag, & Katsiyannis, 2014). For instance, the terms contextual variables, setting events, antecedents, and establishing operations have all been used interchangeably to describe what occurs before the behavior of interest. Furthermore, not only has research not yet been able to define universally effective, efficient, and acceptable methods of data collection, research is mixed regarding the benefits of Behavior Support Plans based on the results of FBA. Moreover, there remains little consensus regarding the level of training needed to ensure school personnel can perform FBA adequately (Couvillion et al., 2009; Gable et al., 2014; Olympia et al., 2002).

In theory, selection of data collection procedures is guided by the nature of the interfering behavior (Cooper et al., 2007). In reality however, trade-offs between indirect and direct assessment methods are more likely to drive selection of a data collection method (Olympia et al., 2002). Research generated by the behavioral analytic community has declared direct assessment as the only method that allows for accurate observation of environmental variables (Chandler & Dahlquist, 2010). Research supports the use of direct methods when the time available for observation is limited and/or in situations that require precise rates of behavior be collected in order to inform decisions (Briesch, Chafouleas, & Riley-Tilman, 2010). In contrast, other research has found that simultaneous use of indirect and direct methods demonstrates higher levels of reliability, likely due to the combination of the capturing not only direct observation of the behavior but also others' perception of it (Scott & Kamps, 2007).

Nonetheless, there are no standards regarding which assessment method is best for a given specific behavior (et al., 2008). Furthermore, critics of SDO point out that its procedures require gathering multiple data points, allocating valuable resources to capture the data, and such specific training that it is often completed by experts external to the classroom (Briesch, Chafouleas, & Riley-Tilman, 2010; Chafouleas, 2011; Miller et al., 2014).

Literature on the barriers to efficient and accurate behavioral assessment identifies overall cost, the type and intensity of training required, and the need for external support among the many challenges schools face when gathering data related to student interfering behavior (Gresham, 2004). On a practical level, school psychologists encounter many challenges to observing and recording behavior. For instance, behavior that occurs at a low rate may be difficult to capture during limited opportunities to observe in the classroom (Steege, Davin, & Hathaway, 2001). It is then recommended teachers and classroom technicians observe and record interfering behavior. However, given other duties, they may find it difficult to reliably perform observation and recording procedures in a manner that produces valid results (Steege et al., 2001).

Collectively, while SDO can lead to the most descriptive data regarding student behavior, it often times is not a feasible option for schools. Therefore, research has been called for that investigates data collection methods that are time and resource-sensitive; yet still offer valid data (Gresham, 2004). As such, recent research has attempted to develop alternate methods that are designed to be less costly and more efficient, while maintaining acceptable levels of reliability and validity (Miller et al., 2014). In doing so, the goal is to develop methods more acceptable to schools (Miller et al., 2014).

Direct Behavior Ratings

Given the significant challenges associated with utilizing direct, systematic observation methods (i.e., Systematic Direct Observation of the frequency of the interfering behavior), schools tend to rely more on indirect methods (i.e., use of behavioral interviews) to collect data regarding interfering behavior (Gable et al., 2014). However, research has been largely mixed with regards to the reliability of indirect methods as scores often vary across time, rater, high- vs. low-incident behaviors, and population (Barton-Arwood et al., 2003). Therefore, researchers have attempted to create alternate approaches that include tools with high "defensibility and usability," referring to their level of acceptability by school staff (Chafouleas, 2011, p. 575).

Aiming to allow for repeated measure of response to behavioral intervention, one such approach is Direct Behavior Ratings (DBR), which has gained popularity in the past decade (Kilgus et al., 2016). Critical of the challenges of traditional SDO methods and careful to avoid the problems that afflict indirect methods (i.e., the rater bias associated with interviews and ratings), DBR is an attempt to not only assess the observer's perception of the magnitude and/or severity of the interfering behavior, but also to do so in a way that captures even slight changes over time (Chafouleas, 2011). According to its proponents, the data captured by DBR is used within a conceptual framework to link intervention with assessment and then gauge a student's responsiveness or resistance to behavior change (Miller et al., 2014). Described as a hybrid data collection method, DBR aims to combine the efficiency of indirect assessment methods with the opportunity to gather data within close proximity to the moment it occurs associated with direct assessment methods (Chafouleas, 2011). Current DBR methodology captures the perceptions of its user regarding the student's behavior as its foundation for making decisions regarding the effectiveness of interventions (Miller et al., 2014).

Within the Direct Behavior Rating data collection process, as described by Chafouleas (2011), three or fewer student behaviors are targeted and defined by the observer, often the classroom teacher, who then conducts observation during a set interval of time (i.e., 10 minutes, an hour, a half day). As immediately as possible following the conclusion of the interval, the observer rates the percentage of time the student engaged in the given behavior using a 10-point scale anchored on either end by opposite extremes of the behavior (i.e., never vs. always, 0% vs. 100%). Research into DBR has suggests it is an accurate and acceptable method of gaining information regarding student behavior (Chafouleas, 2011). DBR has been applied to gather information regarding students' mental health status, as well as to help assess functional relationships (Kilgus, 2016). However, depending on the nature of the identified behavior and the intended purpose of the data, guidelines for use of DBR vary (Briesch, Chafouleas, & Riley-Tilman, 2010). Furthermore, data must be recorded in close proximity to the termination of the recording interval in order to maintain acceptable levels of accuracy (Chafouleas, 2011).

Direct Behavior Rating was created in part to provide a data collection approach that is both cost- and time-sensitive while simultaneously attempting to maintain the usefulness of the data to inform intervention decisions (Chafouleas, 2011). However, DBR has only recently been explored and evaluated systematically as an alternate method to more traditional indirect and direct data collection methods in schools (Chafouleas, 2011; Miller et al., 2014). Research into DBR systems, such as that conducted by Volpe and Briesch (2012), has found that DBR offers information related to only the perceived magnitude of the student's behavior. Preliminary research by Kilgus and colleagues (2016) has only recently explored the application of DBR methodology when assessing functional relationships.

Despite research that supports both Systematic Direct Observation and Direct Behavior Ratings as accurate and acceptable methods for observing and recording student interfering behavior, questions remain regarding several aspects of conducting them in schools. The demands of the time, the effort, and the training of personnel required to perform SDO collection methods often exceeds the capacity of school personnel (Gable et al., 2014). Likewise, DBR offers the efficiency and simplicity of behavior rating scales, yet at the expense of the true direct observation of behaviors in the context in which they occur. As such, questions remain regarding DBR's use in assessing functional relationships. Therefore, additional research is needed that investigates application of DBR methodology within a functional behavior assessment problem-solving model.

The primary purpose of this study is to determine the believability of paraprofessionals' observations using Direct Behavior Ratings to assess functional relationships as compared to Systematic Direct Observation performed by an external observer. Further, this study aims to assess the paraprofessionals' perceptions regarding the acceptability of DBR as a behavioral observation tool.

Chapter 2: Literature Review

Despite a wealth of research regarding the usefulness of Functional Behavior Assessment to detect functional relationships between students' behavior and variables within the classroom environment in which it occurs, there remains limited consensus regarding the most effective and efficient methods for observing and recording behavioral data. Within the literature, challenges often associated with direct observation include the resources and time needed to gather the data, as well as the integrity with which the process is conducted (Kilgus et al., 2016; Gable et al., 2014). Nevertheless, legal mandates and best practices identify the data collected as part of the broader Functional Behavior Assessment process, as the best source of information when developing interventions within a behavioral support plan. As a result, researchers have focused on the development of school-based observation and data collection procedures that are as feasible and accurate as possible.

Functional Behavior Assessment

In order to understand the resource and time intensiveness of Functional Behavior Assessment (FBA), it is important to illustrate the components that make it a behavioral assessment technology. Cooper et al. describe Functional Behavior Assessment (FBA) as "a systematic method of assessment for obtaining information about the purposes (functions) a problem behavior serves for a person…" (2007, p. 696). Steege and Watson identify FBA as "a set of assessment procedures that results in the identification and description of the relationship between the unique characteristics of the individual and the contextual variables that trigger, motivate, and reinforce behavior." (2009, p. 7). Similar to other definitions, O'Neil and colleagues define Functional Behavior Assessment as a comprehensive process of investigation

crucial in devising behavior intervention plans (1997). When completed well, an FBA leads to a clearer understanding of variables that set the stage for and maintain problem behaviors.

Origins of Functional Behavior Assessment

While FBA methodology has grown into its modern incantation over the past 20 years, its origins trace back to E. L. Thorndike in the 1890's and his notion of the Law of Effect (Steege & Watson, 2009). Essentially, through Thorndike's investigation of animal intelligence by using cats and dogs, he observed that an animal's behavior came to be influenced by the consequences that followed it (Thorndike, 1898). Further influencing the development of FBA technology, the work of John B. Watson, considered the father of behaviorism, represented a shift from introspective psychology toward a greater focus on observable behavior (Steege & Watson, 2009). In 1920, Watson and Rayner's publication outlining their work with conditioning fear in an infant illustrated the influence certain environmental stimuli can play on behavior, particularly when those stimuli occur before the behavior (aka respondent conditioning) (Steege & Watson, 2009). By the 1930's, the experiments of B.F. Skinner studied not only the effects of consequences on the behavior of rats, but also the effects of events that preceded behavior. (Steege & Watson, 2009). Further, Skinner pioneered the notion that behavior can be defined in observable terms, measured in a variety of ways, and that relationships can be derived from the results (Skinner, 1953; Steege & Watson, 2009).

By the 1960's, research in behavioral assessment explored the relationship between environmental conditions both pre- and post-intervention (Couvillon et al., 2009). In 1968, in one of the earliest investigations of this topic, Bijou, Peterson, and Ault discussed the use of descriptive observation to record antecedents, behaviors, and consequences and the potential to use that information to determine functional relationships, aka the Three-Term Contingency

(Steege & Watson, 2009). Through establishing operational definitions of observable events and the environments in which they occurred, the problematic behaviors of a preschool student were observed and recorded. Further, Bijou et al. (1968) highlighted the importance of training observers in order to ensure the greatest likelihood of accuracy (i.e., observer reliability).

Research during the 1970's and 1980's mainly involved children and adults with severe disabilities and/or self-injurious behaviors (Couvillon et al., 2009). In 1982, research conducted by Iwata and others described the use of experimental manipulation of operant conditions to determine the functional relationship of self-injurious behaviors in nine individuals with what today would be known as intellectual disabilities. Considered a seminal study, through direct observation of the participants, the occurrence or non-occurrence of target behaviors were recorded during 10-second intervals (Iwata, Dorsey, Slifer, Bauman, & Richman, 1982). Participants were exposed to four separate conditions in which the behavior of the researcher varied. For instance, during the first condition referred to as the Social Condition, the researcher provided social attention upon the occurrence of self-injury (Iwata et al., 1982). In the second condition, an academic demand was removed once the participant began exhibiting self-injurious behavior (Academic Demand Condition). During condition three which involved unstructured play activities, attention was offered non-contingently regardless of the behavior (Unstructured Play Condition). In contrast, the participant did not receive attention or have any demands placed upon them during the fourth condition (Alone Condition).

Based upon their observations, Iwata and his colleagues found different functions, or purposes, for self-injurious behavior across the participants in the study (Iwata et al., 1982). Observations revealed the mean rate at which SIB occurred varied greatly among the participants, ranging from a low of 4.5% to a high of 91.2%. Despite this range, distinct mean

rates of SIB were observed within participants across the four conditions of the study. For instance, all eight participants demonstrated low rates of SIB during the unstructured play condition. For four of the participants, SIB was highest in the Alone condition, whereas for two others participants, SIB was highest in the Academic Demand condition and for one participant, SIB was highest in the Social condition. Collectively, the results from the Iwata study laid the foundation for development of modern Functional Behavior Assessment and Functional Analysis procedures. Seeking to extend Iwata and his colleagues' findings, additional research during the 1980's focused on the behavioral outcomes of systematically manipulating antecedents and consequences (Olympia et al., 2002).

By the late 1990's/early 2000's, research into FBA methodology evolved to include indirect methods of assessment (i.e., interviews conducted with those familiar with the behavior) as well as a variety of direct methods (i.e., direct observation of behavior) (Olympia et al., 2002). However, theses studies typically included individuals with intellectual disabilities and were often performed within clinical settings. Meanwhile, research that focused on the application of FBA methodology when assessing student behavior in classrooms remained limited (Couvillon et al., 2009). Couvillon and colleagues outlined several factors contributing to this limitation (2009). Among them, the legislation mandating FBA be performed within schools lacked specificity regarding its use. Furthermore, the nature of the interfering behavior for which FBA was designed (i.e., high-risk behaviors such as banging one's head) restricted applied research into its use. Moreover, the narrow competency of school personnel to perform FBA contributed to minimal advances within the literature (Couvillon et al., 2009).

The Individuals with Disabilities Education Act (Public Law 105-17) stipulated that school-based intervention teams must develop behavioral intervention plans, based upon

information collected via an FBA, to address student interfering behaviors, such as those identified in relation to a manifest determination, or when a student engaged in drug or even weapon use, or if a student had been suspended for more than 10 days. (Olympia et al., 2002; Steege & Watson, 2009). Within the Individuals with Disabilities Education Improvement Act (Public Law 108-446), the reasons schools performed FBAs were expanded and strengthened, including situations that involve a change in a student's placement. Further, IDEIA identified timeframes in which they are to be completed, typically within 30 days (Steege & Watson, 2009).

The remainder of the chapter that follows describes the rationale behind Functional Behavior Assessment, traditional methods for observing and recording behavioral data, research regarding FBA in schools, as well as the implications of Systematic Direct Observation on collecting behavioral data. Additionally, this chapter focuses on identifying the reported accuracy and acceptability of an alternate data collection procedure (i.e., Direct Behavior Ratings). The review concludes with suggestions for future research in the area of reliable and acceptable behavioral data collection methods to assess functional relationships.

Purpose of Functional Behavior Assessment

Within a problem-solving model, Functional Behavior Assessment (FBA) is commonly thought to be a pre-intervention assessment that leads to the development of a hypothesis regarding the relationship between the student's behavior and the variables within the student's environment. (Anderson, Rodriguez, & Campbell, 2015). Data collected during the process of investigating this relationship is analyzed and used to inform intervention planning (Cooper et al., 2007; O'Neil et al., 1997; Steege & Watson, 2009).

The term FBA represents a wide range of procedures designed to detect identifiable aspects of a student's environment that evoke and maintain interfering behavior. (Anderson, Rodriguez, & Campbell, 2015) The FBA process is made up of many steps including, but not limited to: target behavior identification and selection, observation and recording of data, visual analysis of the data, and development of function-based interventions (Daley et al., 1997; O'Neill et al., 1997). As described by O'Neil and colleagues, the FBA process establishes a description of the behavior(s) that interferes within the academic setting. This description allows the interfering behaviors to be the focus of the assessment procedures (1977). These procedures lead to the identification of variables that both predict the occurrence of the interfering behaviors as well as serve to maintain them. As a result, the FBA produces summary statements that explain the suspected functional relationship between these variables, specific to the individual (O'Neil et al., 1997).

Early approaches to altering interfering behavior focused on modifying factors associated with the behavior (Dunlap & Fox, 2011). Later, with the acceptance that all behavior serves a function, behavioral approaches sought to identify those functions, leading to the emergence of function-based interventions (Dunlap & Fox, 2011). In order to best inform interventions, an FBA must include observation of the individual's behavior within the context it occurs, thereby allowing data to be gathered that supports the hypothesized relationship (O'Neil et al., 2007). Data used to complete school-based FBAs are often gathered via indirect and direct, descriptive assessment procedures. (Alter et al., 2008).

Observation of behavioral contingencies

Within FBA technology, traditional direct observation seeks to observe the interplay between the antecedents and consequences related to the behavior of concern (Cooper et al.,

2009; O'Neil et al., 1997; Steege & Watson, 2009). The student's behavior comes to be controlled by the consequences that maintain the interfering behavior (Dunlap & Fox, 2011; Michael, 2007). Interfering behaviors are preceded by antecedents that signal a greater likelihood of the behavior occurring in their presence (Dunlap & Fox, 2011; Michael, 2007). For example, if a student who finds math worksheets very difficult can reliably predict he will be able to avoid the work (consequence) if he dumps the contents of his desk (behavior), then he will likely exhibit the behavior when given a math worksheet (antecedent). Observation is used to document the events that occurred before (antecedents) and after (consequences) the target behavior (Cooper et al., 2007). Antecedent variables are often found to be uniquely and highly specific across individuals' reactivity to a variety of stimuli including social, physical, ecological, physiological, transitional, and instructional conditions (Cooper et al., 2007; Dunlap & Fox, 2011).

Data collected as part of the assessment process can help establish if patterns exist regarding the times or settings the behavior is more likely to occur (Cooper et al., 2007; O'Neil et al., 1997). Likewise, data can reveal if the interfering behavior is related to the presence or absence of certain individuals with whom the individual may or may not be familiar. As importantly, data can also lead to a determination of consequences that may serve to influence the individual's behavior (Cooper et al., 2007; O'Neil et al., 1997). Based on the observation of these antecedents and consequences that maintain the interfering behavior, a hypothesis can be generated to describe the nature of these relationships (Alberto & Troutman, 2003). In identifying the consequences (function) of the interfering behavior, interventions can be designed to manipulate those consequences in a functionally equivalent way in order to produce a change in behavior. (Daley et al., 1997; Dunlap & Fox, 2011). For instance, in the case of the desk-

dumping student, a functionally equivalent intervention package may include teaching the student ways to request an alternate assignment. A behavior support plan, containing strategies to render the interfering behavior ineffective, inefficient, and irrelevant, is designed to decrease target behavior while simultaneously aiming to increase replacement behaviors (Cooper et al., 2007; McIntosh et al., 2008). However, regardless of the strong empirical support for using FBA methodology to determine behavior contingencies maintaining interfering behavior, without a reliable, efficient, and acceptable method for collecting data describing the functional relationship, these interventions are likely to be minimally effective (Alter et al., 2008).

Data recording procedures

A variety of observation procedures exist for gathering data regarding behavioral contingencies including indirect methods (e. g., ratings, interviews), direct methods (e. g., observation), and experimental manipulations (e. g., functional analysis) (Olympia et al., 2002; Steege & Watson, 2009). Indirect methods of data collection (i.e., interview, questionnaires, rating scales, record reviews) require no direct observation of the student's interfering behavior (Alter et al., 2008). Ratings and interviews are designed to gather information about the behaviors of concern from those who have witnessed the behavior, along with its associated contextual variables, and are assumed to know it well (Olympia et al., 2002; O'Neil et al., 1997). Indirect assessment methods offer potential benefits including assisting with defining interfering behaviors, gaining a measure of their intensity, as well as identifying key observation opportunities (Floyd, Phaneuf, & Wlczynski, 2005). They are less time consuming; require little experience, limited expertise, and minimal staff collaboration. Indirect assessment methods are thought to be most effective with high rate, low intensity behaviors (i.e., off-task behavior), particularly when paired within a multi-informant, multi-source approach to assessment (Gable

et al., 2014). Often, indirect assessment methods are an attractive option for collecting data within schools given these benefits.

Research has been mixed regarding the accuracy of indirect methods in identifying functional relationships. In one study, Alter et al. (2008) compared the function identified via indirect assessment methods to the function identified via direct assessment methods obtained by 12 observers, in four young children (ages 4 – 6 years old) for whom functional analysis was used to establish the true functional relationship. Each child was observed using either the Functional Assessment Interview (FAI), Motivation Assessment Scale (MAS), both considered indirect assessments by the researchers, or an Antecedent-Behavior-Consequence Observation (ABC) sheet, considered the direct assessment measure (Alter, et al., 2008). Results indicated that across the indirect and direct measures, inter-observer agreement was inconsistent (MAS-ABC agreement for 2/4 observers, FAI-MAS and FAI-ABC agreement for 1/4 observers). However, results also indicated inter-observer agreement data was calculated to be an average of 87% to 97% across the four participants when data was collected via an ABC observation worksheet (Alter et al., 2008).

Further, for seven of the 12 participants (56%), there was agreement regarding function among the assessment and the Functional Analysis (Alter et al., 2008). Findings indicated outcomes derived from the FAI and MAS agreed with the outcome of the FA just over half of the time (57%). Agreement between indirect (FAI, MAS) and direct measures (ABC) was observed in only three out of 8 participants (38%). Noteworthy, the function identified via results from the ABC observation worksheet corresponded with the FA 100% of the time.

McIntosh and colleges (2008) reviewed 10 studies, representing 41 participants that used the Functional Assessment Checklist for Teachers and Staff (FACTS) as the reported indirect

assessment method. In their review, the researchers attempted to establish the reliability and validity of FACTS by assessing its technical adequacy among the studies. Upon finishing their analysis, they reported finding strong test-retest reliability and inter-observer agreement as reported in the studies. Further, the findings suggested strong content validity; however, they found no evidence of process validity (McIntosh et al., 2008). McIntosh et al. concluded the FACTS offered a reliable and valid method of assessment when used within a comprehensive approach to FBA. In other words, despite finding support that FACTS offered reliable data regarding functional relationships, the researchers noted it best utilized as a component of a full FBA. However, the number of studies (10) and the extensive training and familiarity of the researchers in use of the FACTS limited the review's findings.

