April 2019

An Evaluation of the Individualized Behavior Rating Scale Tool (IBRST) in Inclusive Classroom Settings

Jessica L. Moore
University of South Florida, jessicalynnm@mail.usf.edu

An Evaluation of the Individualized Behavior Rating Scale Tool (IBRST) in Inclusive Classroom Settings

by

Jessica L. Moore

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Applied Behavior Analysis
Department of Child and Family Studies
College of Behavioral and Community Sciences
University of South Florida

Major Professor: Kimberly Crosland, Ph.D, BCBA-D
Rocky Haynes, Ph.D, BCBA-D
Rose Iovannone, Ph.D, BCBA-D
Terry Blackmon, Ph.D, BCBA-D

Date of Approval:
April 1, 2019

Keywords: concurrent validity, data collection, self-monitoring, problem behavior, on-task behavior

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

One of the greatest challenges facing school staff is problem behavior in the classroom (Public Agenda, 2004). Children who engage in problem behavior in the classroom setting greatly challenge teachers and diminish the ability to learn. This study evaluated the effects of self-monitoring using the Individualized Behavior Rating Scale Tool (IBRST) on problem behavior and on-task behavior in a classroom setting using a multiple-baseline across participants design. This study also evaluated the extent to which students self-rating on the IBRST correlated with direct observation data. Results indicate that self-monitoring using the IBRST was an effective strategy for increasing on-task behavior and decreasing problem behavior for all three students. Results also indicate that the IBRST may be an accurate and reliable means of measuring data when direct observation data may not be feasible or possible. There were 56/60 perfect agreements, with the other four opportunities being only 1-pt value apart. Limitations and future research are discussed.

Keywords: individualized behavior rating scale tool, data collection, self-monitoring, problem behavior, on-task behavior

CHAPTER ONE: INTRODUCTION

Over the course of a 180-day school year, teachers make a concentrated effort to provide learning environments that are stable, positive, and effective. School is a setting which provides the opportunity to learn academics, provide role models, and develop long-lasting social relationships. Although school has many positive qualities, it also provides the context for significant difficulties for educators. One of the greatest challenges facing school staff is managing behavior in the classroom (The New Teacher Project, 2013).

Children who engage in problem behavior in the classroom setting greatly challenge teachers and diminish the ability to learn (Chafee, Briesch, Johson, & Volpe, 2017). According to Public Agenda (2004), 77% of teachers polled stated their teaching practices would be more effective if they did not have to deal with discipline problems. In addition, 1 in 3 teachers have considered quitting the profession entirely because student behavior was such an obstacle. Further, over 40% of teachers report they feel "not at all prepared" or "only somewhat" prepared to handle behavioral issues in the classroom (Greenberg, Putnam, & Walsh, 2014). Although individuals with severe problem behavior represent only 1-5% of the school population, they can account for up to 50% of all incidents handled by school staff (Sugai, Sprague, Horner, & Walker, 2000).

Research shows that teachers of individuals with disabilities may face an even greater challenge. Many studies have indicated that children with disabilities are more likely than their typically-developing peers to present problem behaviors and that these differences in behavior

tend to increase over time (Emerson & Einfeld, 2010; Fauth, Platt, & Parsons, 2017; Landa, Gross, Stuart, & Faherty, 2013). As such, it is evident there is a great need for schools to develop strategies to address behavior concerns for both students with disabilities and teachers alike.

The Individuals with Disabilities Education Act (IDEA) is a federal law first enacted in 1975 and reauthorized in 2004, which requires schools to serve the needs of students with disabilities. By 2017, 6 million students in the United States had received services under this law, including over 1.1 million students with developmental disabilities (US Department of Education, 2017). The purpose of IDEA is to ensure all children with disabilities have a free, appropriate education, which emphasizes services to prepare students for future education, employment, and independent living skills. Among these provisions includes the right of the child whose behavior impedes that child's learning or the learning of others to behavioral interventions and supports to address said behavior (Individuals with Disabilities Education Act, 2004).

Although required by law, many schools lack the ability to identify and implement systems and practices that effectively meet the needs of all students (Sugai & Horner, 1995; Walker et al., 1996). In addition, there has been an increased demand on collecting data related to student's functional progress, especially after behavioral and academic interventions have been implemented. These data are used to make decisions regarding the provision of services for individuals with behavior problems within classrooms, typically within the problem-solving framework. The problem-solving framework was originally developed by Deno in 1989 and includes four major steps (Deno, 1995). The four steps include: 1) identify the problem; 2) analyze the problem; 3) develop and implement interventions that address the problem; and 4)

evaluate response to intervention and adjust as necessary (Iovannone, Greenbaum, Wang, Dunlap, & Kincaid, 2014).

Although the problem-solving framework readily applies to academic goals as data can be collected via standardized tests and other curriculum-based measurements, this is not as simple for evaluating progress towards behavioral objectives. There have been a couple of reasons why a standardized behavioral measurement has been so difficult to obtain. First, behavioral outcomes are often evaluated differently across the levels of the educational context (e.g. teacher, school-wide, district-wide) and can be influenced by political views and culture of a geographical region. A second issue is that many measures that currently exist are not meant to be used as continuous assessment of behavioral progress, rather only as a means of determining whether a child is eligible for special education services (Iovannone et al., 2014). As behavioral interventions are provided for as part of IDEA, it is essential that a feasible, yet reliable measurement strategy to be used in the context of the school environment for evaluating behavioral outcomes of individual students be developed.

Behavioral Interventions in Schools

Under IDEA, it is required that individuals who engage in behavior which disrupts the learning of themselves and/or other children receive behavioral intervention services. These services must be developed based on evidence-based practices and utilize data-driven treatment strategies (IDEA, 2004). In addition, these services should focus on positive, proactive strategies. One framework commonly used in schools is Positive Behavioral Interventions and Support (PBIS) (Sugai et al., 2001). PBIS utilizes multi-tiered levels of support to decrease the likelihood of problem behavior occurring and minimize the need for intensive, individualized

supports for all students (Carr et al., 2002). PBIS is an application of behavior-based systems approach to create environments which make problem behavior less effective, efficient, and relevant while making desired behaviors more functional (There are three main tiers within PBIS including Tier 1 (primary), Tier 2 (secondary), and Tier 3 (tertiary) (Sugai et al., 2001).

Tier 1 consists of school-wide interventions to develop positive social culture (Horner & Sugai, 2015). Tier 1 interventions include the following: creating and defining a small number of behavioral expectations, teaching those expectations, developing a reinforcement system based on those expectations, implementing a system for interrupting and re-directing behavior not consistent with the expectations, and developing an efficient data collection system (Horner & Sugai, 2015). It is expected that 80-85% of all students will respond appropriately to Tier 1 interventions alone; however, all students experience Tier 1 behavior supports.

Tier 2 interventions are moderate intensity interventions and include further structure, additional antecedent prompts, and an elevated level of training and reinforcement of following behavioral expectations (Horner & Sugai, 2015). These interventions are intended for small groups of students. It is estimated that an additional 10-15% of the remaining students will be served adequately under Tier 2 (Debnam, Pas, & Bradshaw, 2012). Finally, students who require additional, individualized support need Tier 3 strategies. One of the core practices that has been shown to be an effective individualized intervention is the functional behavior assessment (FBA) and function-linked behavior intervention plan (BIP). It is estimated that 5% of the school population receive services at this level. Supports at all levels should be evidence- and function-based (Debnam et al., 2012).

According to Horner and colleagues (2005), there are three types of evidence-based strategies used to address problem behavior. The first are antecedent manipulations. These

strategies modify environmental variables to make problem behavior less likely to occur by making the problem behavior less efficient and/or less effective (Carr et al., 2002). This is often accomplished by reducing or eliminating the establishing operation (EO) for the problem behavior, reducing response effort for the appropriate behavior, or re-arrangement of variables that signal reinforcement is available for the problem behavior (eliminating the discriminative stimulus). Common antecedent manipulations in schools include providing more opportunities for students to respond, social stories, offering choices, and/or cuing packages such as a visual schedule (Machalicek, O'Reilly, Beretvas, Sigafoos, Lancioni, 2012; Massey & Wheeler, 2000).

The second evidence-based strategy is instructional strategies. Instructional strategies teach a child a behavior that meets the same function as the problem behavior. Common instructional strategies include functional communication training such as asking a teacher for help or attention for attention-maintained behaviors (Kennedy, Meyer, Knowles, Shukla., 2000; Schindler & Horner, 2005) or asking for a break for escape-maintained behaviors. Finally, the third intervention strategy is contingency management in which appropriate behavior is reinforced while problem behavior is not (i.e. placed on extinction).