Schools often resort to indirect assessment methods given advantages in terms of efficiency (Gable et al., 2014). However, indirect assessment methods rely too heavily on rater memory, are therefore subject to rater bias, and often require inclusion of direct assessment methods anyway. Moreover, as this form of assessment is dependent upon informant report, it is prone to error (O'Neil et al., 1997).

Direct observation procedures, such as Systematic Direct Observation (SDO), have long been considered an essential component of behavioral assessment (Steege & Watson, 2009). Direct assessment methods of data collection require direct observation of the student's interfering behavior within the context in which it occurs. (Alter et al., 2008). Direct observation provides actual contact with the behavior as it occurs within the natural environment, thus greatly reducing the chance of observer bias (which may potentially influence conclusions when using interviews and behavior rating scales) (Olympia et al., 2002). However, the

accurateness of direct observation is often a product of the duration of observation session, the frequency of the behaviors, and the simplicity of the coding procedures (Olympia et al., 2002).

In contrast to indirect methods and direct observation, Functional Analysis involves the direct manipulation of variables (antecedent and consequence) thought to occasion and maintain the interfering behaviors. This manipulation is done in a systematic manner as to permit the detection of the functional relationship (Alter et al., 2008).

Research regarding FBA in school settings

Within schools, data gathered as part of the FBA process is used to generate a hypothesis as to the behavior's purpose. The process typically includes gathering information from multiple sources and in a variety of ways, and using that data to generate a hypothesis about the behavior's function (Losinski et al., 2015). In the end, this process is designed to help school-based intervention teams develop meaningful interventions that reduce or eliminate the interfering behavior and replace it with a more acceptable behavior that serves a similar purpose.

Several studies have investigated the use of school-based Functional Behavior Assessments technology, declaring it a suitable method for addressing challenging behavior (Gable et al., 2014). Early investigations into Functional Behavior Assessment were focused mainly on individuals with severe disabilities, casting doubt on its utility within the classroom (Gable et al., 2014). However, over the last 10 – 15 years, studies have expanded to include students with and without disabilities within school settings. Anderson and colleagues (2015) completed a review of 233 articles published prior to 2013 that utilized either an indirect, descriptive (direct), or experimental method of FBA within a school setting. A total of 640 participants were represented in these studies. Among their review of the research, they found evidence that the yearly average number of studies published on FBA conducted in school

settings has increased over the past three decades. From the period 1981 - 1990, an average of 1 study was published each year. This average increased to 7.3 per year from 1991 - 2001. From 2001 - 2013, that average rose to 11.9 per year. In over half of the studies, participants carried a diagnosis of Autism Spectrum Disorder or Intellectual Disability. Self-injury, elopement, stereotypy, physical aggression, tantrums, or inappropriate vocalizations were more likely to be identified as the interfering behavior for students with ASD or ID as compared to students without either diagnosis (Anderson et al., 2015).

Of this published research on school-based FBAs, 56.7% included use of an indirect form of assessment, 49.3% include use of a direct method of FBA (Anderson et al., 2015). Indirect methods as the sole form of assessment were reported in only 10% of the studies whereas 80% of the studies included a form of descriptive (direct) assessment. In 63.1% of the published studies, researchers used experimental methods (i.e., Functional Analysis) in order to confirm functional relationships. Typically, indirect assessment methods preceded direct methods, both of which were conducted before experimental methods. Of the research using direct methods of FBA, 88.7% used ABC recording (Anderson et al., 2015). Direct methods of FBA were mostly likely to be performed by teachers or other educators.

For descriptive (direct) methods of FBA, 35.9% identified socially mediated, positive reinforcement (i.e., attention, given access) as the maintaining function of the problem behavior and 25.6% identified social-negative reinforcement (i.e., escape) as the function (Anderson et al., 2015). Multiple reinforcers were identified as the function in 14.6% of the studies. There were 212 antecedent-behavior-consequences relation outcomes identified for direct FBA methods. Typically, attention was the most often identified function (33.8%). Anderson et al. (2015)

concluded that researchers interested in school-based FBA often used direct assessment methods as a precursor to more rigorous, experimental methods of FBA.

Not only has research into the use of FBA technology in schools influenced best practice guidelines governing its use, research has also attempted to explore the outcomes of function-based vs. non-function-based interventions. For students with externalizing behaviors (i.e., disruptive behaviors, aggression, poor social skills), a multifaceted approach to application of this traditional notion of FBA has been demonstrated to be effective (Olympia et al., 2002). The approach assumes that interventions based upon determination of the relationship between the antecedents and consequences will be more effective than those not informed by this information (Barton-Arwood, et al., 2003). McIntosh and colleagues (2008) explored research examining the effects of function-based interventions within general education classrooms. Based upon their results, function-based interventions were more successful in the reduction of interfering student behavior than non-function based interventions.

The literature base supports the functional connection between students' interfering behavior and their academic problems. In the case of behavior that is maintained by escape from demands, the academic tasks may be beyond the student's skill level (i.e., a math task that is to difficult). Similarly, if the academic task is too easy, the interfering behavior may also serve to allow the student to escape. Academic problems can also be functionally related to too little or too much attention from a teacher.

In a study by Filter and Horner (2009), the problem behaviors (i.e., out-of-seat, talking out) and task engagement of two 4th grade students were observed via an interval recording observation system. Results of an FBA suggested that for one of the students, interfering behavior followed presentation of a reading task above the student's instructional level. The

other student's behavior was preceded by the support provided by his teacher that allowed him to avoid failing at a difficult academic task. Following confirmation of the functional relationships via use of a Functional Analysis (FA), each student was provided a function-based intervention consistent with the results of the FA (Filter & Horner, 2009). Each student was also exposed to a non-function-based intervention. Results suggested that for one of the students, interfering behaviors occurred at a significantly higher percentage of intervals during baseline (99.3% of intervals) as compared to the intervals during which the student received the non-function based intervention (13.1% of intervals). However, an even greater reduction was observed during the function-based intervention phase (0.1% of intervals). Task engagement was higher during the function-based intervention phase than either the non-function based intervention or baseline phases (76.6%, 69.2%, 43.2%, respectively). Interfering behaviors reduced from 28.2% of intervals during baseline down to 21% during the non-function based phase for the second student (Filter & Horner, 2009). Better yet, it reduced to 3% of intervals during the function-based phase. Task engagement was similar during baseline (51%) and non-function based intervention (56%); however it increased to above 90% of intervals when functional based interventions were implemented (Filter & Horner, 2009). Results such as these have helped establish a growing literature base that indicates function based interventions lead to more meaningful changes in student interfering behaviors than do non-function based interventions.

Several meta-analyses have found function-based interventions implemented within a Positive Behavior Support system to produce better outcomes (i.e. reduction in interfering behaviors) as compared to non-function based interventions. For instance, with regard to reductions in interfering behaviors, Goh and Bambara (2012) found moderate effect sizes across 83 studies published between 1997-2008. Moderate effect sizes were also noted with regard to

increasing students' use of appropriate skills. They also found evidence that these outcomes maintained anywhere in the range of 1 week up to 2 years. Goh and Bambara (2012) also found moderate effect sizes across gender and grade level. For instance, larger effect sizes were noted for high school students with moderate effect sizes noted for elementary and middle school students. Compared to moderate effect sizes for students with developmental disabilities or no disabilities, larger effect sizes were found for students diagnosed with non-developmental disabilities (i.e., Learning Disability, ADHD). Moreover, larger effect sizes were noted for general education classrooms as compared to moderate effect sizes for special education and combined special and general education classrooms (Goh & Bambara, 2012). However, no significant difference was found in the any of the analyses across diagnosis or classroom setting categories. As Goh and Bambara point out, these results suggest that interventions based on the results of an FBA are effective across a wide range of students, grades, and disabilities (2012). Nonetheless, despite research establishing function-based interventions as more effective than non-function based interventions, schools continue to select interventions based on "familiarity, convenience, or a finite amount of time and resources" (Gable et al., 2014, pg 117).

Barriers to conducting FBA in schools

Dunlap and Fox (2011) argue that the complexity of the systems that serve children contributes to the difficulty schools face when attempting to implement FBA technology procedures (2011). Moreover, schools often face doing so with personnel who lack adequate training and experience with implementing function-based interventions. In order to maximize the potential of function-based interventions, school require reliable and acceptable methods for gathering functional assessment data on an on-going basis (Dunlap & Fox, 2011).

Gable and colleagues (2014) note several questions remain regarding the use of FBA in schools. They suggest that behavior intervention plans developed through the FBA process do not reliably lead to behavior change (Gable et al., 2014). Further, the ability of school-based teams to correctly implement FBA technology remains suspect as well, particularly when performed independent of external support. For instance, in a 2009 study of 134 educators, participants took part in an online survey and were asked to indicate their level of training in FBA/BIP (Couvillon et al., 2009). For teachers with one to five years of experience, only 6% of respondents had received any type of training in the use of FBA procedures. In contrast, for those respondents with a decade or more of experience, 62% had some form of training in FBA (Couvillon et al., 2009).

Of particular concern is the effectiveness of the training of school personnel to perform FBAs (Gable et al., 2014). There is little consensus in the research to suggest which methodologies should be taught to school personnel. For instance, McCahill, Healy, Lydon, and Ramey conducted a literature review of the effectiveness of training in FBA for educational staff (2014). In all, FBAs were conducted with 112 students across the twenty-five studies included in the review. Training packages for the school personnel included instruction in indirect (i.e., interviews), direct (i.e. Systematic Direct Observation), and experimental assessment (i.e., Functional Analysis) methodologies, but not in all three. The results indicated the reported effectiveness of any given training package on the perceived skill acquisition of the participants was similar across all training conditions (McCahill, et al., 2014).

Traditional approaches to FBA can place significant drain on school district's resources and personnel (Gable et al., 2014). Criticisms of FBA in schools include the time needed for sufficient observation, skill level of the observer, and applicability in schools (Olympia et al.,

2002). Direct observation techniques can be affected by the reactivity of the student and/or bias of the observer (Miller et al., 2014). Further, there exists a wide range of definitions and characteristics of the various steps and components of FBA methods (Losiniski, Maag, & Katsiyannis, 2014). Despite these types of barriers impacting the application of FBA for use in school settings, additional research over the past decade has demonstrated greater support for the use of FBA as an effective technology for observing and addressing a wide range of behaviors in students with and without disabilities (Couvillon et al., 2009; Gable et al., 2014; Losinski et al., 2015; Steege & Watson, 2009).

Systematic Direct Observation

Given the substantial empirical base establishing the usefulness of FBA in addressing interfering student behavior, the review will now consider the foundation of behavioral assessment – Systematic Direct Observation. Johnston and Pennypacker define measurement as "assigning numbers and units to particular features of objects or events" (2009). Systematic Direction Observation (SDO) has traditionally been the most trusted method of measuring behavior (Briesh et al., 2010). SDO is meant to observe specific pre-determined features of behavior and involves the identification of strategic periods of time for observing and recording behaviors using standardized procedures. (Hintze, Volpe, & Shapiro, 2002). Direct observation of student behavior has long formed the cornerstone of behavioral assessment (Shapiro & Kratochwill, 2000). SDO procedures require that a target behavior be clearly defined in observable and measurable terms. The onset and offset of the behavior should be clear to any observer (Hintze et al., 2002). Given the precision of SDO, student behavior can be repeatedly measured over time (Miller et al., 2014).

Less focused in its procedures, naturalistic observation differs from SDO in that the observer records a variety of potentially related and non-related variables across what ever happens during the observation period (Riley-Tilman, Kalberer, Chafouleas, 2005). Given the technical adequacy required to complete SDO, it typically requires external classroom personnel to complete (Chafouleas, 2011). Systematic direct measurement of student behavior has traditionally served three purposes: to diagnosis the nature of a problem, to monitor response to intervention, and to carry out research (Lewis et al., 2014).

SDO observation procedures

A variety of tools, procedures, and methods exist for collecting data via Systematic Direct Observation. Data collection aimed at informing diagnostic decisions is often gathered via event or duration recording (Lewis et al., 2014). In event recording, every incident of the behavior is captured and recorded (Johnston & Pennypacker, 2009). For example, event recording may be used when observing and recording every time a student dumps the contents of his desk. In duration recording, the time between the onset of the behavior and its offset is captured and recorded. As an example, duration recording may be used to document the amount of time a student engages in tantrum-like behavior in a pre-school. Among the literature base, several variations of these observation and recording procedures have been used with even more modifications having been implemented across clinical settings (Lewis et al., 2014). Event recording is considered best practice for behaviors with a clear beginning and end (Hinze et al., 2002; Johnston & Pennypacker, 2009). For instance, event recording is best used for behaviors such as hand-raising, getting out of one's seat, or even hitting. However, even if the behavior is discrete, if it occurs at a high-frequency, it may be unrealistic to capture every single occurrence

(Johnston & Pennypacker, 2009). For instance, capturing pencil tapping may be too difficult to perform accurately (Hintze et al., 2002).

Interval-based observation is another common observation and recording procedure (Johnston & Pennypacker, 2009). Within these systems, observers record "whether the target behavior is present or absent after a prescribed time interval passes." (Lewis et al., 2014, pg 192). Three common types of interval-based procedures include: whole interval, in which the target behavior must occur for the entire interval; partial interval, in which target behavior must occur at least once during the interval; and momentary time sampling, in which the target behavior must occur at a specific time within the interval (i.e., at the termination of the interval) (Johnston & Pennypacker, 2009; Lewis et al., 2014). Typically, observation periods are divided into equal time intervals (i.e., 10-second interval; 10-minute interval). The total number of intervals in which the behavior occurred is multiplied by 100 and then divided by the total number of intervals the behavior was observed in order to obtain the percentage of intervals the behavior occurred, or occurrence agreement percentage (House, House, & Campbell, 1981; Johnston & Pennypacker, 2009).

Research supports several procedures for increasing the likelihood SDO data accurately represents the behavior observed (Olympia et al., 2002). Among those procedures, the duration of the interval and the time period in which those intervals are collected has been demonstrated to have an impact on observer reliability. When using an interval-based data collection procedure such as partial-interval recording, shorter intervals (i.e., 5 – 10 seconds) over a minimum of 15 minutes per observational window have generally been found to be accurate (Olympia et al., 2002). Even shorter intervals (i.e., less than 5 seconds) would likely make the mechanics of recording the data extremely cumbersome and difficult, thus affecting accuracy of

the data. Moreover, for behavior occurring this frequently, observing using a frequency procedure (i.e., count of every occurrence) would seem to be a reasonable option.

Ideally, selection of direct observation procedures should be matched to the interfering behavior targeted for observation (Johnston & Pennypacker, 2009; Olympia et al, 2002). For instance, high-frequency behaviors, such as hand flapping, should be observed and recorded via interval procedures. Low-frequency behaviors, such as discrete episodes of destruction of property, should be tracked using event-recording procedures (Lewis et al., 2014).

Continuous method vs. discontinuous methods

Continuous observation procedures are meant to detect and record every single occurrence of a target behavior such as afforded by event recording (Johnston & Pennypacker, 2009). For instance, an observer may be interested in capturing the number of times a student raises his hand during a given class period. This type of observation can be potentially very accurate given the discrete nature and typically low frequent of hand-raising. In contrast, discontinuous observation procedures are meant to detect only a sample of the target behaviors within a given observation period (Johnston & Pennypacker, 2009). To illustrate, an observer may be interested in the number of pro-social statements made by a student while on the playground. Depending on the social skills of the student and the length of playtime, it may be difficult to capture each pro-social statement. Dividing the playtime into intervals and recording whether or not the statement was made at any point may be preferable in this type of situation. Typically, discontinuous observation procedures involve interval-recording strategies (Johnston & Pennypacker, 2009).

Continuous methods capture every instance of a behavior (i.e., event recording, duration, latency) and are best used when the behavior is discrete with a clear beginning and end.

Discontinuous methods do not capture each instance and instead offer an estimate of the occurrence of the behavior. These methods observe and record only some instance of the behavior (i.e., partial-interval, whole-interval, momentary time sampling). They are best used when behavior is variable and not marked by a clear beginning and end. Collectively, these various methods make up what is known as Systematic Direct Observation and are considered the gold standard of behavioral observation (Johnston & Pennypacker, 2009). While SDO is considered the most accurate method for gathering behavioral data, it has also been shown to be the most heavily resource dependent.

Challenges to gathering SDO data

As noted within the literature base regarding conducting Functional Behavior Assessments, gathering data via Systematic Direct Observation methods can be very challenging within a school (Gable et al., 2014). Even trained observers can vary in their performance of SDO. Threats to observation accuracy include observer drift (i.e., gradual change in criteria across observations) and measurement reactivity (i.e., change in student behavior as a result of being aware of the observation) (Johnston & Pennypacker, 2009). Observation of the contextual variables related to the interfering behavior (i.e., time of day, location, antecedents) can be difficult due to the complexity of the environment (Lewis et al., 2014). For instance, when observing for antecedents and consequences within a classroom environment, the number of variables that could serve as potential variables is nearly endless. As such, using this strategy, or any other, as the sole source of information is likely to lead to compromised data at best (Lewis et al., 2014). Within schools, SDO requires a significant amount of resources to perform (Riley-Tilman, Kablerer, & Chafouleas, 2005). Observation sessions not only consume the time to conduct them, but there is also a significant amount of time devoted to preparing to conduct

them. As best practice mandates multiple observations, the cumulative impact on available resources increases exponentially. Moreover, as noted by Briesh et al. (2010) inferences based upon SDO are limited by the restricted scope of the observation window (i.e., a 10 minute block of time). Therefore, in order to better understand interfering behavior within the context it occurs, strategies that offer feasible solutions must be discovered (Lewis et al., 2014).

Observers within a classroom

Not only does SDO require time, it also requires an observer dedicated to recording occurrences of the behavior -- at the expense of doing anything else. This increases the chances for measurement reactivity (Riley-Tilman et al., 2005). However, schools must nonetheless collect the data using the personnel available to them. Teachers and paraprofessionals (i.e., classroom aides, paraeductors, behavior technicians) are among potential observers already in the classroom. Interestingly, the role served by paraprofessionals may lend its self to collecting data. The job title 'Paraprofessional' represents individuals who work in a teacher-related position within a school, often performing a variety of duties and fulfilling a number of responsibilities (NEA, 2005). These key components of the educational system in the United States serve as supports for certified/licensed teachers in delivering instructional and other related services. By reducing the number of students to adults in the classroom, teachers can provide more specialized instruction and more individualized intervention. Among their responsibilities, paraprofessionals are responsible for the overall conduct and behavior in the classroom. They play an essential role in monitoring student progress and response to intervention (NEA, 2005). Special education service delivery relies on paraprofessionals for a variety of roles including monitoring behavioral data (Giangreco & Broer, 2005).

Research regarding the use of paraprofessionals in schools has indicated some interesting findings. For instance, for some school districts across the United States, the model of service delivery included more paraprofessionals than special education teachers (Suter & Giangreco, 2009). Nationwide, the ratio of paraprofessionals to teachers has increased over the past decade. According to data from the 2003-2004 school year, the ratio of special education paraprofessionals to students nationally was approximately 1 for every 17 students (Giangreco, Hurley, & Suter, 2009). There was approximately less than one special education paraprofessional Full Time Equivalent–FTE in the United States for every special education teacher FTE (0.9). However, more recent trends in the data indicate there are now more paraprofessionals nationwide than teachers (U.S. Dept of Education, 2016). According to data gathered from the 2013-2014 school year, for every 1 student, there are approximately 15 paraprofessionals. Furthermore, for every special education teacher (FTE) position in the United States, there was slightly more than one special education paraprofessional (1.12).

Giangreco and Broer (2005) argued that some research paints the extensive use of paraprofessionals in a positive light as it allows the special education teacher to be more of a coach and manager within the classroom. However, they also pointed that other research does not share such a positive view citing the findings of several studies that view the special education teacher as the most qualified to meet the student's needs. For instance, one study found special education teacher spent significantly less time on instruction than did the special education paraprofessionals they supervised (Giangreco & Broer, 2005). In a national survey of 737 school professionals, including 153 special education paraprofessionals, results indicated that 19% of paraprofessionals' time was devoted to behavioral support. Nearly half of their day (47%) was devoted to providing instruction to students and less than 5% was spent on activities

not involving direct student contact (Giangreco & Broer, 2005). In a separate survey, Giangreco (2007) found that 53% of paraprofessionals agreed the special education teacher relied on them for information when completing progress reports regarding student performance in the classroom.

These findings suggest paraprofessionals spend a significant portion of each school day providing behavioral support, delivering instruction, and progress monitoring students' academic and behavioral performance in the classroom. As such, they are uniquely positioned to observe and collect data regarding student interfering behavior. Yet, questions remain regarding the feasibility of utilizing SDO in terms of the amount of training and supervision paraprofessionals require in order to optimize their ability to collect behavioral data. Therefore, in order to capitalize on their potential to serve as observers, data collection methods are needed that are no less as accurate as SDO, but offer a more efficient and intuitive, and thereby acceptable, technique for use by paraprofessionals.