One behavioral strategy that has been widely used in classroom settings is self-monitoring. Self-monitoring is defined as a multi-step process where a participant records whether a behavior occurs and may or may not include features of that behavior (Bruhn, McDaniel, & Kreigh, 2015; Mace et al., 2001). Self-monitoring has been used to improve behaviors in all populations such as ADHD (Graham-Day, Gardner, & Hsin, 2010; Gureasko-Moore, DuPaul, & White, 2007), emotional behavioral disorders (Gulchak, 2008), and developmental disabilities (Kartal & Ozkan, 2015; Plavnick, Ferreri, & Maupin, 2010;

Rosenbloom, Mason, Wills, & Mason, 2016). Self-monitoring has been used in classrooms to both increase academic performance (Joseph & Eveleigh, 2011) and decrease problem behavior.

One study by Stasolla and colleagues (2017) compared self-monitoring and differential reinforcement of alternative behavior (DRA) to promote on-task behavior for three children with cerebral palsy and developmental disabilities in a classroom setting using an alternating treatment embedded in a non-concurrent multiple baseline design. This study incorporated the use of an application called Clicker 5. Clicker 5 is typically used as a communication application for individuals with mild to moderate physical challenges. In the DRA condition, participants were provided with a token for each interval in which they were on task. In the self-monitoring condition, participants were prompted every 20s on average to record his/her own behavior using Clicker 5 which utilized a pressure sensor to collect data. Tokens were provided if participants scoring of on/off-task coincided with teacher-rated intervals. Results indicate that while both strategies were effective in increasing on-task behavior, self-monitoring was preferred by all three participants.

A limitation of this study is a token being provided for self-monitoring does not allow for differentiation of results to be attributed to the token or the self-monitoring intervention. Another limitation is that preference allocated towards self-monitoring may be due to the use of assistive technology rather than actual preference. Future research should attempt to separate the effects of receiving a token and self-monitoring.

In a study by Bruhn, Woods-Groves, Fernando, Choi, and Troughton (2017), a self-monitoring application called SCORE IT was used by typically developing middle schoolers to self-monitor three behaviors in a multiple-baseline across settings design (MBL). These behaviors included practice responsibility, respect and safety, and do your best. Students were

prompted by the app after a 10-min interval to rate their own behaviors on a scale from 0 to 4 and three to four intervals were completed throughout a class period. Results were mixed with self-monitoring being successful for one student, variable across class subjects for the second student, and with one student remaining highly variable, although it was noted that that student's attendance was inconsistent due to illnesses.

Limitations of this study include (a) only two tiers in each MBL, (b) no evaluation of academic behaviors, (c) no collection of social validity data, (d) limited diversity among the participants (all were typically developing students), and (e) no demonstration of a functional relationship for two participants.

One benefit of self-monitoring is the ease of implementation for behavioral or educational staff. As the student is responsible for collecting his/her own data, this means fewer demands are placed on others, which may be a more cost-effective use of available time and resources (Moore, Anderson, Glassenbury, Lang, & Didden, 2013). Teachers do not have to use valuable class time to record data and can stay focused on instruction. Students may also benefit from a self-monitoring strategy in that behavior is being recorded immediately across specific academic periods so data may be more accurate and can lead to better intervention decisions (Ganz, 2008). In addition, students may have more buy-in with a system that allows them to actively participate and may serve as an antecedent manipulation to stay on-task to mark their data sheets.

Although self-monitoring has been shown to be effective in increasing on-task behavior, weaker effects have been seen for decreasing problem behaviors such as disruptive behaviors and off-task behavior (Davis, Mason, Davis, Mason, & Crutchfield, 2016). In addition, there are fewer studies that evaluate use of self-monitoring to reduce problem behavior than studies

aiming to increase appropriate behavior. One study that did evaluate the effects of self-monitoring on behavior reduction was Crutchfield, Mason, Chambers, Wills, and Mason (2014). This experiment evaluated the use of I-Connect, a self-monitoring program, to decrease stereotypy in two children with ASD. In an ABAB reversal with an embedded multiple baseline design, two participants were prompted to answer the self-monitoring question "Quiet hands and mouth?" every 30s. Participants were prompted if they had not answered the question within 3 s. Baseline rates of stereotypy were at least 60% of all intervals with an increasing trend in baseline. Although intervention led to an overall decrease in stereotypy, visual analysis did not reveal such trends and a statistical analysis had to be conducted due to the high variability in the data.

Another example of self-monitoring not being effective for problem behavior is a study conducted by Crawley, Lynch, and Vannest in 2006. A 6-year-old with mild intellectual disabilities was asked to self-monitor off-task behavior in a community-based instruction classroom. A reversal design indicated self-monitoring may have been effective in decreasing the duration of off-task behavior; however, both baseline phases had significantly decreasing trends immediately before treatment began so a clear functional relationship was not demonstrated.

A second major limitation of self-monitoring is the frequency required to take data. Often students are prompted multiple times within 1 min to record whether or not they were on task (Crawley et al., 2006; Stasolla et al., 2017). Students who have difficulty focusing on academic work may have further difficulties with academic achievement when they are continuously interrupted to score a data sheet. Although research has shown that self-monitoring has increased on-task behavior, no known study has utilized measures to see if academic

achievement increased as well (Stasolla et al., 2017) In other words, students may need specific, additional instruction to increase achievement in combination with self-monitoring.

A final limitation is a lack of social validity data from both students and teachers. Although some studies assessed the acceptability and likeability of the intervention from one party or the other, no known study has evaluated this for both students and teachers (Bruhn et al., 2017; Crawley et al., 2006; Davis et al., 2016). It is important that both implementers and participants indicate they enjoy using the intervention and think it is effective and efficient. These factors are key indicators of whether the intervention will be used once the researcher is no longer present (Kennedy, 2002).

Data Collection Methods

Self-monitoring interventions involve the use of a data recording system such as an interval recording system (Crutchfield et al., 2014), a checklist (Diegelmann & Test, 2018; Miller & Taber-Doughty, 2014), or a behavior rating scale (Bruhn et al., 2017; Smith, Young, West, Morgan, & Rhode, 1988). Although these measurement strategies may be efficient, there is a lack of research identifying whether data collected in this manner closely aligns with direct observation data taken by skilled data collectors.

Direct observation and repeated measures are considered the gold standard for measuring performance in the field of applied behavior analysis as data collection takes place at the time of the actual occurrence of the behavior (Chafouleas, Riley-Tillman, & McDougal, 2002). This allows for the most accurate measure of a behavior, apart from automatic data collection via a computer. Although direct observation may provide the benefit of objectivity, there are several limitations. Direct observation can often be time consuming, require multiple observers to

ensure the reliability of the data, and can require highly trained observers if the data collection system is complex (Chafouleas, McDougal, Riley-Tillman, Panahon, & Hilt, 2005). These limitations pose a direct challenge for use in the classroom environment, as there are several students and therefore complex behavior interactions occurring at any one time.

Direct Behavior Ratings

Teacher evaluations of student's behavior are the most widely used method to assess classroom conduct problems; however, a complex and time-consuming data collection system often hinders teachers from collecting data appropriately (Chafouleas et al., 2005; Hoge & Andrews, 1992). To try to balance developing an efficient and feasible system while still collecting accurate data, direct behavior ratings (DBRs) were developed (Steege, Davin, & Hathaway, 2001; Chafouleas et al., 2002; Chafouleas, Christ, Riley-Tillman, Briesch, & Chanese, 2007). Direct behavior ratings require at least a daily rating of a specifically defined behavior and that information is then shared across individuals to assist with data-based decision making (Chafouleas et al., 2007).

In one study, Chafouleas and colleagues (2005) compared daily behavior report card ratings to direct observation of off task behavior in a classroom. Teachers used a 5-point, Likert-type scale to rate off task and a specific identified problem behavior for 1 student in his/her classroom. Research assistants also collected data on the same student using a 20-s momentary time sampling recording procedure. The results of the momentary time sample were converted to the daily behavior report card scale and the two scales (direct observation and teacher-rated DBR) were compared. Results indicated that 82-87% of the cases had the same scale rating or only a 1-point difference in agreement. This indicated a moderate correlation between teacher-

rated DBR and direct observation data and accounted for 23-42% of the variance in the data. The agreement of DBRs and systematic direct observation has been evaluated in the educational system many times and results demonstrate moderate correlation between the two (Barnes, 2015; Chafouleas et al., 2005; Chafouleas et al., 2007; Riley-Tillman et al., 2008).