Direct Behavior Ratings

In light of the challenges to performing Systematic Direct Observation in a resource and time-sensitive manner, research has focused on developing more innovative methods of data collection. Today's multi-tiered system of service delivery has driven behavior assessment research to find behavior assessment methods that use classroom participants (such as teachers or paraprofessionals) as observers (Chafouleas, Kilgus, Jaffery, Riley-Tillman, Welsh, & Christ, 2013). Seeking to mirror the type of general outcome measure found in academic skill problem-solving models (i.e., curriculum-based measures), initial research into Direct Behavior Ratings (DBR) sought to establish the tool as not only a defendable, but also a feasible, option within a larger behavioral problem solving framework (Briesch, Chafouleas, & Riley-Tillman, 2016;

Chafouleas, 2011). The aim of DBR was to fill the gap in behavioral assessment, which typically focused on diagnosing problems and not on measuring response to intervention. Secondly, DBR sought to provide a tool that was efficient to implement, allowing for repeated administrations (Briesch, Chafouleas, & Riley-Tillman, 2016; Chafouleas, 2011).

Although noting their lack of applicability within a problem-solving model, Chafouleas (2011) identified systematic direct observation and behavior rating scales as the traditional cornerstones of behavioral assessment. Specifically, Chafouleas pointed to SDO's diagnostic sensitivity while acknowledging the ability of behavior rating scales to serve as screening measures, further noting neither are calibrated to detect short-term behavior change (2011). In contrast, as a behavior assessment methodology, Direct Behavior Rating involves an observer subjectively determining the perceived frequency of a given behavior within a relatively close proximity of having directly observed the behavior (Chafouleas, 2011). In this sense, DBR evolved as an attempt to progress monitor a student's response to behavioral interventions. As such, proponents of DBR identify it as particularly useful in problem-solving models.

Researchers identify DBR as similar to SDO in that behaviors are operationally defined, ratings are made by those who observed the behavior, and the behavior is observed for a specified time period (Kilgus et al., 2014). Moreover, DBR provides the time, resource, and cost efficiency of behavior rating scales (Chafouleas, 2011; Kilgus et al., 2014). As an example of DBR, a teacher might place an 'X' on a line, anchored to reflect zero instances of behavior on one end and extreme rates of behavior on the other, to indicate his or her perception of the frequency of a student's level of disruptive behavior during math class. In this respect, DBR is much like a pain scale index used in hospital settings to subjectively measure a patient's pain level at any given time. Initial research into DBR focused on making key determinations surrounding scale

construction, target behavior identification, and procedural guidelines (Chafouleas, 2011). Noting the influence of the existing literature and citing the goal of capturing data on both problem and pro-social behaviors, several studies led to the adoption of a 10-gradient line used to capture user's perceptions regarding the percentage of time a student displayed academic engagement, respectful, and disruptive behavior within a given time period. These initial studies noted improved observer accuracy for globally defined, positively worded behaviors (i.e., academic engaged vs academically unengaged) (Chafouleas, 2011). As noted by Chafouleas (2011), DBR was initially designed to be used at the Tier 1 (universal assessment of all students) and Tier 2 (targeted assessment of some students). Further, accuracy of DBR relied more on the proximity of the recording to the observation interval (i.e., the closer the proximity to the termination of the interval the better the reliability) and less on the length of the interval (i.e., 5 seconds, 5 minutes).

DBR compared with SDO

Speculating that Systematic Direct Observation may be better matched than Direct Behavior Rating for certain assessment purposes, Briesch et al., (2010) sought to investigate the influence of the psychometric adequacy of Systematic Direct Observation as compared to Direct Behavior Rating on decisions related to observation selection. In the study, SDO and DBR data were gathered regarding the academic engagement of twelve Kindergarten students during a 45-minute time block. Participants included two Kindergarten teachers, both based in the same classroom. Over the course of ten days, video cameras were used to capture the behavior of the students while the teachers recorded DBR data at the conclusion of three 15-minute blocks (Briesch et al., 2010). Researchers used the video in order to gather the SDO data based on a momentary time-sampling procedure. Along with other research regarding SDO, inter-rater

agreement for SDO was strong as was the method's sensitivity to changes in student behavior, indicating it offered a dependable measure of behavior. In terms of the data collected by teachers via the DBR method, their ratings were similar for a single data point when compared to SDO. However, as additional data points were considered, only moderate reliability coefficients were noted (Briesch et al., 2010). Furthermore, based on the data comparing SDO and DBR, observations made via SDO suggested it was appropriate for purposes of progress monitoring and screening. In contrast, the data gathered via DBR supported its use for ranking students relative to one another (Briesch et al., 2010). However, as the study included only two teachers gathering data in only one classroom, the results were limited in their generalizability to other populations. Further, the study only observed for academic engagement and therefore, did not include other behaviors, which may have led to different results.

Accuracy of DBR

Studies have attempted to establish an empirical base for the accurate use of Direct Behavior Rating – Single Item Scale for assessment purposes (Briesch, Chafouleas, Riley-Tillman, 2016). These studies have attempted to investigate the impact of target behavior selection, the reliability across raters and occasions for different behaviors, the level of training that leads to the most accurate results, and concurrent validity of ratings of different behaviors compared to systematic direct observation (Briesch, Chafouleas, & Riley-Tillman, 2016).

Studies have investigated the effects of various training packages on the accuracy of DBR. Harrison et al. (2014) compared the inter-rater accuracy of 67 participants across three behaviors (academic engagement, disruptive behavior, compliance) who received one of three training packages. Those in training package one watched a 5-minute presentation that gave an overview of DBR procedures. Training package two combined the presentation with an

opportunity to practice completing a DBR on four students' behaviors (shown via video clip) and receive feedback about their accuracy. The third session involved the same presentation as package one and two, however, the participants then practiced with eight video clips and received feedback. Results indicated participants from packages two and three rated with similar accuracy (Harrison et al., 2014). Ratings across all packages were similarly accurate as compared to an expert rating. However, ratings of disruptive behavior were significantly more accurate when participants received training that included practice with feedback. Further, when rating academic engagement and compliance, participants who received training package one produced as accurate or better ratings than those who received the more extensive training packages. Further, results indicated that accuracy was best for behaviors that occurred at either a low- or high- rate as compared to those that occurred at a more medium rate (Harrison et al., 2014).

In a study designed to examine the effects of different types of training on the accuracy of teacher-completed DBR, LeBel, Kilgus, Briesch, and Chafouleas (2010) sought to understand the level of training that would produce the most accurate ratings of behavior by teachers. Researchers divided 53 secondary teachers into one of three training conditions: direct training, indirect training, or no training. Teachers viewed 2-minute clips of either actual or simulated classroom footage of student behavior and rated the student's academic engagement and disruptive behavior. Expert ratings of the video clips, used as the criterion, were coded by research assistants. Results indicated that regardless of training condition, teachers were able to approximate the expert ratings when using the DBR form, suggesting minimal levels of training are needed to produce accurate results (LeBel, Kilgus, Briesch, & Chafouleas, 2010).

A study conducted by Chafouleas, Jaffery, Riley-Tillman, Christ, and Sen (2013) sought to investigate the impact of positive vs. negative wording of target behaviors and level of behavior (low, medium, high) for academic engaged behavior, disruptive behavior, and respectful behavior. Systematic direct observation was used as the criterion measure. In the study, 113 undergraduate students viewed video clips of students receiving instruction in a classroom and rated the target behaviors using DBR-SIS. Among the findings, respondent's perceptions of the estimated time the target student engaged in the target behavior were impacted by the wording of the definition and the duration of the behavior within the clip. Accuracy was improved for positively worded definition of academic engaged behavior. In contrast, the negatively worded definition for disruptive behavior led to improved accuracy of respondent ratings (Chafouleas, Jaffery, Riley-Tillman, Christ, & Sen, 2013).

Researchers have also investigated the relationship between rater accuracy and the duration of the observation interval. A study conducted by Riley-Tillman, Christ, Chafouleas, Boice-Mallach, and Briesch (2011) varied the duration of video clips depicting an actual third grade student participating in class (5 minutes, 10 minutes, and 20 minutes). Using a standard DBR form anchored with the words never, sometimes, and always, 81 undergraduate students viewed the clips and rated the student's academically engaged and disruptive behavior. Ratings for half of the clips reflected participant perceptions of the estimated percentage of time the student engaged in a given target behavior. Participants rated the remaining clips according to their perception of the actual time the student engaged in the target behavior. One week later, participants rated the clips again to test the reliability of the ratings after time had passed. Expert ratings of the behavior displayed in the video clips were obtained using a coding procedure performed by two graduate students (inter-rater agreement was 92% for the academically

engaged behavior and 88% for disruptive behavior). Results indicated that DBR tended to over-estimate the occurrence of the both behaviors as compared to SDO data by 1 to 2 points. The type of anchoring system used (estimated time vs. actual time) did not result in a significant difference in the accuracy of the ratings. Results also indicated large effects in regard to the influence the duration of the observation session for disruptive behavior. As session duration increased, the over-estimation of the disruptive behavior obtained via DBR also increased. Data regarding the test-retest reliability gathered in the study unsurprisingly suggested that as the number of raters and the number of ratings increased, the consistency of the ratings also improved (Riley-Tillman, Christ, Chafouleas, Boice-Mallach, & Briesch 2011).

Chafouleas, Briesch, Riley-Tillman, Christ, Black, and Kilgus (2010) investigated the generalizability of previously reported reliability across raters and occasions for academically engaged and disruptive behavior. Four adult raters (2 teachers and 2 research assistants) observed the behavior of seven middle school students within a single classroom during language arts class. A teacher and research assistant simultaneously conducted three 10-minute observations per day over the course of six consecutive school days. Results indicated for both behaviors (academically engaged and disruptive behavior), 53% of the total variance was associated with variance in target students and time of day. The researchers described the findings as significant in that, much like systematic direct observation, DBR ratings were sensitive to behavioral differences exhibited across target students and within repeated observations. However, the researchers noted that 32% of the data was missing due to teacher and student absences. As a result, the decision was made to use multiple imputation to complete the data set and allow for a balanced analysis of the data. Consequently, this limits the generalizability of the findings (Chafouleas et all, 2010).

Other studies have attempted to establish the concurrent validity of DBR in assessment by comparing it to the gold standard of behavior assessment, namely SDO (Riley-Tillman, Chafouleas, Sassu, Chanese, & Glazer, 2008). One study involved 15 elementary and middle school teachers who observed the on-task and disruptive behavior of students over three, separate 15-minute observation sessions. A graduate student serving as the expert observer in the classroom simultaneously collected SDO data. All teachers were trained using similar procedures by the researchers. Results indicated that moderate correlations were found between teacher DBR data and SDO data collected by the external observer. Further, for disruptive behavior, 78% of the variance in the SDO data could be related to a portion of the variance in the DBR rating. The researchers caution that the results simply suggested that DBR procedures might be used in a similar fashion, and not in place of, SDO procedures. Limitations of DBR noted by the researchers included the subjectiveness of DBR in that ratings offered an estimate of the respondent's perception of the target behavior (as compared to a more objective measure afforded by SDO). Further, this study included a small sample size and ratings may have been impacted by rater bias (Riley-Tillman et al., 2008).

Acceptability of DBR methods

Research has found DBR to be acceptable to teachers as a method for gathering data regarding student behavior. Riley-Tillman, Chafouleas, Briesch, and Eckert (2008) analyzed the results of a survey completed by 191 members of the National Association of School Psychologists. Participants were asked to indicate their level of training, use, and acceptability of Daily Behavior Report Cards (similar to DBR) and Systematic Direct Observation procedures. The results indicated the average acceptability of SDO to be in the high range (M = 4.9 on a 6-point Likert scale). For DBRC, the average acceptability rating was similarly high at 4.54.

Collectively, the authors point out that these results suggest school psychologist find DBR-like methodology as acceptable as SDO (Riley-Tillman, Chafouleas, Briesch, & Eckert, 2008).

In a more recent study focused on DBR, Miller and colleagues (2014) measured teacher perceptions of the acceptability of DBR as a behavioral assessment relative to two other commonly used indirect assessments (Behavioral and Emotional Screening System - BESS; Social Skills Improvement System-Performance Screening Guide - SISS). Participants in the study included 65 teachers across 16 different elementary and middle schools. Researchers asked teachers to complete each behavioral assessment at three points (fall, winter, spring) over the course of a school year. The Usage Rating Profile - Assessment (URP–A; see Chafouleas, Miller, Briesch, Neugebaur, & Riley-Tillman, 2012) was used to measure the perceived acceptability of each measure by the teachers in the study. The Usage Rating Profile-Assessment is a 28-item self-report measure designed to capture user perceptions regarding the acceptability of an assessment method (Miller et al., 2014). Based on the user's ratings, which are made by selecting the most appropriate response from a six-point Likert scale, results from the URP-A indicate overall acceptability of the measure, as well as other factors including the amount of knowledge needed to use it, the amount of administrative support needed to implement it, and the ease with which it can be used repeatedly (Mille et al., 2014).

Based upon teacher ratings on the URP – A, no significant differences were found between the overall acceptability of each assessment. However, with regard to the teacher's perception of the amount of knowledge required to use each item, DBR was rated more favorably than either the BESS or SISS. Furthermore, DBR was reported as more favorable than either assessment in terms of the amount of support that would be needed to use the assessment (Miller et al., 2014).

Application of DBR in schools

Studies have investigated the application of DBR methodology for use as a behavioral problem screener within schools (Kilgus et al, 2014). In a study designed to investigate the use of DBR as a screener in Elementary and Middle schools, researchers obtained ratings of 831 students in Kindergarten through 8th grade (Chafouleas, Jaffery, Riley-Tillman, Christ, & Sen, 2013). Teachers completed DBR, Single Item Scales (DBR-SIS) along with the Behavioral and Emotional Screening System (BESS) and the Student Risk Screening Scale (SRSS). The study was conducted in two phases with Phase I focused on K - 5th Grades and Phase 2 focused on 6th through 8th grades. DBR-SIS ratings were recorded by the classroom teacher twice per day regarding three target behaviors for a period of 5 days (30 total data points). After collection of the DBR data, each teacher completed a BESS and a SRSS on each student.

Overall, findings from the study indicated moderate to strong correlations between measures of students' degree of risk for behavioral problems on the DBR-SIS relative to BESS and SRSS. As noted by Chafouleas and colleagues (2013), closer examination of the results suggested most cut scores for DBR-SIS were either higher in sensitivity or higher in specificity. Cut scores varied across grade (i.e., lower vs. higher grade levels) and across single target behaviors versus combined target behaviors (i.e., respect vs a combination of the three target behaviors). Findings indicated that DBR-SIS tended to be more accurate in the elementary grades than in the middle school grades (Chafouleas et al., 2013). Moreover, findings indicated that disruptive behavior was the best individual predictor of student risk for behavior problems in the elementary grades. In the middle school grades, academic engagement was a better predictor than disruptive or respectful behavior (Chafouleas et al., 2013).

DBR has also been used in an applied sense to monitor student response to positive behavior intervention and supports. Interested in addressing mental health concerns within schools, von der Embse, Scott, and Kilgus (2015) found evidence of DBR's applicability for measuring students' response to interventions aimed at reducing anxiety. In the study, 115 participants spent 60 minutes taking a modified GRE General Test in a simulation of a high-stake testing environment. Pre- and post-test anxiety was measured via the Test Anxiety Inventory (TAI). Participants completed a DBR self report at 10-minute intervals throughout the testing session. The DBR broke ratings into three scales based on a bio-psychosocial model of anxiety: social (DBR-S), cognitive (DBR-C), and physiological (DBR-P). Results from the study indicated moderate effect sizes for concurrent validity between DBR-C, DBR-S, and DBR-P anxiety and the TAI (von der Embse et al., 2015).

Questions regarding DBR defensibility

With an aim on communicating information about both assessment and intervention, Direct Behavior Ratings have been purported to allow a flexible, defensible, efficient, and repeatable method for behavioral screening and intervention (Briesch, Chafouleas, & Riley-Tillman, 2016). Literature regarding the flexibility of DBR has demonstrated its adaptability to multiple setting, raters, and conditions. In addition, research supports the efficiency and repeatability of DBR to measure student behavior (Briesch, Chafouleas, & Riley-Tillman, 2016). Notwithstanding this evidence of the utility of DBR, the degree DBR is defensible as an accurate and valid measure of behavior is not without controversy. Research regarding DBR has attempted to build a psychometric base in support of its defensibility as a progress monitoring and screening tool, yet questions remain as to its accuracy relative to SDO.

The use of the term *accuracy* within behavioral assessment, such as the way much of the research presented to this point has used it to refer to the defensibility of DBR to accurately capture data regarding the behavior(s) of interest, is not without debate within the literature. As noted by Gresham (2003), traditional assessment is based upon test theory (e.g., classical test theory). In traditional assessment, reliability is a measure of the correlation between true scores (with an undeniably accurate value) and the obtained scores. Higher reliability correlations reflect stronger estimates of an assessment's accuracy. Furthermore, in traditional assessment, validity refers to how well a test assesses what it purports to measure. Validity is required in order to have confidence in the inferences drawn from the data (Gresham, 2013).

Within behavioral assessment, particularly with regard to behavior observation, the undeniable true value of the target behavior is not known. Therefore, establishing the defensibility of a behavioral assessment such as DBR or FBA is not entirely possible (Gresham, 2003). Within FBA, reliability is a measure of the degree to which observers agree on a behavioral function having viewed the same environmental events at the same time (Gresham, 2003). Reliability is a measure of inter-observer agreement (IOA), or the degree of agreement between two observers regarding occurrence vs. non-occurrence of the target behavior. Validity within functional behavior assessment is a measure of the correspondence between the observed and true function of behavior (Gresham, 2003). However, as with reliability, the true nature of the behavior of interest is not an incontrovertible index. Therefore, inter-observer agreement between the behavior assessment and the generally held best-standard of assessment, systematic direct observation, is used as the standard against which to compare a given procedure. Within the behavioral analytic literature, inter-observer agreement is commonly referred to as the

believability of the data (Shiver, Anderson, & Proctor, 2001). In this sense, use of the term believability may better serve to represent the defensibility of behavioral assessment.

Preliminary investigation of DBR to observe functional relationships

Research examining DBR's applicability within a functional behavior assessment framework has only recently been undertaken. Based upon the previous findings of DBR applicability as a mental health screening instrument and its use to measure response to intervention for student anxiety, DBR methodology was applied to an FBA framework by Kilgus and colleagues (2016). In the study, which aimed to establish DBR as a descriptive FBA tool for students at Tier 2 within a PBiS framework, 213 undergraduate students watched a 2.5-minute video clip and completed ratings of disruptive behavior and the type of consequence (e.g., adult attention, peer attention, escape/avoidance, access to tangibles/activities) that followed. Ratings of disruptive behavior corresponded to the percentage of time the student engaged in the behavior. Ratings of the consequence corresponded to the percentage of time each incident of disruptive behavior was followed by a given consequence. In order to gauge the utility of using DBR to gather data regarding functional relationships, participants were randomly split into four groups: (1) FBA training with performance feedback, (2) FBA with no performance feedback, (3) pre-test only, or (4) pre and post-test measures. Results indicated that the participants within the FBA training with feedback group produced DBR consequence ratings that were as comparably accurate as ratings generated by SDO. These findings suggest DBR methodology can be used for assessment of functional relationships. Despite these promising findings, the researchers noted that the study was limited in that study participants were not typical school personnel, that the application of DBR was outside its intended use, and that the observers themselves played no role in the functional contingency (Kilgus et al., 2016).

Despite the research suggesting DBR holds promise as an approach to assessing functional relationships, questions still remain. For instance, as DBR requires the teacher or other classroom personnel (i.e., paraprofessional) to observe and record their perception of the student's behavior, the ratings are prone to rater error (Briesh et al., 2010). Further, much as is the concern when they perform Systematic Direct Observation, the attention of teachers and classroom personnel is typically divided among several tasks at any given time, producing additional sources of potential error. As such, additional research is needed that investigates the believability of data generated using DBR to assess functional contingencies.

Summary and conclusions

Despite strong empirical support for the use of Functional Behavior Assessment to address the interfering behavior of students, schools often find collecting the necessary data difficult. Methods for detecting functional relationships that are practical and acceptable to school personnel are needed as traditional approaches are too costly and consume too many resources, as well as require expertise beyond the typical repertoires of school psychologists. The practicality and acceptability of Systematic Direct Observation, while forming the foundation of behavioral assessment, is limited given the training of school personnel to perform its procedures and the capacity of schools to facilitate its use. Research, such as that into Direct Behavior Ratings, has attempted to establish data collection procedures that simulate the precision of SDO methods while maintaining the acceptability of indirect methods of observation.

Traditional DBR was designed for screening purposes and/or progress monitoring a student's response to intervention. DBR can be used across all Tiers within a Response To Intervention framework. With regard to observing and recording perceptions of overall disruptive behavior, research regarding rater believability (aka inter-observer agreement)

suggests that respondents who receive training with feedback regarding their performance tend to rate with greater believability than respondents without training or feedback. Research indicates believability is better when multiple ratings are made by the same rater, as opposed to multiple ratings made by different raters. Findings also suggest that believability is better for high rates of disruptive behavior than either low or moderate rates of disruptive behavior. Furthermore, research indicates raters tend to find DBR acceptable as a data collection method.