DBRs have several benefits over direct observation. DBRs are simple to use and are not time-consuming to collect data immediately. DBRs are also flexible to the setting and behavior being assessed and can vary according to the behavior to be rated, the rating frequency, the rater, and the type of rating used (Chafouleas et al., 2002). In addition, DBRs can be used as a treatment component or to monitor the progress of interventions already in place (Riley-Tillman, Chafouleas, Sassu, Chanese, & Glazer, 2008). Based on the benefits of DBRs, researchers have continued to try to develop a standardized measure that is user friendly and sensitive enough to capture changes in their targeted dependent variables. This led to the development of the Individualized Behavior Rating Scale (IBRST).

Individualized Behavior Rating Scale Tool (IBRST)

The IBRST was originally developed as a feasible method for teachers to rate a student's behavior daily in the Prevent-Teach-Reinforce (PTR) model of functional behavior assessment (Dunlap, Iovannone, Wilson, Kincaid, & Strain, 2010). PTR is a systematic, collaborative, multi-step FBA/BIP approach that has been evaluated in two randomized controlled trials, one with students in grades kindergarten through eighth grade (Iovannone et al., 2009) and the other with young children in pre-kindergarten settings (Fronapfel, Dunlap, Flagtvedt, Strain, & Lee, 2018). The IBRST uses a 5-point Likert-type scale that is used to progress monitor specific behaviors daily. The IBRST can monitor both problem and appropriate behaviors (Iovannone, Greenbaum,

Wang, Dunlap, & Kincaid, 2014). As an individual scores the behavior across days, the circled numbers can be connected to generate a line graph for that behavior. This provides a measurement of that behavior across time.

Although similar to a DBR, there are a few differences. First, instead of generic behavior categories such as "on-task" or "disruption," targeted behaviors are selected and operationally defined for each student. In addition, behaviors are rated by circling a rating on the 5-point Likert scale instead of a DBR-type rating of a 100mm line divided into 10, 10 mm intervals and a teacher marking exactly which percentage a child engaged in that behavior (Chafouleas et al., 2002; Riley-Tillman et al., 2008). Third, the anchored ratings of the IBRST are set individually for each student based on estimates of the target behavior occurring. Finally, the daily ratings of the IBRST immediately generate a line graph, which can demonstrate behavior change across time.

Generally, the IBRST is developed in collaboration with a team who knows the child well. For each target behavior, the developer will ask the teacher a series of questions to individualize the IBRST for that child (Dunlap et al., 2010). Once the target behaviors are selected and defined, the time period for monitoring the behavior(s) with the IBRST are determined. Teachers may choose to rate behavior across specific routines or activities, specific time periods, certain class periods, or over an entire school day. The team comes to consensus on an appropriate measurement approach that would matches the target behavior (e.g., frequency, duration, and intensity). To assist with selecting the best measurement type, teachers may be asked a question such as: "Are you most concerned with how often the behavior occurs, how long the behavior lasts, or what percentage of the day/routine the behavior occurs?" Once the measurement approach is selected, the anchors are set in the following order for a problem

behavior: *4* (bad day); *5* (very bad day); *1* (very good day); *2* (good day); *3* (ok day). The anchors for appropriate behavior are reversed such that: *4* (good day); *5* (very good day); *1* (very bad day); *2* (bad day); *3* (ok day). The process of setting the anchors is repeated until they are set for all behaviors targeted for intervention (Iovannone et al, 2014).

Iovannone et al. (2014) was the first published study evaluating the inter-rater agreement of the IBRST. In this study the authors worked with a subset of 19 teacher/student dyad participants from the original PTR randomized controlled trial who agreed to participate in the inter-rater agreement study. The IBRSTs were developed as described above. Teachers and an independent observer each independently scored the IBRST during the specified routine/time period. Kappa coefficients were calculated, and the results showed strong reliability (K_{aw} = .80 and K_{lw} = .62) between the teachers' and independent observer scores. Further analysis suggested that more discrete behaviors might be more accurately measured using the IBRST versus anchors relying on percentage scores and that sometimes, inadequate agreement was found for behaviors that were less salient. Nonetheless, these preliminary results suggest the IBRST is a reliable repeated measure data collection tool for measuring student performance. Limitations of this study include: small sample size, no concurrent validity measured, a variation in the time lengths of observations, and a different number of observations for each student.

An unpublished thesis conducted by Martinez (2016) found similar promising reliability with students recording their own behavior. In this study, three typically-developing, 2nd grade students used the IBRST to self-monitor disruptive behavior and academic engagement during 30-min observation periods. Before baseline began, teachers set anchors with the assistance of the researcher in the manner described previously. During baseline, only teachers recorded IBRST data. Students were then trained on what their individual anchor scores were and were

given an opportunity to practice scoring themselves with feedback from the researcher. Then, they were asked to self-monitor their behavior two times in a 30-min observation period, 3 to 5 times per week. Results indicated that intervention resulted in lower frequencies of problem behavior and greater on-task behavior than in baseline for all three participants; however, only two of the participant's behaviors maintained at follow-up. In addition, both teachers and students rated the intervention as highly effective and with high likeability. Limitations of this study include small sample size, a short time before follow-up was collected, and no concurrent validity data.

Only one known study has evaluated the concurrent validity of direct observation data and the IBRST. In an unpublished thesis conducted by Barnes (2015), the Prevent-Teach-Reinforce (PTR) model was evaluated for three children in an elementary school using a multiple baseline across participants design. In addition, the validity of the IBRST was assessed as a daily monitoring tool. Specifically, the study aimed to evaluate whether the ratings recorded by the teacher on the IBRST agreed with systematic direct observations. Results indicated that the teacher-rated IBRSTs had strong correlations with direct observation data. Limitations of this study include: only one follow-up probe being conducted, generalization was not assessed, and severe time constraints imposed by the school district in which the study was conducted.

Future Research

Future research evaluating the use of self-monitoring should attempt to address some of these aforementioned limitations in the education setting. More research should be conducted using single-case designs, specifically addressing the paucity of self-monitoring interventions to decrease problem behavior in a classroom. In addition, research should also evaluate how often

children may be required to monitor their own behavior to have an effect on behavior. For example, would a child who scored him or herself every 15 s produce the same effect if he/she scored him/herself every 15 min?

To address issues specific to the IBRST, future research should try to maintain similar observation lengths to minimize comparing the reliability of data taken across a day with data taken across a class period as this may affect results. Future research should also evaluate the concurrent validity of self-monitoring using the IBRST to determine if data taken with the IBRST correlates with direct observation data or the "gold standard" of data collection. If a child can score themselves accurately yet intermittently, this may be a less disruptive strategy that can have an important impact on academic performance in the school setting. Finally, although the intervention may be effective and might not require a lot of response effort on behalf of the teacher or student, future research should ensure social validity data is collected from both parties.

Therefore, this study attempted to address three research questions. The first experiment addressed: a) Is there concurrent validity between student self-monitoring data collected using the IBRST and direct observation data? and b) To what extent will student self-monitoring using the IBRST result in reduced problem behavior and increased appropriate behavior in a classroom setting? The second experiment addressed: a) To what extent will students be able to independently set their own anchors and use the IBRST to self-monitor their own behavior and will there be concurrent validity with direct observation data?

CHAPTER TWO: EXPERIMENT 1: METHODS

The purpose of Experiment 1 was to determine the effects of self-monitoring using the Individualized Behavior Rating Scale Tool (IBRST) on problem behavior and on-task behavior of individuals in a small private school. In addition, the concurrent validity of self-recorded IBRST data was evaluated.

Recruitment

Four students were recruited to participate in this study. One student in the pre-baseline assessment did not meet criteria of having at least two problem behaviors occurring during the observation period and was excluded from further participation in the study. Students were nominated to participate by the doctoral level behavior analyst who led the school's ABA program based on: a) at least two non-harmful problem behaviors (such as disruption) occurring at least two times per day for 4 of 5 school days per week, b) at least two academic/behavioral goals, c) aged 7-17, d) verbal assent from participant, d) written informed consent from parents, e) persistent attendance, and f) basic reading skills. Persistent attendance was defined as being in school at least 4 out of 5 days on average for the preceding month at school. Basic reading skills were defined as being on at least a 3rd grade reading level as evaluated by the school's ACE Curriculum. Students were excluded from the study if he/she engaged in serious, harmful problem behavior (e.g. aggression towards others that left lacerations or broken bones, self-

injurious behavior that left a mark lasting longer than 5 min). There were no other exclusion criteria.

For students who met the inclusion criteria, informed consent documents were sent home to his/her parent(s)/guardian(s). On the informed consent document, there was a checkbox to indicate if the parent(s)/guardian(s) would like to be contacted regarding the study as well as a section indicating preferred method of contact. If the informed consent document was returned indicating the family was interested in learning more about the study, the researcher then used the mode indicated on the contact list to schedule a time to review the informed consent document with the parent. Once the form was reviewed with parents, they were given the opportunity to ask questions and either sign the document or be given up to one week to consider whether they would give permission for their child to participate. Parents were told that their child's education and ABA therapy (if the child was receiving this) did not depend upon participation in this study and that they or their child could elect to no longer participate at any point without penalty.

Once the parent signed the consent form, the researcher obtained signed written consent and verbal assent from the child to participate in the research. If a participant indicated he/she did not want to be included in the study, further contact about the study was discontinued and the child was not included in the study. If a child chose to participate, 30-min observations were conducted to identify levels of problem behavior and appropriate behavior. During these observations, the researcher did not interact with anyone within the environment. Participants must have engaged in at least two different problem behaviors at least one time each across two observation periods to be included in the study.

Participants and Setting

Three students, aged 11-15, participated in a private school in a suburban area in West Florida. This school had a population of approximately 80 students, 35% of whom currently required individualized applied behavior analysis (ABA) services and 35% of all students had a DSM-5 diagnosis. All participants names were changed to protect their identities. Each classroom had one teacher and an average of 12 students per class. The school implemented the Accelerated Christian Education (ACE) curriculum, which expects students to work independently at their own pace in workbooks for six core subjects: math, English, word building, science, social studies and Bible. There was no teacher-directed subject at this school, rather, students worked for 40 min periods in each subject's workbook at their own pace in an order assigned by the teacher at the beginning of the school day. As students completed pages, they were prompted in the workbook to self-score their own work and correct errors and re-score as needed. The teacher walked around the room and answered questions students had about the workbooks and gave permission for students to score; however, at no time did the teacher instruct the entire class in any subject.

Danny was an 11-year-old male diagnosed with attention-deficit hyperactivity disorder (ADHD). He was in 5th grade but currently worked on 3-4th grade materials. He had received ABA services for approximately six months prior to the beginning of the study. He was taking Strattera medication for ADHD. His problem behaviors of concern included talking back and disruption.

Deborah was a 13-year-old female diagnosed with ADHD, oppositional defiant disorder (ODD), and bipolar disorder. She was in 8th grade but worked on 5th grade materials. She had received ABA services for approximately one year before the study began. She was currently

taking sertraline for ADHD symptoms and oxcarbazepine for ADHD and bipolar symptoms. Her problem behaviors of concern included verbal outbursts and property destruction.

Bailey was a 15-year-old female diagnosed with ADHD, mood disorder, anxiety, depression, and mild intellectual disability. She was in 10th grade but worked on 4th grade materials. She had received ABA services for approximately one year before the study began. She was taking desmopressin for bedwetting symptoms, metformin for polycystic ovary syndrome, aripiprazole for mood disorder symptoms, sertraline for depression/anxiety, buspirone for anxiety, and focalin XR for ADHD symptoms. Her target behavior of concern was non-compliance.

Measurement

The two dependent variables in this study were problem behavior and on-task behavior. For data analysis purposes, Cohen's kappa coefficient was used to examine the correlation between student-scored Individualized Behavior Rating Scale Tool (IBRST) data and researcher-collected direct observation data. Data were collected via both direct observation of problem behaviors and being on-task as well as via the Individualized Behavior Rating Scale Tool (IBRST) for each targeted behavior.

Direct observation data were collected 3-5 times per week per participant in 30-min observation periods. Direct observation data were collected during all phases of this experiment. IBRST data were only collected during treatment and follow-up conditions of this experiment. All data were collected using paper and pencil (see Appendices A and B).

Research Assistants

The three research assistants (RA) for this study were the child's typical registered behavior technician. Two of the behavior technicians had taken all master's courses required to sit for the BCBA exam and one had taken all bachelor's level courses to sit for the BCaBA exam. All RAs had extensive experience with data collection, particularly for their respective clients. The RA was trained to collect data using behavioral skills training methods including providing instructions, modeling how to mark the data sheet appropriately, role-play of an observation session, and then feedback. The RA had to reach 90% IOA with the researcher to begin scoring participant data. All RAs met correspondence level with one training session. No additional training was required for the duration of the study.

Direct Observation of Problem Behavior

To collect direct observation data, the existing definition in each participant's problem behavior in his/her individualized behavior plan was used as described below.

Danny's problem behavior targeted in this experiment was disruption. Disruption was defined as any instance of Danny talking to or making faces at other students or making comments to the entire classroom such as "This room sucks – I want to go home" above normal conversational level and/or during silent work time. Deborah's targeted problem behavior was verbal outbursts. Verbal outbursts were defined as any vocalization above normal conversational level for that setting including screaming, cursing, calling names, or making threats. Bailey's problem behavior of concern was non-compliance. Non-compliance was defined as engaging in any other behavior than the instruction provided within 10 s of the instruction and/or not completing routine tasks that had been previously described. During the initial 30-min

observation session, Bailey engaged in non-compliance and verbal outbursts. However, once data collection began, only non-compliance was seen; therefore, for Bailey only, there was only one targeted problem behavior.

Problem behavior was measured using frequency count for all participants due to the topography of the behaviors selected. Therefore, each instance of the problem behavior was recorded and scored. The direct observation period was timed using a timer to ensure 30-min of observation occurred. The timer was never stopped for any portion of any observation session as students were expected to remain in their seats and work silently during the observation periods selected.

Direct Observation of Appropriate Behavior

For this study, the appropriate behavior selected for all students was being on-task. On-task was defined as the participant being seated at his or her desk looking at and/or using a writing tool to make appropriate marks in the PACE workbook or walking over to the grading table and self-grading his/her materials by using a red pen and an answer book to mark answers correct or incorrect. If there was a time during the observation period that the child did not have the opportunity to be on-task, such as going to the bathroom or waiting to ask the teacher a question, the observation period timer was stopped until the opportunity for on-task behavior was provided again. The observation period timer was also stopped during the time in which a student was recording data using the IBRST (2 min for each scoring opportunity). The total observation period was about 34 min with 30 min of direct observation.

On-task behavior was measured using duration recording. A stopwatch was used to record the on-set and off-set of each instance of on-task behavior. The timer was started as soon

as the participant was given the instruction to begin working by the teacher. The time was recorded when the participant first met the criteria for being on-task as described above. The time was also recorded as soon as the operational definition for on-task behavior was no longer met. The on-set and off-set of each occurrence of on-task behavior was marked on a data sheet (See Appendix A). The number of seconds of each instance of on-task was calculated and a total duration was obtained by adding up the number of seconds of all instances of on-task behavior. This could then be used to calculate either the total duration of minutes on-task or the percentage of the observation window during which on-task behavior occurred.

IBRST

Students collected data on their own behavior using the IBRST for both problem behavior and on-task behavior (see Appendix B). The IBRST used a 5-point rating scale for both targeted behaviors. For the IBRST, the highest number on the rating scale represented the most behavior occurring, either problem behavior or appropriate behavior. Therefore, for on-task behavior, the scale was rated from 1 to 5, with 5 representing a very good day and a 1 representing a very bad day. Problem behavior was also rated from a 1 to 5; however, a 5 represented a very bad day and a 1 represented a very good day. The IBRST was scored by circling the number on the rating scale which corresponded to how the child perceived he/she behaved during that time period.

Each criterion level was developed by the student and the researcher together as described in the "Student Selection of Anchors" section below. The students were asked to rate themselves using the IBRST two times each observation period for each behavior – the first time at 15 min into the observation period and the second time at the end of the observation period.

The students were given 2 min to score at each 15 min period. Therefore, there were four total circles (two each for problem behavior and on-task behavior) for each observation period.