Current research is only emerging with regard to applying DBR-methodology to the assessment and identification of functional relationships. As a result, several questions remain regarding the believability of DBR ratings as compared to systematic direct observation procedures. Further, there limited research regarding the acceptability of DBR when used to assess functional relationships.

Research questions

The major purpose of this study is to assess paraprofessionals' use of DBR methodology to observe functional contingencies in a classroom, and to determine their perceptions of DBR's acceptability when used in this manner. This examination of DBR application to Functional Behavior Assessment is important given the relatively recent emergence of DBR as a cost- and time-saving data collection method used by school personnel. Given the legal and ethical ramifications of decisions based on FBA, it is essential that data generated by paraprofessionals' use of DBR be compared to an already established gold standard of observation, namely SDO. In order to achieve this, the study will focus on two research questions. First, compared to data obtained by an expert observer using Systematic Direct Observation, can paraprofessionals (i.e., classroom aides) using Direct Behavior Ratings generate believable ratings of the estimated occurrence disruptive behavior? Furthermore, given the occurrence of disruptive behavior, how

similar are DBR ratings and SDO scores for the consequences perceived to maintain the behavior? It is hypothesized that DBR ratings will strongly correlate (greater than .80 Pearson's correlation coefficient) with SDO data across disruptive behavior as well as four consequence targets given the occurrence of the disruptive behavior. Consistent with previous research findings regarding the use of DBR for assessment of functional contingencies, raters will generate scores within 10% to 20% percentage-points of the SDO data.

In addition to the believability of the DBR ratings, the second research question addressed by this study will seek to determine the extent to which classroom aides perceive DBR, used in the context of assessing functional relationships, as an acceptable method of data collection. It is hypothesized that classroom aides will report feeling DBR is an acceptable method of collecting data regarding function relationships, as evident by average scores on the Usage Rating Profile – Assessment falling above 4.0 (on the 1.0-6.0 scale).

Chapter 3: Method

Participants and Setting

Participants included seven paraprofessionals (Education Support Technicians) from an Agency-based, Private School serving the residents of a Residential Treatment Facility (RTF) that provides care to children and adolescents with various mental health and behavioral health disorders. Located within a rural school district in the northeast United States, all residents of the RTF are enrolled in the education program at the school as part of their overall course of treatment. The residents represent a large geographic region of the home state. In addition, students from the local community attend classes at the school as determined by the Local Education Agency (LEA) as outlined in their Individual Education Plans (IEP). This particular setting was selected given the tendency for classrooms to contain students with elevated rates of challenging behaviors that interfere with their academic performance. Four of the paraprofessionals were Caucasian males and three were female. A fourth female paraprofessional was originally selected as the eighth participant; however, she was forced to drop out prior to data collection due to sustaining an injury the day before data collection began. A majority of the paraprofessionals obtained either a High School or equivalent diploma or had some college credit (86%), while one paraprofessional possessed a Bachelor's degree. Of the paraprofessionals, 14% had 6 months to 1 year of experience, 43% had 1 to 1.5 years of experience, while 43% of them had more than 4 years of experience. (See Table 1.) Each of the paraprofessionals was asked to serve as raters of student behavior as part of their assigned classroom duties within the program. Consent for participation was obtained from each paraprofessional prior to participation in the study. Refer to Appendix A for a copy of the participant (paraprofessional) consent form.

In addition, the students in the school served as potential participants whose behavior was observed and given ratings by the classroom paraprofessionals. In an effort to simulate the reasons an FBA may be performed, not all students from each class were eligible to serve as participants in the study. Through a review of student records and consultation with classroom teachers, only students who met the following inclusion criteria were identified as potential candidates for observation during the study: a) students on whom either an FBA was on file that targeted disruptive behavior (see definition for disruptive behavior below including examples and non-examples); b) students who had been referred to the office for disciplinary reasons related to disruptive behavior in the classroom during the previous school year; and/or, c) students for whom anecdotal reports identified as displaying disruptive behavior in the classroom two or more times each week. Only those students who met at least two of the three criteria for inclusion in the study were selected as potential participants for observation within a particular classroom.

Prior to contacting potential participants' caregivers/guardians, Institutional Review Board (IRB) approval from Alfred University's Human Subject Research Committee was obtained in June 2016. Prior to data collection, caregivers/guardians of 29 students were mailed a permission packet consisting of an introductory letter, an information sheet, and two copies of the opt-out permission form. The caregiver/guardian opt-out permission packet can be found in Appendix B. Caregivers/guardians were provided with an opportunity to return one copy of the opt-out permission form if they did not wish for their child to participate in the study. None of the caregiver/guardians chose to opt out of the study; therefore, none of the students were removed from the potential participant pool. A total of at least four students from 5 separate classrooms for whom the opt-out permission form was not returned, and who met the inclusion

criteria described above, were considered eligible to be observed during the study. Folders, containing study-related observation sheets, were placed on a table in each classroom on that classroom's day of observation. Blind to the contents of the folders, three paraprofessionals were instructed to randomly select three folders from the array, while three others selected two from the array. Lastly, one paraprofessional selected only one folder. Those students selected served as targets for observation during the study.

A total of sixteen students were selected for observation. In this study, 10 students were male and 6 students were female. The students ranged in grade level as follows: 3 were in Elementary (Grades 3 – 6); 6 were in Middle School (Grades 7 – 9); and 7 were in High School (Grades 10 – 12). A majority of the students were classified with an Emotional Disturbance (11 students), while the other students carried classifications of Intellectual Disability (3 students), Other Health Impaired, and Autism (1 student each). Demographic information related to student participants is summarized in Table 2.

Design

This study is an experimental design wherein paraprofessionals' perceptions regarding the estimated percentage of time a student exhibited disruptive behavior, as well as the perceived estimated percentage of disruptive behavior met by that consequence (i.e., the function of the behavior), were collected via Direct Behavior Ratings and compared to data collected by an external observer using Systematic Direct Observation in order to assess the level of agreement between DBR and SDO data. The design was also intended to investigate the paraprofessionals' perceptions regarding the acceptability of the DBR methodology.

Independent/Dependent Variables

Two dependent variables, disruptive behavior and its perceived consequence, were observed. Disruptive behavior was selected as it is a frequently observed behavior within the DBR literature base and was the target behavior observed in a previous study investigating the application of DBR methodology to determine functional relationships (Briesch, Chafouleas, & Riley-Tillman, 2016; Kilgus et al., 2016). The definition from the Kilgus and colleagues (2016) study was adopted for use in this study. Therefore, disruptive behavior was defined as any student actions that interfered with their typical school or classroom activities. Examples of disruptive behavior included: getting out of one's seat, making noise such as when rapidly tapping a pencil, fidgeting, touching others, playing with objects, taking objects that do not belong to them, acting aggressively toward others, and talking or yelling about things unrelated to classroom instruction. Non-examples of disruptive behavior included: responding to teacher directives, asking questions related to class materials or instruction, or raising one's hand and waiting to be called upon before responding.

When selecting and defining the consequence variables in this study, multiple sources of information were considered including review of Cooper and colleagues' book on Applied Behavioral Analysis (2007), as well as a recent study on use of DBR for determining functional relationships (Bijou et al., 1968; Carr, 1994; Kilgus et al., 2016; Michael, 2004). Consistent with the literature on FBA conducted in schools, consequences were identified to simulate a large topographic response class of commonly occurring maintaining variables typically observed in school settings. For instance, escape/avoidance was constructed to include a range of behaviors including, but not limited to, leaving the room, hiding under a desk, or tearing up materials.

To provide a rating of the perceived consequence, participants selected from the following potential consequences – adult attention, peer attention, escape/avoidance, and access

to tangibles/activities. Adult attention was defined as verbal or nonverbal reaction by an adult. Examples included: being issued a verbal reprimand (e.g., "You know better than that"), being given a verbal directive (e.g., "Go to the office"), being issued a consequence as part of the implementation of a class wide behavior management system (e.g., clip down a color), or being given a momentary glance from a teacher. Peer attention was defined as verbal or nonverbal reaction by a peer. Examples include: a verbal reaction (e.g., talking with the peer; a peer commenting about the behavior, the peer laughing about the behavior) or a non-verbal reaction (e.g., a peer rolling their eyes, a peer turning their head toward the student). Escape/avoidance was defined as removal or avoidance of a task, activity, or expectation. Examples include student self-directed actions (e.g., walking out of the room, leaving assigned area/desk) as well as staff directed actions (e.g., removal of materials, reduction of expectations). Access to tangibles was defined as acquisition of items or activities. Examples include: attempts, successful or otherwise, to gain something (e.g., obtaining toys, food, gaining close proximity of person) regardless of student's preference for the tangible.

The independent variable was the type of data collection procedure used (Direct Behavior Rating vs. Systematic Direct Observation). Systematic Direction Observation (SDO) was chosen as the alternate method (criterion measure) given its acceptance as the commonly held gold standard method of observation within the literature base (Kilgus, 2016).

A final research question addressed by this study involved paraprofessionals' perception of the acceptability of the DBR methodology as a tool for observing and recording functional contingencies. As a third dependent variable, each aide completed the Usage Rating Profile – Assessment (URP – A) with the mean Total Usability and each separate Cluster score used to understand each paraprofessional's perception of DBR's overall acceptability.

Measures

Direct Behavior Rating of Disruptive Behavior

Direct Behavior Ratings (DBR) are used to capture the user's perception of the estimated percentage of time the student exhibited a target behavior (Miller, et al., 2014). For purposes of this study, DBR data was captured using a form based on the Direct Behavior Rating-Single Item Scale (DBR-SIS) as reported in the Kilgus and colleagues (2016) study investigating the application of DBR methodology when determining functional relationships. The DBR consisted of five separate graphic horizontal lines, each divided into 10 equal sections by 11 vertical lines. Anchors designated one end of each line as 0% (never), the middle of each line as 50% (some of the time), and the opposite end of each line as 100% of the time (always). Thus, a lower rating was more desirable for disruptive behavior. Observations were recorded within a close proximity of the termination of each interval.

Participants were asked to indicate two types of ratings on the DBR form. First, participants were asked to rate the estimated percentage of time the target student exhibited disruptive behavior during the 5-minute observation window. Disruptive behavior was the only targeted behavior included on the form. Secondly, each participant was asked to rate the percentage of time they perceived the disruptive behavior was met with a particular consequence. Participants rated each of the four consequence targets using the same lines and anchors used when rating disruptive behavior. Along with brief instructions regarding DBR procedures, definitions of the target behavior as well as each consequence were included on the rating form. Ratings were recorded within a close proximity of the termination of the interval. A copy of the DBR form used in this study, as well as a sample completed form, can be found in Appendix C.

Systematic Direct Observation

Systematic Direct Observation procedures are used to capture precise measurement of behavior as the behavior occurs (Johnston & Pennypacker, 2009). Depending on the nature of the observation technique (i.e., continuous vs. discontinuous), data is observed for a specified period of time and recorded on a form. For this study, an SDO data collection form was created in Microsoft Excel to include space to rate disruptive behavior using a 15-second partial interval procedure in order to obtain an estimate of behavioral occurrence over a 5-minute period (20 total 15-second intervals per 5-minute observation window). The duration of 15 seconds per interval was selected based on previous research that suggested shorter intervals offer increased accuracy (Olympia et al., 2002). At the end of each 15-second interval, the expert observer entered either a zero to indicate the student did not, or a '1' to indicate the student did, exhibit disruptive behavior according to the same operational definition used for the corresponding DBR procedure.

The SDO form also included a space to record data throughout each 15-second interval regarding the consequences perceived to have followed any incident of disruptive behavior. This data was collected using a frequency count procedure. That is, a mark was made following each occurrence of disruptive behavior that best reflected the perceived consequence of the student's behavior. The definitions with regard to each consequence target were the same as described for the DBR procedure. The decision to use a combination of partial interval recording and frequency counts was used based on previous research examining the agreement between DBR and SDO for engagement and disruptive behavior conducted by Riley-Tillman and colleagues (2008). Refer to Appendix D for a copy of the SDO form. A sample of page one from a completed form is also included.

Usage Rating Profile – Assessment (URP-A)

The Usage Rating Profile – Assessment is a 28 item, self-report measure based on an instrument used to assess the perceptions regarding the usability of a particular intervention (Miller et al., 2014). Respondents selected from a six-point Likert-type scale indicating their level of agreement (1 = strongly agree; 6 = strongly disagree). Sample items include "I understand the procedures of this assessment," "I would be committed to carrying out this assessment," and "This assessment is too complex to carry out accurately." A six-factor model has been demonstrated in studies investigating the construct validity of the URP–A (α = .63 - .90). The six factors are as follows: Acceptability – respondent's interest and eagerness to use the measure; Understanding – respondent's knowledge of the measure and the procedure governing its use; Home-School Collaboration – the degree to which the measure requires collaboration with home; Feasibility – the degree to which the measure can intuitively be used; System Climate – the likelihood the measure can be used within the school setting; System Support – the degree to which respondents feel an external support is needed. A copy of the URP–A form can be found in Appendix E.

Procedures

Pre-Observation Training

Paraprofessionals. Prior to collecting data, an approximately 1-hour training session was held to train participants (paraprofessionals) in all study procedures. After the participants were given information regarding an overview of the study, an explanation of Functional Behavior Assessment and its purpose in schools was presented. Next, a three-step program was used to train participants on the Direct Behavior Rating and Usage Rating Profile – Assessment forms. The first step reviewed components of the DBR form, described the definitions of the target behavior, as well as each consequence target, and explained the instructions for completing the

Usage Rating Profile – Assessment form. The next step included a demonstration that illustrated how to rate disruptive behavior using the DBR form. As part of the demonstration, the participants watched a brief video clip and were instructed to focus on the behavior of a target student. The participants then listened as an explanation was offered regarding the selection of DBR ratings of that behavior. The third step included a practice session that required participants to view an additional video clip and then use the DBR form to rate the student's disruptive behavior. Once ratings were complete, the participants were asked to reveal their ratings before the researcher reviewed the actual percentage of time the student was disruptive as well as the percentage of disruptive behavior that was followed by a particular consequence. A brief discussion was held to explore any discrepancies in ratings between the paraprofessionals and the researcher. Participants then watched a third video clip and independently completed the DBR form. At that point, all participants reached at least 80% agreement with the researcher's DBR coding. Salvia, Ysseldyke, and Bolt (as cited in Burns and Parker, 2014) stipulate that inter-rater coefficients of .90 or higher be used for high-stakes decisions whereas coefficients of .80 are sufficient for screening decisions about individual students. The inter-rater reliability coefficient of .80 or above was deemed acceptable for this study given DBR is not meant to be the sole source of information regarding behavior within the process of completing an FBA on an individual student. Although a 90% agreement (.90 inter-rater reliability coefficient) was preferred, training concluded once all participants reached at least 80% agreement (.80 coefficient) with the researcher's ratings. Training participants using a practice-with-feedback model was selected based on previous DBR research findings suggesting this type of training is associated with improved rater accuracy (Kilgus et al., 2016). Following the practice session, the specific procedures (detailed previously) for selecting students whose behavior was going to

be the target during observation was explained. All participants were given the opportunity to ask questions throughout the training session. Following training, aides began data collection the next day.

Expert Training. A Board Certified Behavior Analyst (BCBA) with over three years of experience working with students with co-occurring intellectual disabilities and mental health disorders was used as the expert observer to capture SDO data during the study. Before beginning data collection, the expert observer was trained in one session. This session included an overview of the study as well as each of the following components: a) review of the operational definitions of the target behavior and identified consequences; b) explanation of the SDO data collection form; c) practice coding using the SDO form. Practice consisted of using the same video clips used during the paraprofessional training session that depicted students engaged in disruptive behavior in a school setting. Prior to the training session, the researcher coded each segment using a partial interval procedure for determining the percentage of intervals of disruptive behavior and an antecedent-behavior-consequence observation procedure for determining behavior-consequence relationships. During the training session, the expert observer used the SDO form (described in next section) to record his observations of the students' behavior. By the conclusion of the training session, inter-rater agreement between the researcher and the expert was 93%.

Data Collection

In June 2017, data was collected over the course of four consecutive days during a 1-week observation window. Paraprofessionals were responsible for collecting data using the DBR form for the identified target student. The expert observer simultaneously recorded data on the SDO form for the same identified target student.

At the beginning of each school day data was collected, paraprofessionals randomly selected folders from an array of at least four folders placed on a table in the classroom. Each folder contained a packet of six blank DBR forms used to rate the identified target student's behavior. The folder also contained a blank copy of the URP-A form. Each single-sheet of paper DBR form contained the name of the target student for whom the rating was to be made, the scale to rate the target student's disruptive behavior, and the four other, separate scales to rate the percentage of the disruptive behavior followed by each of the identified consequence targets. A brief definition of the target behavior and each of the consequence targets was included on the form. Directions instructed paraprofessionals to place an 'X' on the line that best reflected the percentage of total time that the target student exhibited the disruptive behavior during the 5-minute observation interval. Directions for the other scales instructed the paraprofessional to place an 'X' on the line that best reflected the percentage of total disruptive behavior that the paraprofessionals perceived was followed by the line's corresponding consequence. The paraprofessionals were instructed to only mark on gradients and not between them.

Following completion of their final 30-minute observation window for the day, each participant completed one copy of the URP–A. Directions for the URP–A form asked the paraprofessionals to circle the number that best reflected their agreement or disagreement with each corresponding statement on the form.

The expert observer was given separate folders for each classroom that contained a blank SDO form. Once the paraprofessional selected the target student by choosing a folder, the expert observer consulted with the paraprofessional to learn the identity and seat location of the target student for whom the data was to be recorded. A Motiv-Aider was provided to the expert observer that was pre-set to a fixed 15-second interval with its vibration strength set to "4."

This device was used to cue the expert observer to record data on the SDO form. The expert was also responsible for cueing the paraprofessional at the termination of every twenty 15-second intervals (5-minutes). In order to provide this cue to the paraprofessional, the expert observer raised and waved his right hand above his head.

In order to control for order effects, a coin flip determined the order in which each of the students were observed by the aide. Each student was observed for a total of 30 minutes. The 30-minute period was divided into 6 consecutive, 5-minute intervals at the end of which the paraprofessional marked the DBR form. Consistent with DBR recording procedures, the paraprofessionals were instructed to mark the form within a close termination of the interval (Briesch, Chafouleas, & Riley-Tillman, 2016). They were also instructed to skip an interval should the next 5-minute interval lapse prior to making their rating on the DBR form. A short break was used in-between observation windows of each student to allow for reorganization of paperwork and re-starting of the Moti-vator.

Each paraprofessional placed all DBR forms into the student's folder at the conclusion of the observation window. They then completed the URP–A form before placing it into the same folder that contained the completed DBR forms. The folder was then collected by the expert observer. The expert observer then combined the DBR folder with all completed SDO forms and returned them to the researcher at the end of the school day.

Inter-observer agreement was assessed for expert ratings of disruptive behavior and its related consequences as recorded on the SDO form using a second independent observer. The independent observer was a Master Level-trained Behavior Therapist with more than 5 years of experience completing Functional Behavior Assessments of the interfering behavior of students with co-occurring mental health and intellectual disabilities. The independent observer was

trained according to the same procedure described above for the expert observer. The independent observer collected data using the SDO form during 480 of the total 1920 15-second intervals (a total of 25% of intervals observed). The second observer collected data on each of the days student behavior was observed. Agreement was determined by computing kappa values in order to correct for occurrence agreement due to chance (Hinze, 2005). The overall observed kappa values were as follows: Disruptive behavior (.875); Adult attention (.819); and Peer attention (.804) – each indicating substantial agreement (Hinze, 2005). Due to the minimal number of intervals during which disruptive behavior was perceived as maintained by escape/avoidance or access to tangibles, Kappa values could not be calculated.

As an incentive, all participating paraprofessionals and expert and independent observers who completed the study were provided with a $30.00 gift card to a nearby national retail store at the end of the data collection period. All data was entered into SPSS for analysis.