Concurrent Validity of IBRST

Student IBRST ratings were compared to direct observation to evaluate concurrent validity. All direct observation data were converted to align with the anchors selected by the students for the IBRST. For example, if a child engaged in talking back 12 times according to direct observation data and the anchors for the IBRST were selected as follows *1* (0-1 times), *2* (2-5 times), *3* (6-9 times), *4* (10-13 times), *5* (14-17 times), the direct observation data was converted to a *4* on the IBRST. Cohen's Kappa co-efficient was calculated to examine agreements between the direct observation data collected by the research team and the IBRST score selected by the students. This determined the concurrent validity of the IBRST ratings.

Treatment Integrity

Treatment integrity was assessed for 38% of all intervention sessions. This measure focused on the researcher's implementation of IBRST procedures as planned. A 13-step fidelity checklist was used to assess treatment integrity (see Appendix C). As the intervention was implemented, the research assistant marked "yes" if the step was implemented correctly or "no" if it was not. Some of the steps evaluated included: a) Giving the IBRST data sheet to the student before the observation period, b) setting a 15 min timer, c) stating the timer has started, d) instructing the student to complete the IBRST, and e) praise for marking the IBRST at the end of the timer. Treatment integrity was calculated by taking the total number of steps marked "yes" and dividing by the total number of steps, then taking the quotient and multiplying by

100%. For example, if 7/8 steps were marked "yes," then the quotient would be .875 x 100% = 88%. Treatment integrity for intervention implementation was 100% across all scored sessions.

Treatment integrity data were also collected on the training of students to use the IBRST. A 9-step checklist was used to assess treatment integrity (see Appendix J) for 100% of training sessions. It was calculated in the same manner as described above. Treatment integrity for training students to use the IBRST was 100% across all scored sessions.

Social Validity

Social validity was collected after the follow-up phase to assess the perceived effectiveness and likeability of the intervention. Social validity was gathered from two sources a) teachers and b) students. The teacher's social validity was assessed using an 8-question, 5-point Likert-type scale. The teacher was asked questions such as: rate how effective you think the intervention to be and rate how easy you think the intervention is to implement. The student's social validity assessment was also assessed using a 5-question, 5-point Likert-type scale. An example of each questionnaire can be seen in Appendices D and E.

Inter-Observer Agreement (IOA)

Two independent observers (primary data collector being the researcher and secondary collector being the RA) collected data simultaneously to assess IOA. IOA was calculated for all targeted behaviors for 58% of all baseline sessions and 46% of all intervention sessions. IOA was calculated for direct observation data only. IOA for problem behavior was calculated using the total-count IOA method in which the smaller number was divided by the larger number and then multiplied by 100%. For example, if one observer recorded 9 instances of problem

behavior and the other observer recorded 7 instances, IOA was (7/9) x 100% or 78%. IOA for on-task behavior was calculated using total agreement method in which the smaller percentage was divided by the larger percentage and then multiplied by 100. For example, if one observer marked 80% and the other observer marked 85%, 80/85 = .94 x 100% is 94% (see Appendix F). For Danny in baseline, IOA for problem behavior was 100% and IOA for on-task behavior was 99.5% (range 99-100%). For Danny in treatment, IOA was 100% for problem behavior and 100% for on-task behavior. For Bailey in baseline, IOA was 97.8% for problem behavior (range 89-100%) and 99% for on-task behavior (range 98-100%). For Bailey in treatment, IOA was 100% for problem behavior and 99.5% for treatment (range 99-100%). For Deborah in baseline, IOA was 100% for problem behavior and 95% for on-task behavior (range 86-100%). For Deborah in treatment, IOA was 100% for problem behavior and 99.5% in treatment (range 99-100%).

IOA was also calculated for implementation fidelity. IOA for implementation fidelity was measured for 38% of all intervention sessions. IOA for implementation fidelity was calculated in the same manner as on-task behavior. Two observers observed the researcher engage in the treatment steps. They individually scored fidelity using Appendix C. IOA of implementation fidelity was calculated by dividing the smaller percentage by the larger percentage and then multiplying by 100%. IOA for implementation fidelity was 100% across all scored sessions.

Experimental Design

This study used a multiple-baseline design across participants. The phases of this experiment were baseline, self-monitoring with the IBRST, and follow-up.

Procedure

Baseline

During baseline, teachers were instructed to conduct their classrooms as usual. Behavioral staff were instructed to implement procedures as usual. No self-monitoring with the IBRST was conducted during this phase; however, direct observation data were collected using the methods described above for both problem behavior and on-task behavior. Data collected during baseline were used to help identify anchors for the IBRST during the next phase of the experiment. For all students, a goal was set of at least a 25% reduction in problem behavior and a 25% increase in on-task behavior from baseline as measured by direct observation data. Baseline was considered complete once data were stable or were trending contra-therapeutically for at least three data points in a row.

Student Selection of Anchors

The students participating in the experiment worked in collaboration with the researcher to select appropriate anchors for the intervention. Unlike Iovannone et al. (2014) or Martinez et al. (2016), the IBRST anchors were made in collaboration with the student instead of the teachers and the measurement system and routine was selected for the participant. The participants were asked to estimate behavior occurrences to set the scale points for each target behavior. In a further deviation from Iovannone et al. (2014), the IBRST anchors for the problem behavior were set in the following order: *5* (very difficult day), *4* (difficult day), *1* (very easy day), *2* (easy day), *3* (ok day).

The student was provided with their baseline data to help guide them through the selection process, including the highest and lowest frequency of problem behavior and the average frequency of problem behavior. As the participant baseline data were collected across a 30-min observation session and the students were rating themselves every 15 min, the participants were guided to divide their highest frequency of problem behavior in baseline by two. They then would select the highest quotient as anchor 5. For example, if a participant had 20 instances of problem behavior in baseline, this score was divided by 2 for a total of 10. Ten was therefore set as anchor 5 for problem behavior. Then the participant was told to divide their highest baseline score for the 15 min by the 5 anchors. For example, 10/5 = 2. The participant would then create a range from that score. Using the example above, anchor 5 would be a range from 9-10. Then anchor 4 was set by subtracting 2 from the highest point of the anchor above it and creating a range as previously described. For example, 10-2 = 8 so anchor 4 would be 7-8. The rest of the anchors were set in the order described above so that all anchors had an even distribution.

If a student selected an anchor that did not correlate well with baseline data (i.e. average frequency was 12 but student selects a very difficult day occurring 1-4 times), the student was guided through selecting an anchor that was based on the data being shown. The student was asked leading questions to come to an agreement on anchor points. First, the student was prompted with a broad question such as "What was the most difficult day you had according to the data?" If a broad question did not lead the student to select an appropriate anchor, the student was prompted with a detailed question such as "What is the range of numbers that would be less than what you put for anchor 5? If agreement was still not reached or there was a lack of response, the student was provided with the appropriate anchor score. Students were then given

the opportunity to set the next anchor individually. The participants and the researcher continued working back and forth with each other until agreement was reached over the anchors. Appropriate behavior scale points were set as well, although the anchors were reversed as described previously. Two of the students, Deborah and Bailey, required all anchors to be provided to them. Danny was able to set anchor 5 individually, needed anchor 4 set for him, but then was able to set the rest of the anchors independently.

For Experiment 1, Danny's anchors for disruption were set as follows: *5* (5 + instances), *4* (4 instances), 3 (3 instances), *2* (2 instances), *1* (0-1 instance). Danny's anchors for on-task behavior were set as follows: *5* (13-15 min), *4* (10-12 min), 3 (7-9 min), *2* (4-6 min), *1* (less than 4 min). Deborah's anchors for verbal outbursts were set as follows: *5* (8-10 instances), *4* (5-7 instances), 3 (3-5 instances), *2* (1-2 instances), *1* (no instances). Deborah's anchors for on-task behavior were set as follows: *5* (13-15 min), *4* (10-12 min), 3 (7-9 min), *2* (4-6 min), *1* (less than 4 min). Bailey's anchors for non-compliance were set as follows: *5* (8-10 instances), *4* (5-7 instances), 3 (3-5 instances), *2* (1-2 instances), *1* (no instances). Deborah's anchors for on-task behavior were set as follows: *5* (13-15 min), *4* (10-12 min), 3 (7-9 min), *2* (4-6 min), *1* (less than 4 min).