Data Analysis

The major objectives of this study are to describe paraprofessionals' ability to generate believable (similar) data using DBR to assess functional relationships relative to SDO and to examine their perceptions regarding DBR's acceptability. The major variables of concern are 1) believability, or level of inter-rater agreement between DBR and SDO data and 2) DBR's perceived acceptability. Prior to analyses, it was necessary to calculate scores based upon the data gathered by the expert observer in this study (SDO) to serve as the dependent (criterion) variable. These scores were regarded as the estimated true occurrence of disruptive behavior and each consequence target (Kilgus et al., 2016). Specifically, the occurrence of overall disruptive behavior was recorded using a 15-second partial interval recording system that ran continuously during each five-minute observation window. In total, the expert observer completed six 5-

minute observations of each student. A percentage of intervals during which disruptive behavior occurred was calculated for each observation window by dividing the number of 15-second intervals disruptive behavior occurred by the total number of intervals during each of the 5-minute observation windows [$SDO_{DISR} = X_{INTERVALS\ DISRUPTIVE\ BEHAVIOR\ OCCURRED} / 20$]. Consistent with the development of true score estimates by Kilgus and colleagues (2016), this score then served as a true score estimate of the percentage of time the student engaged in disruptive behavior within each 5-minute observation window. Data regarding the consequences that followed the occurrence of disruptive behavior was collected via a frequency count. True score estimates for each potential consequence were developed by calculating the percentage of total disruptive behaviors met with each consequence [$SDO_{CONSEQUENCE} = X_{NUMBER\ OF\ TIMES\ CONSEQUENCE\ OCCURRED} / X_{TOTAL\ NUMBER\ OF\ DISRUPTIVE\ BEHAVIORS}$]. DBR scores represented the average estimated percentage of time during each 5-minute observation window the paraprofessional observers (DBR) observed and recorded disruptive behavior, as well as the overall average estimated percentage of student disruptive behavior met with each consequence.

Based upon previous research findings, the researcher in this study hypothesized that the analysis of the data would reveal high levels of concordance between DBR and SDO data and that the paraprofessionals would find DBR an acceptable data collection method. To test these hypotheses, a number of different analyses were performed. These included Pearson's correlations, examination of occurrence agreement percentages and point differences in agreement between the DBR and SDO data, as well as visual analysis of graphed data.

A combination of analyses were performed in order to answer the first question, "Compared to data obtained by an expert observer using SDO, can paraprofessionals using DBR generate believable ratings of disruptive behavior?" Additionally, the analyses were performed to

answer the related question of, “Given the occurrence of disruptive behavior, how similar are the DBR and SDO ratings of the percentage of disruptive behavior met with each consequence?” First, Pearson correlations were performed to understand the relationship between the DBR and SDO data. Given the occurrence of disruptive behavior, correlations were also used to understand the relationship between the DBR and SDO data for each consequence target. Moreover, to further understand the level of agreement between the paraprofessional and expert observers, occurrence agreement percentages were calculated. Additionally, point differences in agreement between DBR and SDO data were examined on an interval-by-interval basis. Finally, to simulate FBA best-practice, observation results were graphed to allow for visual analysis of the DBR and SDO data.

To answer the second research question, “To what extent do paraprofessionals perceive DBR, used in the context of assessing functional relationships, as an acceptable method of data collection?,” URP – A cluster scores were used to determine if the mean response was different from a null hypothesis of “no acceptability.”

Chapter 4: Results

Multiple procedures were used to investigate the similarity of the data generated by the paraprofessionals' and expert observer's observations. Table 3 presents a summary of the estimated occurrence of disruptive behavior as recorded on the Systematic Direct Observation (SDO) and Direct Behavior Rating (DBR) forms, including the total number of 5-minute observation windows each observer recorded the occurrence of a disruptive behavior. This table also includes the number of observation windows met by each consequence, given the occurrence of disruptive behavior, across both observers.

Table 4 presents a summary of the overall average percentage of 15-second intervals the expert observer (SDO) recorded disruptive behavior during each 5-minute observation window, as well as the overall average percentage of student disruptive behavior met with each consequence. This table also summarizes the average estimated percentage of each 5-minute observation window the paraprofessional observers (DBR) recorded disruptive behavior, as well as the overall average estimated percentage of student disruptive behavior met with each consequence.

As a reference, the data generated by the expert observer in this study was considered the standard by which the DBR data was compared given the general acceptance of SDO as the "gold standard" of behavioral observation. Therefore, in order to establish a baseline of the estimated true occurrence of student disruptive behavior and its consequences, the SDO data was considered first.

SDO Data

Based upon the expert's observations, disruptive behavior occurred during 47 of the 96 (48.9%) five-minute observation windows. The data also indicated that based upon the expert's

observations, adult attention was an observed consequence during 38 of those 47 intervals (80.8%). Peer attention was a consequence during 23 of the 47 intervals (48.9%). Escape/avoidance was not observed to be the consequence for any interval. Access to tangibles was the consequence during 2 of the 47 intervals (4.26%). See Table 3.

The average percentage of 15-second intervals the student exhibited disruptive behavior for each observation window was 27.8%. In terms of consequence targets, the average percentage of 15-seconds intervals the student disruptive behavior was met with adult attention during each observation window was 78.6%. The percentage of intervals behavior met with peer attention was 41.3%, escape/avoidance was 0.0%, and access to tangibles was 75%. See Table 4.

DBR data

While SDO data was gathered by the expert observer, the paraprofessional simultaneously observed the same student and recorded data via the Direct Behavior Ratings (DBR) form. Based upon the paraprofessionals' observations, disruptive behavior was observed during 36 of the 96 (37.5%) five-minute observation windows. Based upon the paraprofessionals' observations, adult attention was an observed consequence during 32 of those 36 intervals (88.8%). Peer attention was a consequence during 13 of the intervals (36.1%). Escape/avoidance was an observed consequence during 7 of the intervals (19.4%). Gaining access to tangibles was an observed consequence during 4 of the intervals (11.11%). See Table 3.

The DBR data also indicated that the average estimated percentage of observation window the student exhibited disruptive behavior was 34.4%. The average estimated percentage of disruptive behavior met with adult attention during each observation interval was

66.6%. The percentage of disruptive behavior met with peer attention was 23.5%, escape/avoidance was 42.1%, and access to tangible was 45%. See Table 4.

Correlations among observation methods

Pearson's test of correlation was conducted to assess the strength and nature of the relationship between DBR ratings and SDO scores for disruptive behavior and all consequence targets. Correlations can be found in Table 5. There was a significant, positive correlation between observations of overall disruptive behavior made using Direct Behavior Rating and those made using Standard Direct Observation ($r = .886$, $N = 96$, $p < .001$, two-tailed). It is a strong correlation: calculation of r^2 indicated 73% of the variation is explained. With regard to consequence targets, correlations were calculated using the DBR ratings and SDO scores only for the intervals in which the expert observer indicated disruptive behavior occurred ($n = 47$). It was revealed that given the occurrence of disruptive behavior, there was a moderately strong, significant positive correlation for adult attention ($r = .470$, $N = 96$, $p = .001$, two-tailed). Calculation of r^2 for the adult attention target indicated 22% of the variance in DBR data can be interpreted as a portion of the variance in SDO data. The results revealed DBR data was not significantly correlated with SDO data for the peer attention, escape/avoidance, or the access to tangible consequence targets.

Examination of occurrence agreement

House, House, and Campbell (1981) discuss several measures of association for examining inter-observer agreement, noting the advantages and disadvantages of each. Among those measures, the authors discussed the use of occurrence agreement percentage formulas as a means for determining the level of agreement between observers regarding the occurrence (behavior occurred) and non-occurrence (behavior did not occur) between two sets of data during

a set interval of time. Hinze (2005) stated that occurrence agreement formulas should be used whenever target behaviors occur at very low rates (i.e., less than 20%). Occurrence agreement percentage formulas are considered to be a conservative measure of association between two data sets as they account for the possibility of both agreement and disagreement. Typically, occurrence agreements are calculated by dividing the total number of agreements by the number of agreements plus disagreements and multiplying by 100%. However, House and colleagues caution that these formulas are subject to bias due to the potential that the number of intervals the observers agree differs from the number they disagree (1981). This difference affects the probability of chance agreement. Therefore, to account for this in the current study, the occurrence agreement percentage (level of agreement regarding the number of observation windows the behavior did occur and did not occur) between DBR scores and SDO ratings was calculated using weighted occurrence agreement percentage (House et al., 1981). The rationale for using this measure was that the expert observer was assigned "criterion" status given the general acceptance of SDO as the gold standard of direct observation. Therefore, only the intervals in which the paraprofessional failed to record a behavior were counted as errors (i.e., expert recorded an occurrence but the paraprofessional did not). In order to calculate the weighted occurrence agreement percentage, the total number of agreements was divided by the number of agreements plus disagreements (intervals in which the paraprofessional failed to record a behavior) and multiplying by 100%. When calculating occurrence agreement percentage between two data sets, results can range from 0% (no association) and 100% (total association) (House et al., 1981). Generally speaking, "average agreement percentages at or above 70% are necessary [to suggest inter-observer reliability], above 80% is adequate, and above 90% is good" (House et al., 1981, pg. 46).

Overall disruptive behavior

Of the ninety-six 5-minute intervals, the paraprofessional agreed with the expert observer that disruptive behavior occurred during 29 intervals. Both observers agreed a behavior did not occur during 42 intervals. However, during 25 additional intervals, the observers disagreed that disruptive behavior occurred (DBR recorded behavior during 7 intervals; SDO during 18 intervals). Refer to Table 6 for a two by two matrix notation of inter-observer agreement for overall disruptive behavior. With regard to the level of agreement between DBR and SDO for overall disruptive behavior, the weighted occurrence agreement was 62%, or below the minimal average agreement percentage of 70%. Occurrence agreement percentages, both weighted and non-weighted, for overall disruptive behavior, and each consequence target, can be found in Table 7.

Consequence targets

The weighted occurrence agreement percentage formula, as discussed by House and colleagues' (1981), was used to consider the level of agreement between SDO and DBR data with regard to consequence targets for the 47 intervals the expert observer indicated disruptive behavior occurred. Analysis of this data revealed multiple findings. First, with regard to disruptive behavior maintained by adult attention, the weighted occurrence agreement was 82%, or adequate. Secondly, with regard to peer attention as the perceived maintaining consequence, the weighted occurrence agreement was 42%, or below the minimal necessary to be considered reliable. Next, with regard to disruptive behavior maintained by escape or avoidance, the weighted occurrence agreement was 0%, or below the minimum threshold required for reliability. Similarly, for disruptive behavior perceived as maintained by access to tangibles, the weighted occurrence agreement percentage was 0%.

Examination of percentage point differences

In order to gain an alternate understanding of the similarity between DBR and SDO data, percentage-point differences in agreement between DBR ratings and SDO scores were examined (Kilgus et al., 2016; Riley-Tillman et al., 2008). In previous research investigating the similarity between DBR ratings and SDO scores, researchers chose to convert rating estimates from a 0% - 100% percentage scale to a 0 – 10-point scale. However, rating estimates from the current study were not converted in an attempt to minimize any potential confusion when comparing DBR ratings to SDO scores. Nonetheless, a review of previous research has suggested that when recording observations with accuracy, the DBR ratings would be expected to fall within a range of 0 to 2 points of the SDO scores (Kilgus et al., 2016). Given the scale of the current study, DBR ratings were expected to fall within a range of 0% to 20% of the SDO scores.

Overall Disruptive Behavior

With regard to overall disruptive behavior, the paraprofessional and expert observer agreed that no disruptive behavior occurred during 42 of the 96, 5-minute observation windows. However, examination of the remaining 54 observation windows revealed some interesting findings. First, the data was considered in terms of the 29 observation windows the expert observer and paraprofessional agreed on the occurrence of disruptive behavior. Of these instances, six of the DBR ratings of disruptive behavior exactly matched the SDO ratings of disruptive behavior. The data also revealed that 14 of the observation windows had a 10% difference in agreement between ratings, whereas 5 observation windows had a 20% difference in agreement between DBR and SDO ratings.

The second interesting finding was that when the observers disagreed a disruptive behavior occurred (25 observation windows), the data revealed 20 observation windows had a

10% difference in agreement with an additional 2 observation windows having a 20% difference in agreement. Collectively, the DBR ratings from 84% of the intervals either matched exactly or were within 10% or 20% difference in agreement of the SDO ratings, whenever the paraprofessional and/or expert observer recorded the occurrence of disruptive behavior. Information regarding the number of observation windows with percentage-point differences in agreement between DBR and SDO data for disruptive behavior and can be found in Table 8.

Consequence Targets

In order to investigate the similarity of the data for each consequence target, examination of percentage-point differences in agreement for each consequence target was restricted to the 47 observation windows during which the expert observer recorded the occurrence of disruptive behavior. The data suggested adult attention was functionally related to disruptive behavior for 39 of the intervals. Of these intervals, DBR ratings matched the SDO ratings for six of the intervals. For 7 of the observation windows, the difference in agreement was within 10% and within 20% for an additional 4 observation windows. However, for 22 of the observation windows in which adult attention was perceived as the consequence for the disruptive behavior (56.4%), the difference in agreement was equal to or greater than 30%. Information regarding point differences in agreement for each consequence target, given the occurrence of disruptive behavior, can be found in Table 9.

The data indicated that peer attention was the potential consequence for disruptive behavior in 25 of the 47 observation windows. DBR ratings did not match SDO ratings for any of the observation windows. For 3 of the observation windows, differences in agreement between the DBR and SDO ratings were within 10% and for 8 observation windows the differences in agreement were within 20%. For 14 observation windows, differences in

agreement between DBR and SDO data were 30% or greater (56%). The data also revealed escape/avoidance served as a potential consequence for 7 observation windows during which disruptive behavior was recorded as having occurred. For 3 of the observation windows, the point difference in agreement was within 10% to 20%. For four windows, the difference in agreement was equal to or greater than 30% (57%). Similarly, access to tangibles was the perceived consequence during six of the observation windows. The difference in agreement between DBR and SDO data was within 10% to 20% for 2 of the windows. For 4 of the observation windows, the difference in agreement was 30% or greater (66.7%).

Comparison of Identified Function

Within an Applied Behavioral Analysis framework, procedures for reporting the results of a Functional Behavior Assessment typically include visual inspection of the data, particularly as it pertains to identifying the function of the interfering behavior. In order to simulate this process with the data obtained from the current study, the percentage of 5-minute observation intervals a given consequence was recorded as the strongest maintaining variable within the functional contingency, for both DBR and SDO, are illustrated in Figure 1. Based on this graph, both DBR and SDO identified adult attention as the strongest consequence likely maintaining the disruptive behavior across all observations.

Acceptability of Direct Behavior Rating

To answer the second research question, "To what extent do paraprofessionals perceive DBR, used in the context of assessing functional relationships, as an acceptable method of data collection?," mean scores for total usability and for each of the separate cluster scores on the Usage Rating Profile – Assessment (URP – A) were reviewed. The URP – A (see Appendix E) contained 28 items that were aimed at gathering information about each participant's experience

with DBR. The goal was to measure overall usability, that is the overall acceptability, of DBR as well as the six cluster scores of: Acceptability, Understanding, Home-School Collaboration, Feasibility, System Climate, and System Support.

Paraprofessionals (n = 7) selected a single point from a 6-point Likert type scale to indicate their level of agreement or disagreement with each of the 28 statements. Possible scores on the URP – A ranged from 1 – 6 with a score for all clusters closer to zero reflecting the perception of "no acceptability," with the exception of the system support cluster. On the URP – A, paraprofessionals who report scores closer to 6 perceive DBR to be an acceptable data collection method. With regard to system support, scores closer to 0 reflect stronger agreement regarding acceptability. A summary of the results is presented in Table 10.

Participants slightly agreed in their overall perception of the usability of DBR (Usability Rating M=4.17; SD=.41) suggesting the participants found the tool somewhat acceptable as a tool for assessing student disruptive behavior. Overall, participants most agreed that they felt confident they understood how to implement DBR procedures (Understanding M=5.14; SD=.42). The results also indicated they agreed that the time and effort to conduct DBR was manageable, while also agreeing that DBR would lend itself well to the culture and environment of the school (Feasibility M=4.48; SD=.48; System Climate M=4.29; SD=.34). Participants also agreed DBR was suitable for use in the classroom (Acceptability M=4.29; SD=.64). Furthermore, the results suggest the participants felt DBR could be implemented with a minimal amount of administrative support (System Support M=3.33; SD=.58). Finally, results from participants' ratings on the URP-A indicated that they slightly agreed partnering with the home was required in order to carry out DBR (Home-School Collaboration M=3.48; SD=1.26).

Chapter 5: Discussion

The purpose of the present study was to evaluate paraprofessionals' ability to generate believable data using Direct Behavior Rating to assess functional relationships relative to Systematic Direct Observation and to examine their perceptions regarding DBR's acceptability. Observational data regarding behavior is considered "believable" whenever it demonstrates high levels of agreement and concordance with a commonly accepted gold standard of measurement, such as SDO (Johnston & Pennypacker, 2009). Believability reflects the degree of association between the data sets of two observers, often measured and reported as inter-observer agreement (Hinze & Matthews, 2004). In the case of this study, the DBR data gathered by paraprofessionals was compared to SDO data gathered by an expert observer in order to evaluate the level of agreement, or believability, of the data. Secondly, this study sought to assess users' perceptions regarding the acceptability of DBR.

There exist a growing body of literature documenting the flexibility, efficiency, repeatability, and defensibility of Direct Behavior Rating for use as a tool when collecting data regarding interfering behaviors in schools (Briesch et al., 2016). There is also preliminary evidence to support the applied use of DBR to gather data regarding the function of disruptive behavior, including behavior maintained by attention, escape/avoidance, or access to tangibles (Kilgus, et al., 2016). Results from the current study further supports DBR's use as a means for collecting data regarding overall disruptive behavior. However, the results suggest DBR should be used more cautiously when observing behavior within the framework of Functional Behavior Assessment (FBA), particularly when raters are required to self-evaluate their role within the functional contingency.

Believability of DBR data

The first research question focused on examining the believability between DBR and SDO data regarding disruptive behavior and the consequence(s) perceived most likely related to its occurrence. With regard to overall disruptive behavior, results suggested that the ratings of paraprofessionals using DBR were strongly associated with those recorded by an expert using SDO. However, it is important to note that the results did not indicate precise agreement between observation methods. Nevertheless, this study offers further support for DBR's use as a behavioral observation tool, particularly when monitoring a student's response to a behavioral intervention. Consistent with prior research findings, DBR ratings tended to fall within a range of 10% - 20% of the SDO data, even while paraprofessionals remained responsible for performing their typical classroom duties. This finding is all the more promising given this study observed the behavior of students who display high rates of behavioral disruption in a rather uncontrolled classroom setting. Results of the present study therefore suggest paraprofessionals can use DBR to obtain accurate estimates of the occurrence of disruptive behavior. This finding supports the potential for DBR to be a cost- and time-efficient tool for progressing monitoring student behavior.

However, other important findings from this study suggest that conclusions regarding DBR's use for assessing the potential function of disruptive behavior (i.e., conducting an FBA) must be made more cautiously. Despite a moderate association between the DBR and SDO data sets for behavior maintained by attention from an adult, along with solid agreement regarding occurrence and non-occurrence of the functional contingency, an analysis of interval-by-interval percentage-point differences in agreement between raters suggested the potential exists for large discrepancies between the two observation methods. Results from the current study indicated that given the occurrence of disruptive behavior, nearly half of all DBR ratings differed from

SDO ratings by 30% or more. With regard to behavior maintained by peer attention, the weak association between DBR and SDO data suggested the potential for differences in agreement is even greater. These findings are inconsistent with preliminary research investigating the application of DBR within an FBA framework. Given this lack of consistent agreement regarding the potential consequence targets in this study, inferences made regarding the function of disruptive behavior based upon DBR data may be invalid. Clinical situations that require FBA's be conducted with greater levels of accuracy (i.e., self-injurious behavior, initial assessments) may find SDO offers a more believable and thus, defendable explanation of functional relationships.

A few possible explanations exist as to why stronger levels of agreement were not found between DBR ratings and SDO scores regarding potential functional relationships. First, previous research investigating the application of DBR to assess functional contingencies has tended to rely on having participants view 1 – 2.5-minute-long video clips and then rate the target students' disruptive behavior. Noteworthy, these participants played no direct role within the functional contingency. Because participants in the current study conducted observations of students within an actual classroom and the duration of a given interval was longer (5-minutes), it is possible the longer observation in a live setting resulted in increased demands of participant attention that impacted rater accuracy.

Next, the results from this study indicated paraprofessionals agreed with the expert observer that attention from an adult was the most commonly occurring consequence following disruptive behavior. However, the paraprofessionals' perceptions regarding the behavior and its functional contingency differed from the expert observers' in important ways. First, the paraprofessionals recorded the occurrence of disruptive behavior in fewer observation windows

than did the expert observer. However, the paraprofessionals' perception was that the average percentage of time disruptive behavior occurred during each observation window was greater than that recorded by the expert observer. This suggests that despite observing disruptive behavior less often, the paraprofessionals over-estimated the total percentage of time disruptive behavior occurred per five-minute observation window. In other words, even though the paraprofessionals recorded less incidents of the disruptive behavior, their perception was that the behavior occurred for a larger percentage of time than did the expert observer.

Secondly, when the paraprofessionals observed disruptive behavior, they identified adult attention as the maintaining variable in a larger percentage of observation windows than did the expert observer. As such, their perception was that when disruptive behavior occurred, an adult responded during more incidents of the behavior. However, the paraprofessionals' estimate of the percentage of disruptive behavior met with adult attention per five-minute observation window was less than the expert observer's rating. This suggests that given disruptive behavior, the paraprofessionals indicated attention from an adult had less to do with the occurrence of the behavior than was observed by the expert observer. Collectively, given the occurrence of disruptive behavior, the paraprofessionals may have over-estimated the frequency with which adults responded, while also underestimating the strength with which adult attention influenced the functional contingency.