Student Training to Use the IBRST

Students were trained individually to use the IBRST by the researcher. The steps for training included: a) showing the student how to circle the correct anchor on the IBRST according to the behavior that he/she engaged in during the observation period, b) practicing the skill with the child by asking the student to score the researcher engaging in the target behaviors, c) providing feedback for the correct use of the IBRST during the role-play scenarios. The

IBRSTs that were used during training were collected from the student immediately following the training. Training continued until students reach 100% correct use of the IBRST. This only took one training session for all three students. Once 100% fidelity was reached, the self-monitoring intervention began.

Self-Monitoring with the IBRST

This phase of the study was conducted exactly the same as baseline (i.e. teacher conducted class as normal and other behavioral procedures were implemented as normal) except for the self-monitoring component of scoring the IBRST. During the 30-min observation period, students were asked to record their own behavior using the IBRST two times, 15 min into the observation period and at the end of the observation period.

At the beginning of the observation period, the researcher passed out a clean copy of the IBRST to the participant. The participants were given 2 min to review the definitions of the behaviors being targeted as well as the anchors. Then, the participant was asked if he or she had any questions. If the child answered "no" or after the question(s) was (were) answered, the first 15 min timer was set along with the prompt "Your timer has started. You will be monitoring your own behavior now." After 15 min elapsed, the timer sounded, and the participant was prompted to score his/her IBRST for both on-task and the child-specific problem behavior. The child was given 2 min to score his/her two targeted behaviors. Behavior-specific praise was provided for completing the IBRST, but not for accuracy of the recording, in the form of "Excellent work completing the IBRST."

Following scoring, the participants were prompted to put the IBRST sheet to the side of his/her desk. Then, the second 15 min timer was set along with the same prompt that was

delivered before. After the timer went off a second time, marking the end of 2nd interval, participants were prompted to complete the IBRST as before. The students were again given 2 min to mark his/her data sheet. Praise was provided as described previously. After the student scored his/her IBRST, it was collected from the student. The IBRST phase was complete after at least three consecutive sessions with stable data (i.e., no large change in level or trend). Following the intervention condition, the IBRST plus reinforcement or the follow-up condition was implemented.

CHAPTER THREE: EXPERIMENT 1: RESULTS

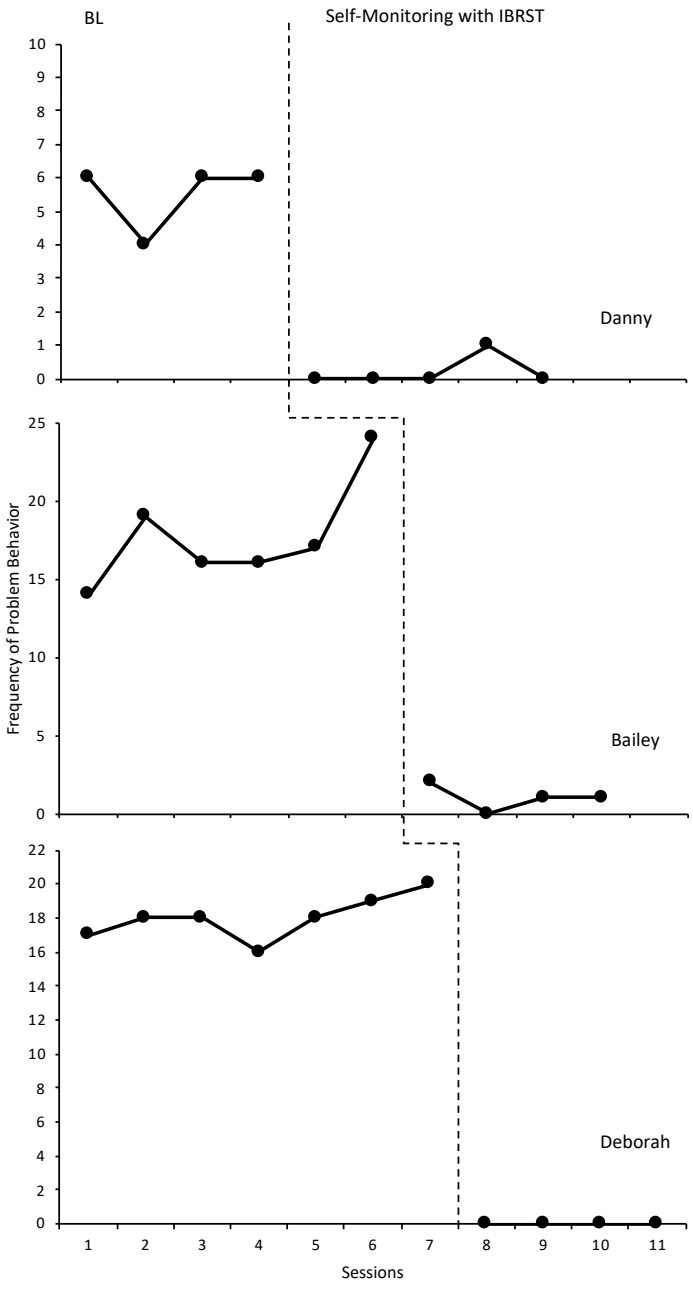

Figure 1. Results of self-monitoring using the IBRST on problem behavior.

As indicated in Figure 1, there was an immediate decrease in problem behavior when and only when the intervention was implemented. The average frequency of disruption for the last three sessions of baseline for Danny was 5.33 (range 4-6 instances). In treatment, this decreased to an average of 0.33 (range 0-1 instance) for the last three sessions. Bailey's average frequency of non-compliance for the last three sessions in baseline was 19 (range 16-24 instances). After implementing the self-monitoring intervention, non-compliance decreased to .67 instances (range 0-1 instance) for the last three sessions of intervention. Finally, for Deborah, verbal outbursts occurred on average 19.5 instances (range 18-20) for the last three sessions of baseline. Following intervention, verbal outbursts decreased to 0 instances.

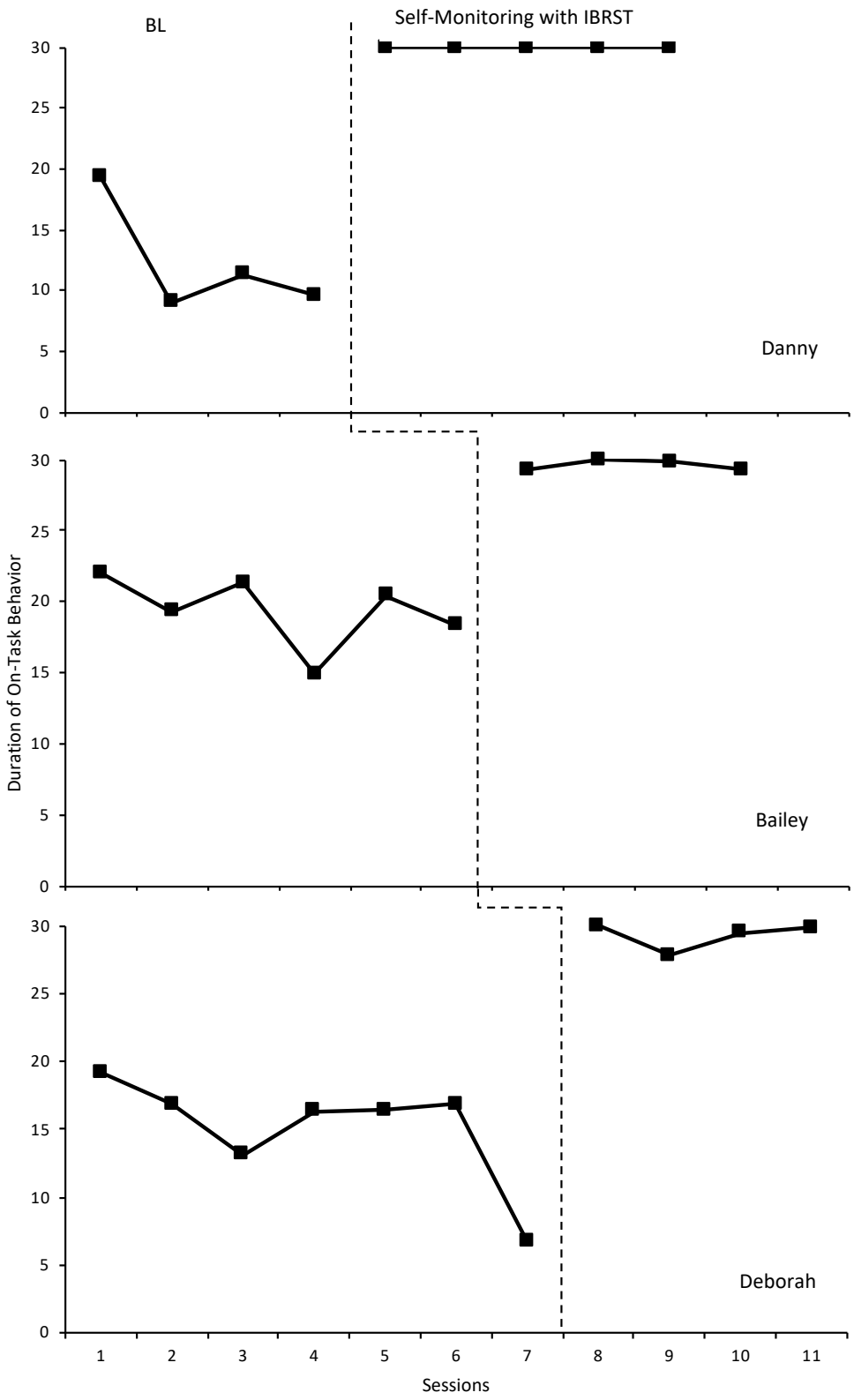

Figure 2. Results of self-monitoring using the IBRST on on-task behavior (SDO data).