It seems reasonable that rater error (i.e., perception was accurate but recorded an incorrect rating) may have contributed to this finding. However, it is also possible that rater perception (i.e., perception was inaccurate resulting in an incorrect rating) likely influenced this discrepancy. As such, it seems plausible that the paraprofessional failed to accurately observe the occurrence of the behavior as well as the functional relationship, in this case the role of

attention from an adult. One explanation may be related to the paraprofessional's ability to self-detect their potential role in the functional contingency. As noted above, participants in previous studies generated ratings having viewed video clips of student disruptive behavior and did not have any direct involvement in the situation. In the current study, paraprofessionals remained responsible for their typical job duties while simultaneously completing the DBR ratings. This presumably included responding to student disruptive behavior. It is possible the expert more easily detected this type of response and identified it as a consequence (i.e., attention from an adult) than did the paraprofessional, who may have categorized this as simply a part of their job. Therefore, it is necessary to engage in further discussion of how paraprofessionals' ability to self-detect and evaluate their role within functional contingencies can contribute to errors in rater perception. This line of research could possibly offer important recommendations regarding the nature of the content of training programs necessary for training raters competently.

Another important finding was related to the occurrence and role of peer attention within the functional contingency. Based upon the data, the paraprofessionals underestimated the percentage of observation windows maintained by peer attention, given the occurrence of disruptive behavior. Moreover, they also under-estimated the influence peer attention had on the functional contingency. It is possible that the complexities of observing not only the target student's behavior, but also that of the peers in the room while simultaneously performing other job tasks, may have influenced their capacity to observe and record contingencies involving peer attention. In addition, as it was beyond the scope of this study to gather data regarding the nature of the peer's behavior in the classroom, it is unclear to what extend the behavior of peers in the room may have influenced the believability of the data. Future research should investigate the impact of peer-displayed disruptive behavior on rater perception of target student behavior.

Given the occurrence of disruptive behavior, both the expert and paraprofessional observers recorded a minimal number of instances that were followed by escape/avoidance or by gaining access to a tangible. Therefore, limited conclusions can be drawn regarding the agreement between the SDO and DBR data for disruptive behavior maintained by either contingency. It is important to note there may have been limited opportunities for students to escape/avoid or gain access to a tangible given the nature of the paraprofessionals' training and the climate of the classroom. Paraprofessionals were expected to intervene at the first sign of student disruptive behavior, which most often included verbal redirection designed to avoid an escalation in behavior. Additionally, the highly structured schedule and closely monitored environment likely limited student access to a variety of items and activities. Furthermore, this study did not attempt to control for the type of task students were engaged in during the observation. As such, students may not have been observed during tasks they found to be aversive, thereby reducing the likelihood of the occurrence of escape/avoidance-maintained behavior. These factors may have served to abolish the reinforcing value of escape/avoidance and gaining access to a tangible as part of the functional contingency.

Guidelines for implementation of DBR

Despite the noted imprecision with regard to agreement between DBR and SDO, the data from this study indicated that a majority of the DBR ratings tended to fall within a range of 0% to 20% of the SDO data (comparison scores), particularly with regard to overall disruptive behavior, which is consistent with previous research. Furthermore, a visual analysis of the data (see Figure 1) suggests practitioners would reach similar conclusions regarding the potential function of disruptive behavior regardless of the observation method employed (i.e., DBR or SDO). This lends further support for the potential of DBR as means for assessing functional

relationships (via behavior and consequence ratings) as proposed by Kilgus and others (2016). However, given the findings regarding the level of occurrence agreement and percentage-point differences in agreement between DBR and SDO of the present study, the following guidelines for implementation of DBR should be considered:

1) The use of DBR to assess functional relationships is not supported by the results of this study in situations that demand a high degree of clinical precision and thoroughness in the behavioral assessment process. In such cases, practitioners typically require more precise descriptive and contextual information about student behavior than can be inferred from the results of DBR ratings. Furthermore, the data gathered for these types of situations often carry legal and ethical ramifications, especially when used to assess student behavior that presents significant safety risks (i.e., self-injurious behavior). Therefore, initial Functional Behavior Assessments, as well as those for high-risk target behaviors, should rely on a combination of indirect assessment and direct systematic observation in order to determine functional relationships (Cooper et al., 2007). Given DBR offers estimates of the perceived nature and strength of the function of the target behavior, DBR appears to be better suited to situations in which less precision is required, such as when making low to medium stake decisions and/or monitoring a student's response to intervention over a period of time.
2) The flexibility of conducting DBR makes it better suited to situations in which the occurrence of the targeted student behavior is of low to moderate frequency and/or intensity. A common criticism of SDO is that given the difficulty of predicting the occurrence of a low-frequency behavior, practitioners may fail to observe and record data regarding the target behavior despite conducting multiple observation sessions.

DBR's flexibility to conduct and record on-going, continuous observations across a substantial amount of time increases the probability of collecting data regarding the occurrence of the behavior and its functional relationships. As a result, the risk of conducting an observation during which the target student never displays the target behavior should be minimized. Thus, the limited time and resources in a school may be better allocated. Furthermore, the results of this study suggested paraprofessionals may struggle to self-detect their role within functional contingencies and the influence it plays on the occurrence of the target behavior. Additionally, DBR accuracy may be improved at lower levels of frequency/intensity as demands for intervention are likely to be less intensive. As a result, demands on rater attention may be lower and thus reduce the likelihood of rater error, although additional research to confirm this would need to be conducted.

3) DBR procedures demonstrate promise as an efficient, reliable, and defendable method for recording observations of interfering behavior. Because of its perceived ease of use, applying DBR methodology to the assessment of functional relationships appears to result in a more simplified method of data collection as compared to direct observation procedures. However, the need for a thorough understanding of Functional Behavior Assessment should not be underestimated. Often the goal of FBA is not only to assess the consequences that maintain the behavior (i.e., control the behavior), but also the potential antecedent events that increase the likelihood of the behavior (i.e., predict the behavior). Providing training and support to paraprofessionals prior to and during the data collection process by school personnel

trained in the principals of Applied Behavioral Analysis will likely enhance the reliability of DBR data.

4) Given the results of this study appear to have raised questions regarding the utility of DBR for accurately assessing functional contingencies, it does not mean practitioners should refrain from using it all together. Rather, practitioners are encouraged to recognize the benefits and limitations of Direct Behavior Rating as they relate to any given clinical situation in which it may be used. Although DBR may be useful in some situations that require FBA, more research is needed in order to fully evaluate its utility within a Functional Behavior Assessment framework. When conducting an FBA, consideration of this need for further research is essential when selecting its use as a data collection tool. As such, practitioners should incorporate DBR's use within a multi-method approach to data collection when conducting FBA.

Acceptability of DBR by Paraprofessionals

Finally, the second research question explored paraprofessionals' perceptions of the usability of Direct Behavior Ratings for assessing functional relationships. Results indicated the paraprofessionals found DBR to be generally acceptable as a data collection tool. Other findings based upon the URP – A data included that paraprofessionals reported they felt moderately motivated to use DBR and that its use would be beneficial for the student. Previous research indicates that the acceptability of an assessment may be potentially related to the likelihood it will be used (Miller, et al., 2014). Therefore, this finding that paraprofessionals found DBR acceptable implies they perceive it as usable. These findings support previous research documenting moderate to high rates of acceptability of DBR procedures among school personnel (Briesch, et al., 2016, Miller, et al., 2014; Riley-Tillman et al., 2008).

The results of this study also suggested that paraprofessionals found the procedures and directions for using DBR understandable and that they felt confident about using DBR to assess the students' behavior. Thus, given paraprofessionals appear to understand how to use DBR, they likely find it usable. Paraprofessionals additionally reported that they found DBR methods suitable for the classroom, its procedures feasible, and that it can be performed independent of administrative support. Previous research encourages school personnel to consider the feasibility of an assessment for the setting in which it is to be used (Riley-Tillman et al., 2005). As such, these results suggest paraprofessionals find DBR feasible and therefore, useable as a data collection tool.

Paraprofessionals also felt that DBR would lend itself well to the culture and environment of the school. Furthermore, the paraprofessionals indicated the support of their coworkers when implementing the assessment may enhance its use. However, participants did not indicate they would require collaboration with students' parents in order to implement DBR. This is understandable given that it is not common for paraprofessionals themselves from the setting in which this study was conducted to have regular contact with the student's caregivers.

Overall, paraprofessional scores on the URP-A indicated they perceived DBR as a generally usable tool for recording behavioral observations. This is an important finding as perceived usability is likely related to actual acceptability of an assessment tool and, ultimately, related to ensuring its procedures are implemented with integrity.

Limitations

Results of this study must be interpreted with caution due to many of its limitations. First, the current study included observation of a relatively small sample size of students by a limited number of paraprofessionals. The sample utilized was not diverse in terms of paraprofessionals'

identified ethnicity and students' identified primary educational exceptionality. All paraprofessionals identified as Caucasian. Students were predominately classified with Emotional Disturbance. Furthermore, students included in this study were purposefully selected based upon the presence of a history of significant behavioral concerns. All students were enrolled in a highly specialized setting within a school designed to meet the needs of individuals with severe behavioral issues. The level of support and supervision afforded by this setting is likely not reflective of typical special education classrooms in rural school districts. For instance, paraprofessionals were expected to proactively intervene in student disruptive behavior as to avoid an escalation in the intensity of the behavior. It is plausible these factors influenced the occurrence/ nonoccurrence of student behavior.

Secondly, 43% of the paraprofessionals in this study have more than 4 years of experience working with students who demonstrate severe disruptive behavior. Another 43% have between one and one-and-a-half years of experience. Moreover, these paraprofessionals receive annual training specifically addressing classroom management, de-escalation techniques, and the use of proactive strategies to decrease disruptive behavior. Previous research examining the level of agreement between DBR and SDO for assessing functional relationships has typically used undergraduate and graduate students as participants. The rationale is their backgrounds tend to mirror the background and experience of typical paraprofessionals. Because the paraprofessionals in the current study are more seasoned and receive specialized training, their familiarity with managing disruptive student behavior may be drastically different than a typical paraprofessional. Thusly, their ratings may have been influenced. Together, these factors may limit the generalizability of these findings to students from more typical special education classrooms served by traditional paraprofessionals. Additional research comparing

SDO and DBR used to assess functional relationships in settings more reflective of typical special education settings is warranted.

The next limitation may have stemmed from the version of the DBR form utilized by the paraprofessionals. Despite pre-observation training that instructed them to indicate the estimated percentage of disruptive behavior met with each consequence, the prompt on the DBR form next to each consequence target line read "% of total time." This ambiguity may have influenced paraprofessionals' ratings. However, an informal survey of each paraprofessional conducted post-observation indicated the paraprofessionals understood the instruction (i.e., to rate the percentage of disruptive behavior). Future versions of the DBR form should seek to avoid this potential source of ambiguity by more clearly specifying the intent of the consequence rating.

Another limitation of the current study involves the limited occurrence of disruptive behavior. In addition, the results suggested that a majority of the behavior was followed by attention from an adult or peer, with very little to no disruptive behavior maintained by escape/avoidance or access to tangibles. As discussed earlier, the classroom environment was highly structured and closely monitored, therefore there may have been limited opportunity for students to gain access to preferred items or activities. Moreover, given instructional tasks across classrooms were not controlled, students may not have found the tasks aversive, thus reducing the power of escape/avoidance to serve as a reinforcer. Therefore, limited conclusions can be drawn regarding the use of DBR to assess those functional contingencies.

The way student behavior was observed may have also influenced the occurrence of the behavior. At least one external observer was present during all observational windows. As such, the potential exists for a reactivity effect which may have influenced the ratings recorded by the paraprofessionals as they may have attempted to be more careful in their observations.

Nonetheless, the paraprofessionals were aware that the data they collected was not going to be used to inform treatment decisions in any way. The impact on the accuracy of paraprofessional observations had the paraprofessionals perceived the data was going to be used to drive intervention selection remains unclear at this point. In addition, the presence of an external observer may have influenced the behavior of the students. Despite the external observers regularly entering the classroom on an on-going basis, there remained a possibility their presence could have influenced rates of student behavior.

Finally, another limitation of this study is that observations focused solely on a global definition of disruptive behavior. Given this broad definition of the behavior, additional research is needed that investigates more specific definitions of target behaviors (i.e., physical vs. verbal aggression) that interfere with student success in school. Additionally, this study recorded data related to only four socially mediated forms of reinforcement (consequences). Future research should investigate the reliability of DBR for assessing other potential consequences of interfering behavior.

Future Research

Although the results from the current study demonstrated support for the use of DBR to collect data regarding overall disruptive behavior, less support was found with regard to its potential to inform Functional Behavior Assessments. Thus, decisions based upon DBR data regarding the hypothesized function of disruptive behavior should be made with caution. Relatedly, given that the results suggested similar conclusions regarding the perceived function of the behavior were likely to be made when the data was graphed and visually analyzed, an additional line of research could compare the relative accuracy of functional hypotheses derived from DBR data with those made using descriptive assessment (SDO). Extending this research to

investigation of the effectiveness of interventions informed by analysis of DBR, relative to those based on SDO data, may provide important implications for practice. Future research efforts may wish to determine if interventions derived from hypothesized functional contingencies based on DBR data lead to similar or even more meaningful behavior change than do interventions based upon SDO data.

Additional research should also attempt to develop a better understanding of how diversion of rater attention in a classroom can impact the believability of DBR scores generated by paraprofessionals. A line of research investigating the conditions necessary to optimize the believability of DBR scores when determining functional relationships may provide useful suggestions for its use in the classroom.

Future research should also attempt to investigate application of DBR-like methodology for capturing data regarding antecedent level variables. Improving the efficiency and practicality of reliably observing and recording data related to variables that precede target behavior may allow schools to more proactively intervene in interfering behaviors.

References

Alberto, P. A., & Troutman, A. C. (2003). *Applied behavioral analysis for teachers* (6th ed.). Upper Saddle River, New Jersey: Merrill Prentice Hall.

Alter, P. J., Conroy, M. A., Mancil, G. R., & Haydon, T. (2008). A comparison of functional behavioral assessment methods with young children: Descriptive methods and functional analysis. *Journal of Behavioral Education, 17*, 200-219. doi:10.1007/s10864-008-9064-3

Anderson, C. M., Rodriguez, B. J., & Campbell, A. (2015). Functional Behavior Assessments in schools: Current status and future directions. *Journal of Behavioral Education, 24*. 338-371. doi: 10.1007/s10864-01509226-z

Barton-Arwood, J. H., Wehby, P. L., Gunter, P. L., & Lane, K. L. (2003). Functional behavior rating scales: Inter-rater reliability with students with emotional or behavioral disorders. *Behavioral Disorders, 28*(4), 386-400.

Bijou, S. W., Peterson, R. F., & Ault, M. H. (1968). A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. *Journal of Applied Behavior Analysis, 1*, 175-191.

Briesch, A. M., Chafouleas, S. M., & Riley-Tilman, T. C. (2016). *Direct Behavior Rating: Linking assessment, communication, and intervention.* New York, New York: The Guilford Press.

Briesch, A. M., Chafouleas, S. M., & Riley-Tilman, 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*(3), 408-421.

Burns, M. K., & Parker, D. C. (2014). *Curriculum-based assessment for instructional design:*

Using data to individualize instruction. New York, NY: Guilford Publications.

Carr, E. (1994). Emerging themes in functional analysis of problem behavior. *Journal of Applied Behavior Analysis, 27*, 393-399.

Chafouleas, S. M. (2011). Direct Behavior Rating: A review of the issues and research in its development. *Education and Treatment of Children, 34*(4), 575-591. doi: 10.1353/etc.2011.0034

Chafouleas, S.M., Briesch, A.M., Riley-Tillman, T.C., Christ, T. J., Black, A.C., Kilgus, S.P. (2010). An investigation of the generalizability and dependability of Direct Behavior Rating Single Item Scale (BRS-SIS) to measure academic engagement and disruptive behavior of middle school students. *Journal of School Psychology, 48*. 219-246. doi: 10.1016/j.jsp.2010.02.001

Chafouleas, S. M., Jaffery, R., Riley-Tilman, T.C., Christ, T.J., Sen, R. (2013). The impact of target, wording, and duration on rating accuracy for direct behavior rating. *Assessment for Effective Intervention, 39*(1), 39-53. doi: 10.1177/1534508413489335

Chafouleas, S. M., Kilgus, S. P., Jaffery, R. Riley-Tilman, T. C., Welsh, M., & Christ, T. (2013). Direct Behavior Rating as a school-based behavior screener for elementary and middle grades. *Journal of School Psychology, 51*, 367-385. doi: 10.1016/j.jsp.2013.04.002

Chafouleas, S. M., Miller, F. G., Briesch, A. M., Neugebauer, S. R., & Riley-Tillman T. C. (2012). URP – A: Usage Rating Profile – Assessment. (Assessment and Scoring Guide). Storrs CT: University of Connecticut. Retrieved from http://wp.urp.uconn.edu

Chandler, L. K., & Dahlquist, C. M. (2010). *Functional assessment: Strategies to prevent and*

remediate challenging behavior in school settings (3rd ed.). Upper Saddler River, NJ: Pearson.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied Behavior Analysis*. (2nd ed.) Upper Saddle River, New Jersey: Merrill Prentice Hall.

Couvillon, M. A., Bullock, L. M., & Gable, M. A. (2009). Tracking behavior assessment methodology and support strategies: A national survey of how schools utilize functional behavioral assessments and behavior intervention plans. *Emotional & Behavioral Difficulties*, *14*(3), 215-228. doi: 10.1020/13632750903073459

Daley, E. J., Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review, 26*(4), 554-574.

Dunlap, G., & Fox, L. (2011). Function-based interventions for children with challenging behavior. *Journal of Early Intervention, 33*(4*),* 333-343. doi: 10.1177/1053815111429971

Filter, K. J., & Horner, R. H. (2009). Function-based academic intervention for problem behavior. *Education and Treatment of Children, 32*(1), 1-19.

Fisk, K., & Delmolino, L. (2012). Use of discontinuous methods of data collection in behavioral intervention: Guidelines for practitioners. *Behavior Analysis in Practice, 5*(2), 77-81.

Floyd, R. G., Phaneuf, R. L., & Wilczynski, S. M. (2005). Measurement properties of indirect assessment methods for functional behavioral assessment: A review of research. *School Psychology Review, 34*(1), 58-73.

Gable, R. A., Park, K.L., & Scott, T. M. (2014). Functional behavioral assessment and students

at risk for or with emotional disabilities: Current issues and considerations. *Education and Treatment of Children, 37*(1), 111-135. doi: 10.1353/etc.2-14.0011

Giangreco, M.F. (2007). Paraprofessional Supports in Inclusive Classrooms: Why we need alternatives. http:www.pealcenter.org/images/mgiancreco_powerpoint.pdf accessed 7/24/2017

Giangreco, M.F. & Broer, S.M. (2005).Questionable utilization of paraprofessionals in inclusive schools: Are we addressing symptoms or causes? *Focus on Autism and Other Developmental Disabilities, 20*(1). 10-26.

Giangreco, M. F., Hurley, S. M., & Suter, J. C. (2009). Special education personnel utilization and general class placement of students with disabilities: Ranges and ratios. *Intellectual and Developmental Disabilities, 47*(1), 53-56. doi: 10.1352/2009.47:53-56

Goh, A.E., & Bambara, L.M. (2012). Individualized positive behavior support in school settings: A meta-analysis. *Remedial and Special Education, 33*(5), 271-286. doi: 10.1177/0741932510383990.

Gresham, F.M. (2003). Establishing the technical adequacy of Functional Behavioral Assessment: Conceptual and measurement challenges. *Behavioral Disorders, 28*(3), 282-298.

Gresham, F. M. (2004). Current status and future directions: Directions of school-based behavioral interventions. *School Psychology Review, 33*(3), 326-343.

Harrison, S. E., Riley-Tilman, T. C., & Chafouleas, S. M. (2014). Direct behavior rating: Considerations for rater accuracy. *Canadian Journal of School Psychology, 29*(1), 3-20. doi: 10.1177/0829573513515424

Hinze, J. H., & 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.

Hinze, J. M. (2005). Psychometrics of direct observation. *School Psychology Review, 34*, 507-519.

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

Hinze, J. H., Volpe, R. J., & Shapiro, E. S. (2002). Best practices in the systematic direct observation of student behavior. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV*. Bethesda, MD: National Association of School Psychologist.

Individuals with Disabilities Education Improvement Act (2004). Pub L. No. 108-446, 118 Stat 2647. retrieved from https://www.congress.gov/bill/108th-congress/house-bill/1350/

Iwata, B. A., Dorsey, M. G., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1982). Toward a functional analysis of self-injury. *Analysis and Intervention in Developmental Disabilities, 2*, 3-20.

Johnson, J. M., & Pennypacker, H. S. (2009). *Strategies and tactics of behavioral research* (3rd ed.). New York, NY: Routledge.

Kilgus, S. P., Kazmerski, J. S., Taylor, C. N., von der Embse, N. P. (2016). Use of Direct Behavior Ratings to collect functional assessment data. *School Psychology Quarterly.* Advance online publication. http://dx.doi.org/10.1037/sq0000156.