As indicated in Figure 2, on-task behavior increased after sequential introduction of the intervention. The average number of minutes of on-task behavior for the last three sessions of baseline for Danny was 10 min (range 9-11.3 min). In treatment, this increased to an average of 30 min for the last three sessions. Bailey's average number of minutes of on-task for the last three sessions in baseline was 17.9 min (range 14.9-20.4 min). After implementing the self-monitoring intervention, on-task increased to an average of 29.7 min (range 29.3-30 min) for the last three sessions of intervention. Deborah was on-task on average 13.3 min (range 6.7-16.8 min) for the last three sessions of baseline. Following intervention, on-task increased to 29 min (range 27.8-29.8 min).

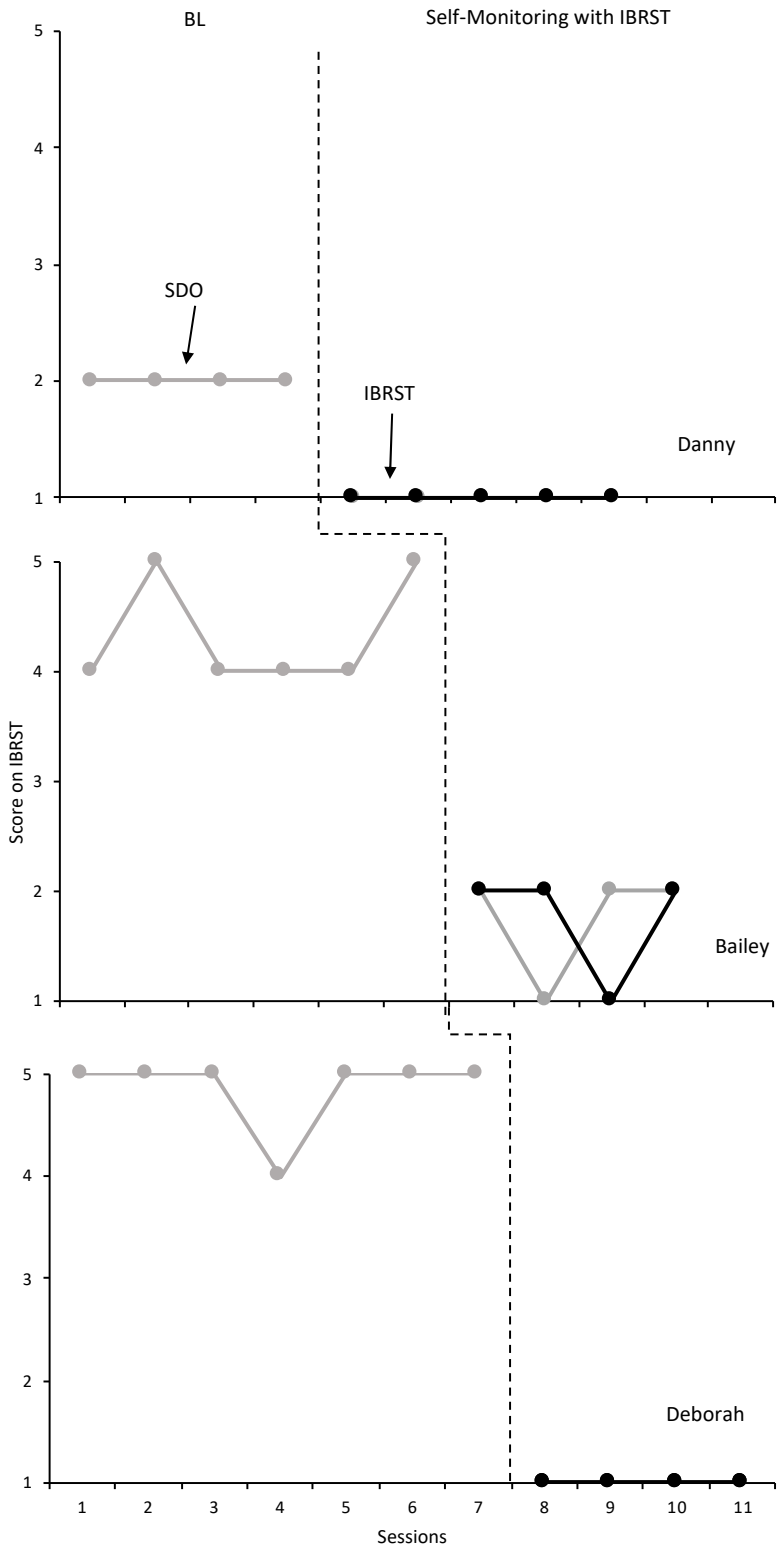

Figure 3. Correlation between converted direct observation scores (SDO) and student-scored IBRST scores for problem behavior.

As indicated in Figure 3, only three data points did not correlate precisely with direct observation scores. The only participant in which there were differences between direct observation data and IBRST scores was Bailey. Therefore, there were perfect agreements 11/13 times or 85% of the time. When differences occurred, it was only by 1 point on the IBRST scale. At no point did scores differ from SDO data for Danny or for Deborah, indicating high levels of agreement between the IBRST and direct observation data. Cohen's Kappa co-efficient was calculated to examine the agreement between SDO data and IBRST scores. The result was k = .56, indicating a moderate correlation. An explanation as to why the k value may have been so low despite a high percentage of agreements will be described in the discussion section.