LeBel, T.J., Kilgus, S.P., Briesch, A.M., Chafouleas, S. (2010). The impact of training on the accuracy of teacher-completed Direct Behavior Ratings (DBRs). *Journal of Positive Behavior Interventions, 12*(1), 55-63. doi: 10.1177/1098300708325265

Lee, Y., Sugai, G., Horner, R. (1999). Using an instructional intervention to reduce problem and off-task behaviors. *Journal of Positive Behavior Interventions, 1*,(4), 195-204.

Lewis, T. J., Scott, T. M., Wehby, J. H., & Wills, H. P. (2014). Direct observation of teacher and student behavior in school settings: Trends, issues, and future directions. *Behavioral Disorders, 39*(4), 190-200.

Losinski, M. Maag, J. W., & Katysiyannis, A. (2014). Examining the effects and quality of interventions based on the assessment of contextual variables: A meta-analysis. *Exceptional Children, 80*(4), 407-422. doi: 10.1177/0014402914527243.

Losinski, M., Maag, J. W., Katsiyannis, A. & Ryan, J. B. (2015). The use of structural behavioral assessment for secondary students exhibiting challenging behaviors. *Education and Treatment of Children, 38*,(2), 149-174.

McCahill, J., Healy, O., Lyndon, S., & Ramy, D. (2014). Training educational staff in functional behavior assessment: A systematic review. *Journal of Developmental Physical Disabilities, 26*, 479-505. doi: 10.1007/s10882-014-9378-0

McIntosh, K., Brown, J. A., & Borgmeier, C. (2008). Validity of functional behavior assessment within a response to intervention framework: Evidence, recommended practice, and future directions. *Assessment for Effective Intervention, 34*, 6-14.

Michael, J. L. (2004). Concepts and principles of behavior analysis. Kalamazoo, MI: Association of Behavior Analysis International.

Miller, F. G., Chafouleas, S. M., Riley-Tillman, C., & Fabiano, G. A. (2014). Teacher perceptions of the usability of school-based behavior assessments. *Council for Exceptional Children, 39*(4), 201-210. doi: 10.1353/etc.2014.0011

National Education Association (2005). *NEA Paraeducator Handbook*. Retrieved from

http:www.nea.org/assets/docs/05espparahandbook.pdf

O'Neill, R.O., Horner, R. H., Albin, R. W., Storey, K. Sprague, J. (1996). *Functional assessment and program development for problem behavior* (2nd Ed.). Belmont, CA: Cengage Learning.

Olympia, D. E., Heathfield, L. T., Jenson, W. R., & Clark, E. (2002). Multifaceted functional behavior assessment for students with externalizing behavior disorders. *Psychology in the Schools, 39*(2), 139-155.

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*, 313-317. doi: 10.1007/s10869-008-9070-5

Riley-Tillman, T.C., Chafouleas, S.M., Sassu, K.A., Chanese, J.A., & Glazer, A.D. (2008). Examining the agreement of Direct Behavior Ratings and Systematic Direct Observation for on-task and disruptive behavior. *Journal of Positive Behavior Interventions, 10*(2), 136-143. doi: 10.1177/1098300707312542

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. doi: 10.1177/1098300710361954

Riley-Tillman, T. C., Kabler, S. M., & Chafouleas, S. M. (2005). Selecting the right tool for the job: A review of behavior monitoring tools used to assess student response to intervention. *The California School Psychologist, 10*, 81-91.

Scott, T. M., & Kamps, D. M. (2007). The future of functional behavior assessment in school

settings. *Behavioral Disorders, 32*, 146-157.

Shaprio, E. S., & Kratochwill, T. R. (2000). *Behavioral assessment in schools: Theory, research, and clinical foundations.* (2nd ed.). New York: Guilford Press.

Shrier, M.D., Anderson, C.M., & Proctor, B. (2001). Evaluating he validity of Functional Behavioral Assessment. *School Psychology Review, 30*(2), 180-192.

Skinner, B. F. (1953). *Science and Human Behavior*. New York, NY: The Free Press.

Steege, M. W., & Watson, T. S. (2009). *Conducting school-based functional behavioral assessments: A practitioners guide* (2nd ed.). New York, NY: Guilford Press.

Steege, M. W., Davin, T., & Hathaway, M. (2001). Reliability and accuracy of a performance-based behavior recording procedure. *School Psychology Review, 30*(2), 252-261.

Suter, J.C. & Giangreco, M.F. (2009). Exploring special education and paraprofessional service delivery in inclusion-oriented schools. *The Journal of Special Education 43*(2). 81-93. doi: 10.1177/0022466907313353

U.S. Department of Education. Office of Special Education and Rehabilitative Services, Office of Special Education Programs, *38th Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2016*, Washington, D.C. 2016. Available from https://www2.ed.gov/about/reports/annual/osep/2016/partsb-c/38th-arc-for-idea.pdf Accessed 8/22/2017

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.

von der Embse, N. P., Scott, E., & Kilgus, S. P. (2014). Sensitivity to change and concurrent

validity of direct behavior ratings for academic anxiety. *School Psychology Quarterly 30*(2), 244-259. doi: 10.1037/spq0000083

Table 1

Demographic Statistics for Participants (Para.) Demographic Variables

Variable and category	*N*	*%*
N	7	--
Gender		
Male	4	57
Female	3	43
Ethnicity		
White	7	100
Black or African American	0	0
Asian	0	0
Other	0	0
Highest educational degree obtained		
High School or equivalent	2	29
Some college credit	4	57
Associates	0	0
Bachelor's	1	14
Master's or Beyond	0	0
Years of experience		
Less than 6 months		
6 months to 1 year	1	14
1 year to 1.5 years	3	43
1.5 years to 3 years	0	0
4 years to 5 years	1	14
More than 5 years	2	29

Table 2

Demographic Statistics for Participants (Students) Demographic Variables

Variable and category	*N*	*%*
N	16	--
Gender		
Male	10	63
Female	6	37
Grade		
Primary (K – 2)	0	0
Elementary (3 – 6)	3	19
Middle (7 – 9)	6	38
High (10 – 12)	7	44
Primary Classification		
Emotional Disturbance	11	69
Intellectual Disability	3	19
Other Health Impaired	1	6
Autism	1	6

Table 3

Number and Percentage of Observation Windows Behavior/Consequence Was Recorded Across Observers

	Disruptive Behavior		Adult Attention		Peer Attention		Escape/ Avoidance		Access to Tangible	
Observer	no.[a]	%	no.[b]	%	no.[b]	%	no.[b]	%	no.[b]	%
Expert (SDO)	47	(48.9)	38	(80.8)	23	(48.9)	0	(0.0)	2	(4.3)
Paraprofessional (DBR)	36	(37.5)	32	(88.9)	13	(36.1)	7	(19.4)	4	(11.1)

Note: SDO = Systematic Direct Observation; DBR = Direct Behavior Rating; no. = number of observation windows

[a]Total number of observation windows each observer recorded the occurrence of disruptive behavior during the 96, 5-minute observations.

[b]Total number of 5-minute observation windows during which disruptive behavior occurred and was followed by the given consequence.

Table 4

Mean Estimated Rating and Standard Deviation for Disruptive Behavior and Consequence Targets Across Observers

	Disruptive Behavior		Adult Attention		Peer Attention		Escape/ Avoidance		Access to Tangible	
Observer	M[a]	SD	M[b]	SD	M[b]	SD	M[b]	SD	M[b]	SD
Expert (SDO)	27.8	(27.2)	78.6	(24.4)	41.3	(27.3)	0.0	(0.0)	75	(35.4)
Paraprofessional (DBR)	34.3	(33.1)	66.6	(38.0)	23.5	(13.8)	42.1	(34.1)	45	(40.4)

Note: SDO = Systematic Direct Observation; DBR = Direct Behavior Rating

[a]For SDO, to calculate the percentage of observation window the disruptive behavior occurred (estimated rating), partial interval scores were calculated for each window by dividing the number of 15-second intervals disruptive behavior was recorded by the total number of intervals observed during each of the 5-minute observation windows. For DBR, paraprofessionals recorded their perception of the estimated percentage of each 5-minute observation window the disruptive behavior occurred.

[b]For each consequence target, SDO scores were calculated by dividing the number of times disruptive behavior was met by each consequence by the total number of disruptive behaviors recorded during each 5-minute observation window. For DBR, scores reflect the paraprofessional's perception of the estimated percentage of disruptive behavior met by each consequence target.

Table 5

Correlations Between Expert and Paraprofessional Ratings of Disruptive Behavior and Consequence Targets

	Expert ratings				
	SDO_{DISR}	SDO_{AA}	SDO_{PA}	SDO_{EA}	SDO_{TAN}
Paraprofessional ratings					
DBR_{DISR}	.886*				
DBR_{AA}		.470**			
DBR_{PA}			.093		
DBR_{EA}				---	
DBR_{TAN}					-.047

Note: DBR = Direct Behavior Rating; SDO = Systematic Direct Observation; AA = Adult Attention; PA = Peer Attention; EA = Escape/Avoidance; TAN = Access to Tangibles; Overall disruptive behavior $n = 96$; Each consequence target $n = 47$

*$p < .001$, two-tailed

**$p = .001$, two-tailed

Table 6

Two by Two Matrix of Inter-Observer Agreement for Disruptive Behavior

	Expert (SDO) Ratings	
Paraprofessional (DBR) Ratings	Occurrence	Nonoccurrence
Occurrence	29	7
Nonoccurrence	18	42

Note: DBR = Direct Behavior Rating; SDO = Systematic Direct Observation

Table 7

Occurrence Agreement Percentage Between DBR and SDO Ratings: Non-weighted and Weighted

Observation Target	Occurrence Agreement Percentage (Non-weighted[a])	Occurrence Agreement Percentage (Weighted[b])
Behavior		
Disruptive Behavior	54%	62%
Consequence		
Adult Attention	79%	82%
Peer Attention	39%	42%
Escape/Avoidance	0%	0%
Access to Tangibles	0%	0%

Note: DBR = Direct Behavior Rating; SDO = Systematic Direct Observation
[a]Non-Weighted Occurrence Agreement Percentage calculated as follows: agreements divided by agreements + disagreements.
[b]Weighted Occurrence Agreement Percentage calculated as follows: agreements divided by agreements + disagreements-criterion (observation windows only the Expert Observer recorded occurrence).
.

Table 8

Number of Observation Windows with Percentage-Point Difference in Agreement Between Expert and Paraprofessional Ratings of Disruptive Behavior

	Percentage-Point Difference in Agreement											
Observation Target	NO	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Disruptive Behavior[a]	42	6	34	7	2	4	1	0	0	0	0	0
Both recorded[b] (Occurrence agreement)		6	14	5	1	2	1	0	0	0	0	0
Only Expert recorded[c] (Occurrence disagreement)		0	20	2	1	2	0	0	0	0	0	0

Note: NO = nonoccurrence;
[a]Reflects the number of observation windows with percentage-point difference between rating estimates for the 96, 5-minute observation windows, regardless of agreement between observers.
[b]Reflects the number of observation windows with percentage-point difference between rating estimates for the 29, 5-minute observation windows both the Expert and Paraprofessional recorded the occurrence of disruptive behavior (occurrence agreement).
[c]Reflects the number of intervals with percentage-point differences between rating estimates for the 25, 5-minute observation windows the Expert and Paraprofessional disagreed regarding the occurrence of disruptive behavior (occurrence disagreement). Includes only the observation windows during which the Expert Observer recorded the occurrence of disruptive behavior.

Table 9

Number of Observation Windows with Percentage-Point Difference in Agreement Between Expert and Paraprofessional Ratings for Each Consequence Target

		Percentage-Point Difference in Agreement[a]										
	NR	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Consequence Target												
Adult Attention	8	6	7	4	5	1	4	2	2	3	1	4
Peer Attention	22	0	3	8	2	3	3	2	1	1	0	2
Escape/Avoidance	40	0	1	2	2	0	0	0	0	1	0	1
Access to Tangibles	41	0	1	1	0	0	2	0	0	0	0	2

Note: NR = no relationship observed between behavior and consequence

[a]Reflects the number of observation windows with percentage-point difference between rating estimates for the 47, 5-minute observation windows during which at least the Expert Observer recorded the occurrence of disruptive behavior.

Table 10

Descriptive Statistics for Usage Rating Profile - Assessment

Usability Dimension	M	SD
Total Usability	4.17	0.41
Acceptability	4.29	0.64
Understanding	5.14	0.42
Home-School Collaboration	3.48	1.26
Feasibility	4.48	0.48
System Climate	4.29	0.34
System Support[a]	3.33[a]	0.58

Note: Aides responded to questions on the URP-A using a 6-point scale: 1= strong disagree; 2= disagree; 3= slightly disagree; 4= slightly agree; 5= agree; 6= strongly agree. n = 7

[a]Items composing the System Support factor were reverse scored.

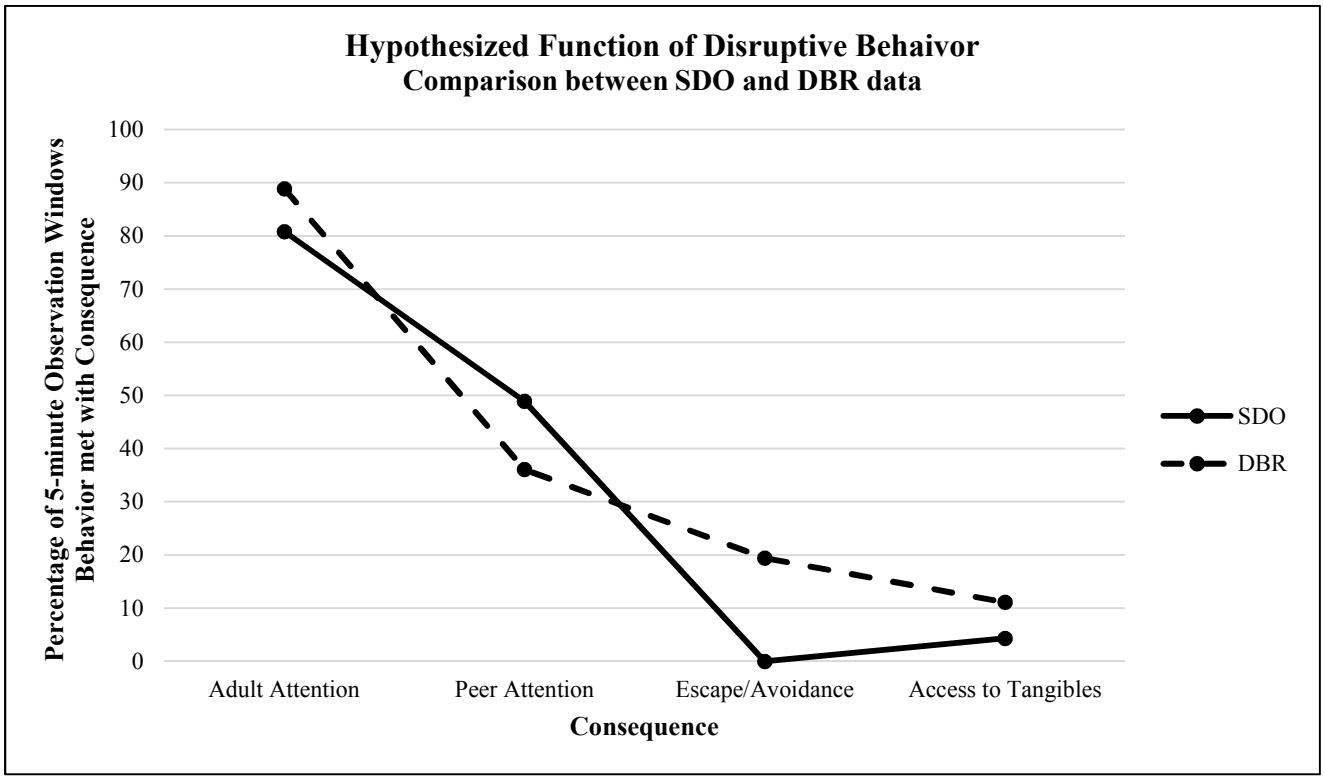


Figure 1. Comparison of the percentage of 5-minute observation windows disruptive behavior was met with each consequence.

Appendix A

Participant (Para) Consent Form

Participation of Educational Paraprofessionals (Classroom Aides)

You are invited to be in a research study of paraprofessionals' use of Direct Behavior Ratings (DBR) to collect data regarding functional behavior assessment. You were selected as a possible participant because you are 21 years of age or older and an employee at the School operated by Agency. We ask that you read this form before agreeing to be in this study.

This study is being conducted by Brian South, MA, Alfred University, Alfred, NY, 14802 under the supervision of Mark Fugate, Ph. D, Licensed Psychologist, Dissertation Chair, Alfred University, Alfred, NY, 14802.

Background Information

The current study will investigate the level of agreement between data regarding functional relationships collected by paraprofessionals using Direct Behavior Ratings and data gathered by an external observer using Systematic Direct Observation. Further, the study seeks to determine paraprofessionals' perceived acceptability of DBR as a behavioral observation tool.

Procedures

If you agree to participate in this study, we ask that you agree to actively participate in an approximate one-hour training session regarding functional behavior assessment data collection methodology, to diligently observe and record information regarding student behavior, and to fill out a questionnaire in full and be forthright in your answers. Completion of this study is estimated to involve participation in a one-hour training session, two 30-minute observations of student behavior, and an additional 5-minute period to complete a 2-page questionnaire.

Risks and Benefits of Being in the Study

While unlikely, it is possible that you may feel discomfort while observing student behavior, filling out the DBR form, and/or when considering some of the questionnaire items. You are free to discontinue your participation at any time during the study simply by discontinuing completion of the DBR form and/or exiting the survey. In the unlikely event that participation in this study causes mild stress, the researchers suggest that you stop collecting data. Only resume data collection if you feel you are able to do so. Participation in this study may provide you with some additional knowledge about research related to functional behavior assessment data collection methodology and your participation will hopefully add to this knowledge base.

Confidentiality

The records of this study will be kept private. In any sort of report that might be published, it will not include any information that will make it possible to identify any participant. Research records will be kept in a locked cabinet and office; only the researcher will have access to the records. Electronic data files will be password-protected. Records will be kept for at least three years after completion of the study, after which records may be destroyed at the discretion of the researcher.

Research information that identifies participants may be shared with the Human Subjects Review Committee (HSRC) and others who are responsible for ensuring compliance with laws and regulations related to research, including members of the researcher's Dissertation Committee. The Dissertation Committee will also have access to the results of the study (data) and commentary regarding those results will be included in the final Dissertation Report, which will be published by the researcher.

Voluntary Nature of the Study
Your decision whether or not to participate will not affect your current or future relations with School, Agency, and/or Alfred University. If you decide to participate you are free to withdraw at any time without penalty.

Subject compensation for participation
As compensation for your participation in this study, you will receive a $30.00 gift card to a local, national retail store.

Contacts and Questions
If you have questions about your participation in this study that you would like to ask before participating, please contact the researcher electronically at bs7@alfred.edu or faculty sponsor Dr. Mark Fugate, Licensed Psychologist, Dissertation Chair at ffugate@alfred.edu. If you have any questions now, or later, related to the integrity of the research, (the rights of research subjects or research-related injuries, where applicable), you are encouraged to contact Dr. Steve Byrne, Chair of the Alfred University Human Subjects Research Committee, at (607) 871-2212 or electronically at HSRC@alfred.edu.

Statement of Consent: I have read the above information. I consent to participate in the study.

_____	_____	_____
Printed Name	Signature	Date

Appendix B

Parent/Guardian Opt Out Permission Letter

OPT OUT PERMISSION LETTER

Brian South
Doctoral Candidate
School Psychology
Alfred University
bs7@alfred.edu

RE: Important information about a research project being conducted at your child's school
Alfred University Human Subjects Research Committee Approval # [INSERT NUMBER HERE]

Dear Parent/Guardian,

Paraprofessionals use of Direct Behavior Ratings (DBR) to collect functional behavior assessment data: Believability and acceptability

I am writing to you about the research I am conducting as part of my dissertation at Alfred University. I am also a long-term employee of Agency most recently serving as a Clinical Consultant to several programs.

I am interested in helping school personnel find ways to accurately observe student behavior in classrooms. I want to find ways to help classroom aides observe student behavior that are as accurate as possible while also making the best use of time and the resources available in schools.

I have approached the school your child attends and explained the purpose of the study, and the school has kindly agreed to distribute these letters to you.

Please read the information sheet attached to this letter. You will see that my research involves observation of students during normal lessons and there will be no direct contact with any of the children. I hope therefore that you will agree to your child being involved in my research.

If you have any further questions about the research, please contact me at bs7@alfred.edu. If you have any concerns about the research please contact my supervisor: Mark Fugate, Ph. D, Alfred University, ffugate@alfred.edu.

If you would prefer that your child does not take part, please sign and return the form enclosed.

Yours sincerely,

Brian South, MA
Doctoral Candidate
School Psychology

INFORMATION SHEET

Paraprofessionals use of Direct Behavior Ratings (DBR) to collect functional behavior assessment data: Believability and acceptability

Researcher: Brian South, MA

Supervisor: Dr. Mark Fugate, Ph D.

This information sheet explains why I am doing this research and what it will involve. Please take time to read this information carefully together with your child. Please contact me if there is anything that is not clear or if you would like more information. Thank you for reading this.

What is this study about?

I am trying to find out more about improving the way schools go about observing student behavior when performing a Functional Behavior Assessment. I want to compare different methods of observation while also trying to learn about things that might impact aides' accuracy.

How will my child be involved?

Your child will be working in a normal lesson and your child's learning will not be affected in any way. Two students will be randomly selected per classroom. The classroom aide will observe each student separately and then make a mark on a sheet to indicate if the student's behavior was disruptive during the class period. Your child may or may not be selected for observation. Likely, your child will not be able to tell if their behavior is being observed. Additionally, an expert observer (who is a Board Certified Behavior Analyst employed by (Agency) will also be observing the student's behavior. I will coordinate with teachers to ensure minimal disruption within the classroom.

Who will have the access to the research information (data)?

Any information collected will be used for the purposes of this research project only. None of the information gathered will be used to inform any educational or treatment decisions concerning your child. I will not keep information about your child that could identify them to someone else. The data will be stored safely and will be destroyed at my discretion after 3 years. The data will only be used for my work and will only be seen by myself, my supervisor and members of my dissertation committee. Results of my work will be published in my final Dissertation report; however, no information that could identify any participant will be included.

Will my participation affect my child's relationship with Agency in any way?

Whether or not you decide to allow your child to participate in this study will have no impact on your current or any future relationship with Agency, School, and/or Alfred University. Participation is entirely voluntary.

Who has reviewed the study?
The research study has been approved under the regulations of Alfred University Human Subjects Research Committee. Studies involving children are subject to the fullest review by the committee. I have also obtained consent from the classroom aides and the school principal before beginning the study.

Who do I speak to if I have questions about this research?
If you would like more information or have any problems with this research, please let me know. You can contact me via the University at the following address:

Brian South, MA
bs7@alfred.edu

If you would like to speak to someone else you can contact my supervisor:

Mark Fugate, Ph. D.
ffugate@alfred.edu

If you have any complaints about the research, please contact _____.

What do I do next?
If you agree for your child to be involved in my research you do not need to do anything. Please keep this information for reference.

If you **do not** want your child to be involved in this research, please complete one copy of the attached form and return it in the self-addressed, stamped envelope no later than 6/19/2017, Please keep the letter, information sheet and the 2nd copy of the form for your information. If I do not receive a signed copy of the slip by 6/19/2017, I will assume you consent for your child to take part.

Can you change your mind?
You and your child have the right to withdraw from the research at any time. Should you decide after the study that you no longer want your child's data included, simply contact me and I will withdraw it.

Thank you very much for your time.

PARENT OPT-OUT FORM
(1ST COPY FOR RETURN TO RESEARCHER)

Paraprofessionals use of Direct Behavior Ratings (DBR) to collect functional behavior assessment data: Believability and acceptability

I have read the information about the study and talked about this with my child.

Please check the box below.

I am not willing for my child to take part in the study. ☐

Name of child: ……

School: …….

Signature of parent/guardian: …….

Date: ……..

PARENT OPT-OUT FORM
(2ND COPY FOR PARENT/GUARDIAN RECORDS)

Paraprofessionals use of Direct Behavior Ratings (DBR) to collect functional behavior assessment data: Believability and acceptability

I have read the information about the study and talked about this with my child.

Please check the box below.

I am not willing for my child to take part in the study. ☐

Name of child: ……

School: …….

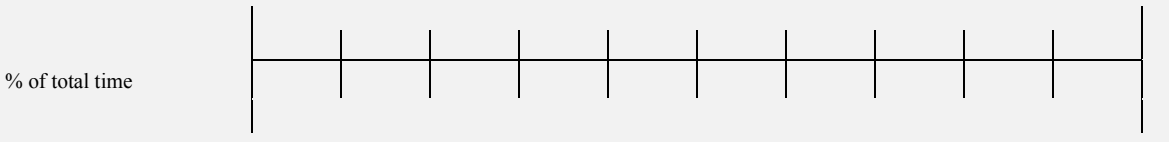
Signature of parent/guardian: …….

Date: ……..

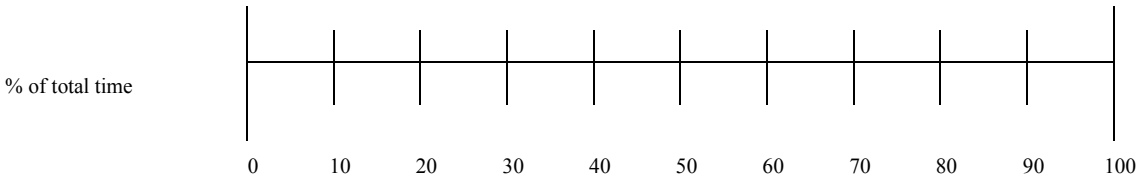
Appendix C: Direct Behavior Rating (DBR) Form

Date: M T W Th F	**Student:**		**Activity Description:**		
	Aide:				
Observation time:	*Start time:*		*Stop Time:*	*No observation*	
Behavior Description:	**Disruptive Behavior:** any student action that interrupts regular school or classroom activity. For example: out of seat, fidgeting, playing with objects, acting aggressively, talking/yelling about things that are unrelated to classroom instruction, taking things that do not belong to them.				
Possible functions:	**Adult Attention** is positive, negative, or neutral adult reaction that can be verbal or nonverbal (e.g., reprimand, redirection to work). **Peer Attention** is positive, negative, or neutral peer reaction that can be verbal or nonverbal (e.g., talking. laughing, arguing). **Escape/Avoidance** is the removal or avoidance of a task, activity, or performance expectation(s) (e.g., removal of academic materials, given permission to leave room). **Access to tangibles/activities** is the acquisition of items or activities (e.g., toys, food, prizes, preferred tasks).				
Directions: Place a mark along the line that best reflects the percentage of total time the student exhibited disruptive behavior. Note that the percentages do not need to total 100% across consequences.					

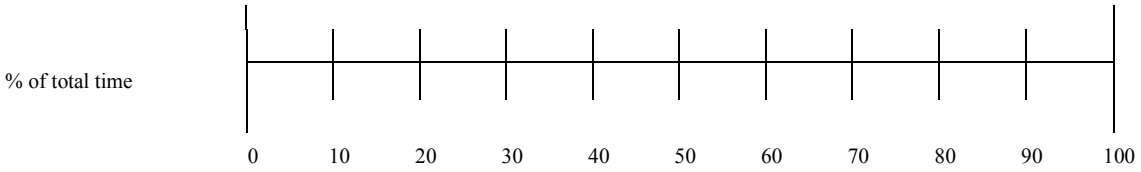
Disruptive



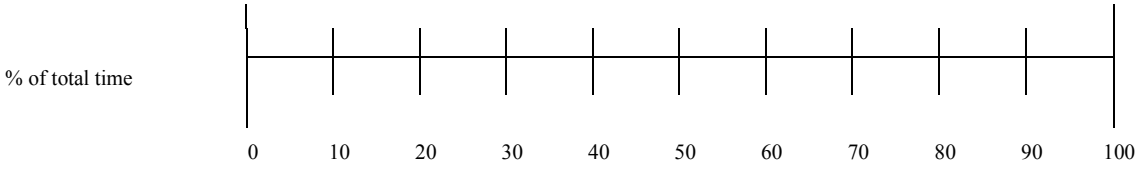
Adult Attention



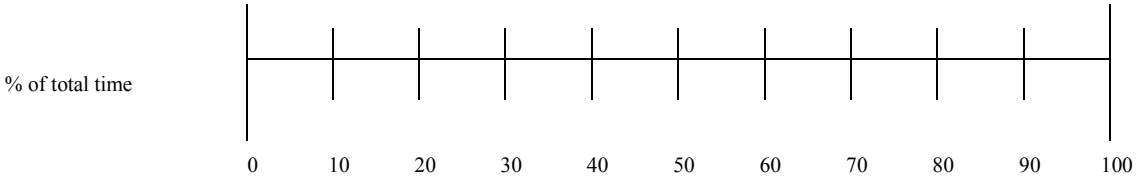
Peer Attention



Escape/Avoidance

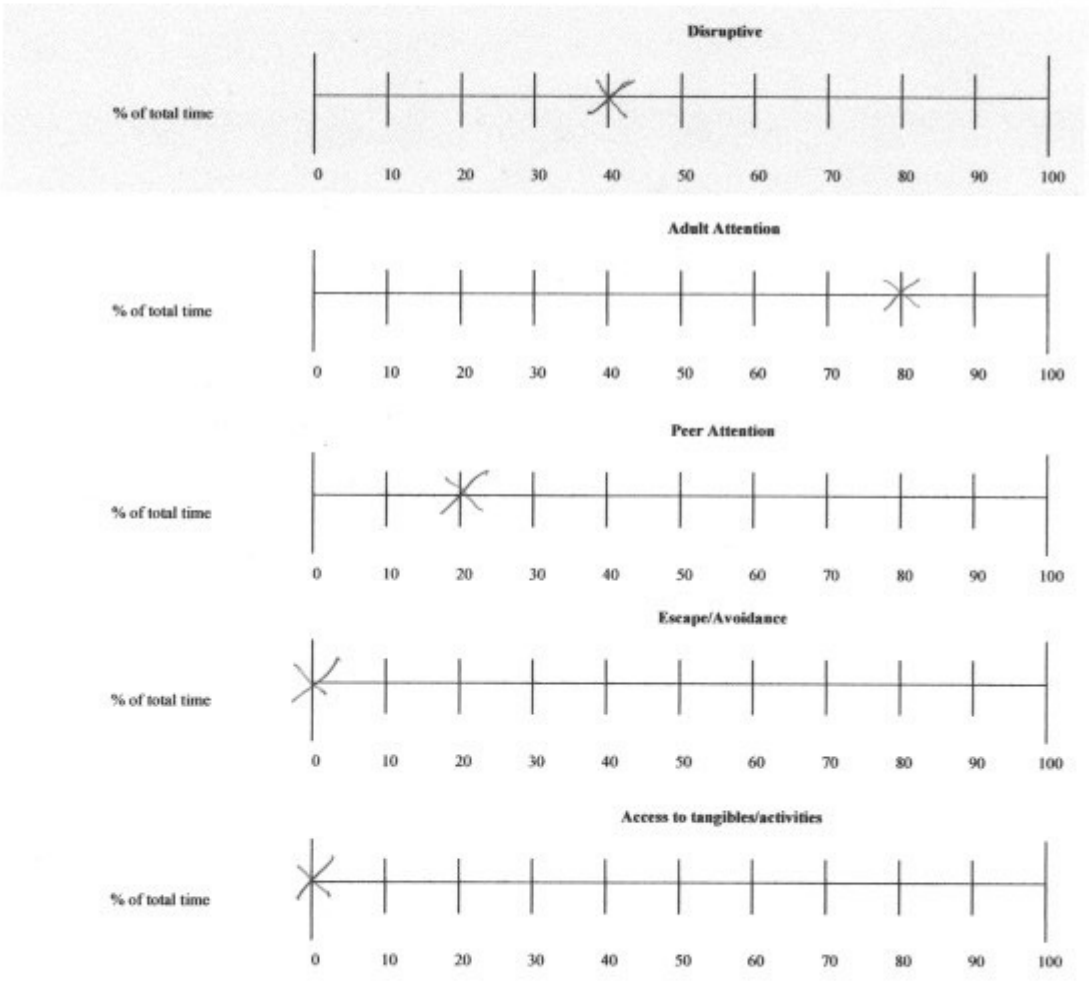


Access to tangibles/activities



Direct Behavior Rating (DBR) Form – Sample of completed form

Date: 06/26/17 M T (W) Th F	**Student:** Billy S. **Aide:** Sally S.		**Activity Description:** Small group, math; 3 students		
Observation time:	*Start time:* 9:00 AM		*Stop Time:* 9:05 AM	*No observation*	
Behavior Description:	**Disruptive Behavior:** any student action that interrupts regular school or classroom activity. For example: out of seat, fidgeting, playing with objects, acting aggressively, talking/yelling about things that are unrelated to classroom instruction, taking things that do not belong to them.				
Possible functions:	**Adult Attention** is positive, negative, or neutral adult reaction that can be verbal or nonverbal (e.g., reprimand, redirection to work). **Peer Attention** is positive, negative, or neutral peer reaction that can be verbal or nonverbal (e.g., talking. laughing, arguing). **Escape/Avoidance** is the removal or avoidance of a task, activity, or performance expectation(s) (e.g., removal of academic materials, given permission to leave room). **Access to tangibles/activities** is the acquisition of items or activities (e.g., toys, food, prizes, preferred tasks).				
Directions: Place a mark along the line that best reflects the percentage of total time the student exhibited disruptive behavior. Note that the percentages do not need to total 100% across consequences.					



Appendix D

SDO Observation Form

SYSTEMATIC DIRECT OBSERVATION FORM Observation date: _____

EXPERT OBSERVER FORM Start time: _____

Student: _____ End time: _____

Clrm: _____ Observation Type: Inter-Aggreement Expert

PERIOD	INTERVAL (15-seconds)	DISRUPTIVE BEHAVIOR	CONSEQUENCE			
			Adult Attention	*Peer Attention*	*Escape/Avoidance*	*Tangible*
		0 = no; 1 =yes	FREQUENCY	FREQUENCY	FREQUENCY	FREQUENCY
1	1					
	2					
	3					
	4					
	5					
	6					
	7					
	8					
	9					
	10					
	11					
	12					
	13					
	14					
	15					
	16					
	17					
	18					
	19					
	20					
2	1					
	2					
	3					
	4					
	5					
	6					
	7					
	8					
	9					
	10					
	11					
	12					
	13					
	14					
	15					
	16					
	17					
	18					
	19					
	20					

Student: _____ Observation Date: _____

PERIOD	INTERVAL (15-seconds)	DISRUPTIVE BEHAVIOR	CONSEQUENCE			
			Adult Attention	*Peer Attention*	*Escape/Avoidance*	*Tangible*
		0 = no; 1 =yes	FREQUENCY	FREQUENCY	FREQUENCY	FREQUENCY
3	1					
	2					
	3					
	4					
	5					
	6					
	7					
	8					
	9					
	10					
	11					
	12					
	13					
	14					
	15					
	16					
	17					
	18					
	19					
	20					
4	1					
	2					
	3					
	4					
	5					
	6					
	7					
	8					
	9					
	10					
	11					
	12					
	13					
	14					
	15					
	16					
	17					
	18					
	19					
	20					

Student: _____ Observation Date: _____

PERIOD	INTERVAL (15-seconds)	DISRUPTIVE BEHAVIOR	CONSEQUENCE			
			Adult Attention	*Peer Attention*	*Escape/Avoidance*	*Tangible*
		0 = no; 1 =yes	FREQUENCY	FREQUENCY	FREQUENCY	FREQUENCY
5	1					
	2					
	3					
	4					
	5					
	6					
	7					
	8					
	9					
	10					
	11					
	12					
	13					
	14					
	15					
	16					
	17					
	18					
	19					
	20					
6	1					
	2					
	3					
	4					
	5					
	6					
	7					
	8					
	9					
	10					
	11					
	12					
	13					
	14					
	15					
	16					
	17					
	18					
	19					
	20					

SDO Observation form – Sample of page 1 of completed form

SYSTEMATIC DIRECT OBSERVATION FORM
EXPERT OBSERVER FORM
Student: BILLY S.
Clrm: 216

Observation date: 06/26/17
Start time: 9:00 AM
End time: 9:00 AM
Observation Type: Inter-Agreement (Expert)

PERIOD	INTERVAL (15-seconds)	DISRUPTIVE BEHAVIOR	CONSEQUENCE			
			Adult Attention	*Peer Attention*	*Escape/Avoidance*	*Tangible*
		0 = no; 1 =yes	FREQUENCY	FREQUENCY	FREQUENCY	FREQUENCY
1	1	0				
	2	0				
	3	1	1			
	4	1	1			
	5	1	1			
	6	0				
	7	0				
	8	0				
	9	0				
	10	1		1		
	11	1		1		
	12	1	1			
	13	1	1 1			
	14	1	1 1			
	15	0				
	16	0				
	17	0				
	18	0				
	19	0				
	20	0				
2	1	0				
	2	0				
	3	0				
	4	0				
	5	0				
	6	0				
	7	0				
	8	0				
	9	0				
	10	0				
	11	0				
	12	0				
	13	0				
	14	0				
	15	0				
	16	0				
	17	0				
	18	0				
	19	0				
	20	0				

Appendix E

Usage Rating Profile-Assessment

Directions: Consider the Direct Behavior Rating form when answering each of the following statements. Circle the number that best reflects your agreement with the statement, using the scale provided below.

		Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
1.	This assessment is an effective choice for understanding a variety of problems.	1	2	3	4	5	6
2.	I would need additional resources to carry out this assessment.	1	2	3	4	5	6
3.	I would be able to allocate my time to implement this assessment.	1	2	3	4	5	6
4.	I understand how to use this assessment.	1	2	3	4	5	6
5.	A positive home-school relationship is needed to use this assessment.	1	2	3	4	5	6
6.	I am knowledgeable about the assessment procedures.	1	2	3	4	5	6
7.	The assessment is a fair way to evaluate the child's behavior problem.	1	2	3	4	5	6
8.	The total time required to implement the assessment procedures would be manageable.	1	2	3	4	5	6
9.	I would not be interested in implementing this assessment.	1	2	3	4	5	6
10.	My administrator would be supportive of my use of this assessment.	1	2	3	4	5	6
11.	I would have positive attitudes about implementing this assessment.	1	2	3	4	5	6
12.	This is a good way to assess the child's behavior problem.	1	2	3	4	5	6
13.	Preparation of materials needed for this assessment would be minimal.	1	2	3	4	5	6

14.	Use of this assessment would be consistent with the mission of my school.	1	2	3	4	5	6
15.	Parental collaboration is required in order to use this assessment.	1	2	3	4	5	6
16.	Material resources needed for this assessment are reasonable.	1	2	3	4	5	6
17.	I would implement this assessment with a good deal of enthusiasm.	1	2	3	4	5	6
18.	This assessment is too complex to carry out accurately.	1	2	3	4	5	6
19.	These assessment procedures are consistent with the ways things are done in my system.	1	2	3	4	5	6
20.	Use of this assessment would not be disruptive to students.	1	2	3	4	5	6
21.	I would be committed to carrying out this assessment.	1	2	3	4	5	6
22.	The assessment procedures easily fit in with my current practices.	1	2	3	4	5	6
23.	I would need consultative support to implement this assessment.	1	2	3	4	5	6
24.	I understand the procedures of this assessment.	1	2	3	4	5	6
25.	My work environment is conducive to implementation of an assessment like this one.	1	2	3	4	5	6
26.	The amount of time required for record keeping would be reasonable.	1	2	3	4	5	6
27.	Regular home-school communication is needed to implement these assessment procedures.	1	2	3	4	5	6
28.	I would require additional professional development in order to implement this assessment.	1	2	3	4	5	6

BRIAN N. SOUTH, MS Ed., M.A., C.A.S.
southb49@gmail.com

EDUCATION

Alfred University, Alfred, NY (NASP & APA accredited)
Doctor of Psychology in School Psychology December 2017
Certificate of Advance Study in School Psychology May 2017
Masters of Arts in School Psychology May 2015

St. Bonaventure University, Allegany, NY
Master of Science in Education, Community Counseling Track May 2003

University of Pittsburg at Bradford, Bradford, PA
Bachelor of Science in Psychology May 1998

PROFESSIONAL CREDENTIALS

New York State Provisional Certificate: School Psychologist
Board Certified Behavioral Analyst (BCBA)

SELECT CLINICAL EXPERIENCE

The Rehabilitation Center, Allegany, NY
Psychologist September 2017 – present

Portville Central School District, Portville, NY.
School Psychologist Doctoral Intern August 2016 – July 2017

Lea R. Powell Institute's Child and Family Services Center, Alfred, NY
Graduate Clinician Supervisor, Advanced Practicum August 2015 – May 2016
Graduate Clinician August 2014 – May 2015

SELECT RELATED RESEARCH EXPERIENCE

Presentation:

South, B. & Fugate, M. (2017). *Small Steps in Treating Co-Morbid Psychosocial Disorders and Intellectual Disabilities.* Mini-skills proposal accepted at the 2017 National Association of School Psychologist (NASP) Annual Conference in San Antonio, Tx.

Fugate, M., South, B. N., & Brimstein, A. (2016). *Connecting the Dots: Applied Behavioral Analysis and Academic Assessment.* Poster accepted at the 2016 National Association of School Psychologist (NASP) Annual Conference in New Orleans, La.

Lecture:

PSYC 629: Social, Emotional, & Behavioral Assessment (Graduate). Topic: *Functional Behavioral Assessment Training*. Alfred University, Alfred, NY. 2014, 2015, 2016, & 2017.

SELECT AWARDS

Lea R. Powell Honors Award, Excellence Among Graduate Students May 2017