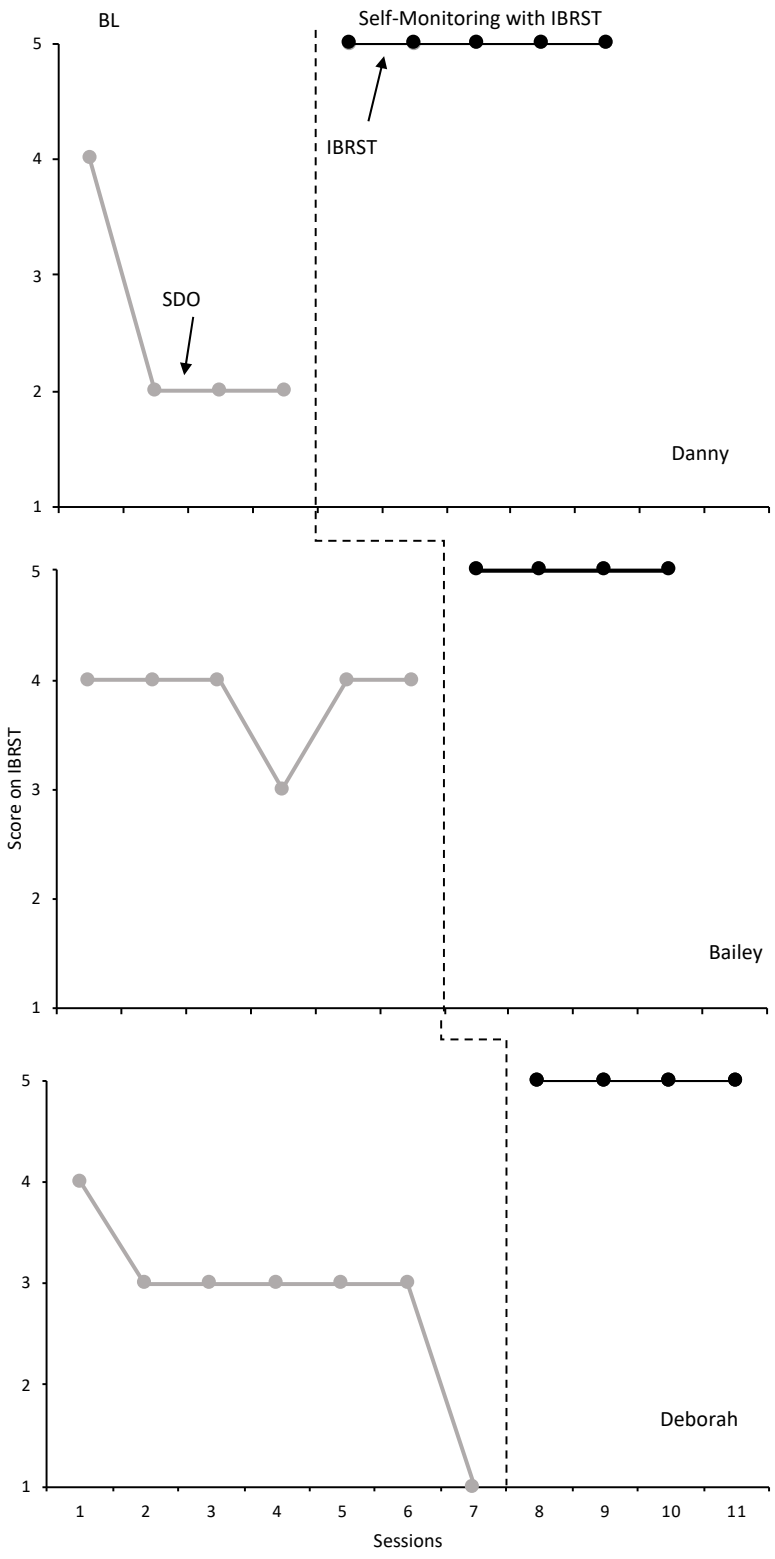

Figure 4. Correlation between converted direct observation scores (SDO) and student-scored IBRST scores for on-task behavior.

As indicated in Figure 4, all data points overlapped between researcher-scored and converted SDO scores and student-scored IBRST data, again indicating strong correlation between SDO data and IBRST scores. Cohen's Kappa co-efficient was calculated to examine the agreement between SDO data and IBRST scores. The result was k = 1, indicating perfect agreement.

Table 1. Student social validity results for Experiment 1.

	The IBRST helps me stay on-task in class.	The IBRST is easy to use	The IBRST helps my behavior in class.	I think the IBRST would help other kids.	I liked using the IBRST.
Danny	4	4	5	5	4
Deborah	5	4	5	4	5
Bailey	5	5	4	5	4

Student social validity results indicated high-likeability and perceived effectiveness of the intervention. The average scores were as follows: 1. *4.7* (range 4-5), 2. *4.3* (range 4-5), 3. *4.7* (range 4-5), 4. *4.7* (range 4-5), 5. *4.3* (range 4-5). Students scored the intervention as easy to use, that they liked using it, that the IBRST helped them stay on task, and that it would help other children stay on-task during class.

Table 2. Teacher social validity results for Experiment 1.

	Self-monitoring with IBRST was acceptable to decrease problem behavior.	Self-monitoring with IBRST was acceptable for increasing on-task behavior.	Self-monitoring was effective at decreasing problem behavior.	Self-monitoring was effective at increasing on-task behavior.	Suggest IBRST to other teachers.	Student's behavior warranted need for treatment.	I would use IBRST during other instructional times.	The student appeared to enjoy self-monitoring using IBRST.
Teacher 1	5	5	4	4	4	5	5	5
Teacher 2	5	5	4	4	4	5	5	5

Teacher social validity results also indicated high-likeability and perceived effectiveness of the intervention. The teachers scored the self-monitoring intervention using the IBRST a 5 as an acceptable option for decreasing problem behavior, a 5 as an acceptable option for increasing on-task behavior, a 5 as an acceptable option for increasing on-task and decreasing problem behavior, a 4 that she would suggest self-monitoring with the IBRST to other teachers, and a 5 that the student appeared to enjoy using the IBRST.

CHAPTER FOUR: EXPERIMENT 2: METHODS

The purpose of Experiment 2 was to determine if participants were able to set anchors and collect data for the IBRST independently and the effects of independent self-monitoring on a 2nd problem behavior of interest. The concurrent validity of the independently-developed IBRSTs was evaluated as well.

Participants and Setting

The participants for this experiment were the same participants from Experiment 1. The only inclusion criteria for this study was the participant must have shown decreases in problem behavior during Experiment 1 to goal level as measured by visual analysis. The setting was the same as in Experiment 1.

Measurement

Direct Observation of Problem Behavior and On-Task Behavior

Direct observation of problem behavior and on-task behavior occurred in the same manner as Experiment 1. A second, different problem behavior was selected for each student. As mentioned previously, only one problem behavior was targeted for Bailey as she did not exhibit any instances of verbal outbursts after the initial observation period.

Danny's problem behavior of concern was talking back. Talking back was defined as verbally questioning/challenging a task or instruction provided by a staff member. Deborah's problem behavior was property destruction. Property destruction was defined as breaking, throwing, or knocking objects over as well as pounding objects with hands. Bailey's problem behavior of concern was non-compliance. Non-compliance was defined the same as in experiment 1 (e.g., engaging in any other behavior than the instruction provided within 10 s of the instruction and/or not completing routine tasks that have been previously described).

For on-task behavior, the difference for this experiment was the routine during which problem behavior and on-task behavior was observed. For example, if on-task was measured during math time in Experiment 1, on-task was measured during science in Experiment 2. The students followed a strict schedule of what books are worked on at what part of the day with Math, English, and Word Building being morning subjects and Science, Social Studies, and Bible being afternoon subjects. If a morning subject was selected for a participant in Experiment 1, an afternoon subject was selected for the participant in Experiment 2. The length of the observation period was the same as Experiment 1 (30 min).

Individualized Behavior Rating Scale Tool (IBRST)

The IBRST was scored in the same manner as Experiment 1. The difference in the IBRST was that the anchors were created by the participants independently after being given the minimum, maximum, and average frequency of his/her problem behavior in baseline. In addition, a timer was provided but no prompts were given to participants to set the timer or score themselves at 15 min and 30 min. Therefore, at the end of each observation period, there was four total circles (two each for problem behavior and on-task behavior) if the student scored

themselves appropriately. All students scored themselves at the appropriate time throughout the study.

Treatment Integrity

Treatment integrity was assessed for 40% of all intervention sessions. This measure focused on the researcher's implementation of Experiment 2's IBRST procedures as planned. An 11-step fidelity checklist was used to assess treatment integrity (see Appendix C). As the intervention was implemented, the research assistant marked "yes" if the step was implemented correctly or "no" if it was not. Some of the steps evaluated included: a) Giving the IBRST data sheet to the student before the observation period, b) providing the child the opportunity to ask questions before the observation period begins, c) ensuring researcher does not provide prompts to start timer or mark the IBRST, and d) collecting the IBRST 2 min after the observation period has ended and providing praise if all 4 circles are completed. Treatment integrity was calculated in the same manner as Experiment 1. Treatment integrity was 100% across all scored sessions.

Social Validity

The same social validity questionnaire was provided at the end of Experiment 2 with the addition of the following questions: 1) I think I could set the levels on the IBRST by myself, 2) Doing this on my own is easy, 3) When I use the IBRST on my own, I think it helps me stay focused on my work, 4) I think I could set goals by myself using the IBRST. These questions were asked in the same format as before (5-point Likert-type scale) (see Appendix E).

Inter-Observer Agreement (IOA)

IOA was calculated in the same format as described in Experiment 1 for both problem behavior and on-task behavior (see Appendix F). IOA was calculated for 53% of all baseline sessions across participants. IOA was calculated for 40% of all intervention conditions across participants. For Danny in baseline, IOA for problem behavior was 100% and IOA for on-task behavior was 99.3% (range 99-100%). For Danny in treatment, IOA was 100% for problem behavior and 100% for on-task behavior. For Bailey in baseline, IOA was 100% for problem behavior and 98% for on-task behavior (range 96-100%). For Bailey in treatment, IOA was 100% for problem behavior and 100% for treatment. For Deborah in baseline, IOA was 100% for problem behavior and 97.7% for on-task behavior (range 93-100%). For Deborah in treatment, IOA was 100% for problem behavior and 100% in treatment.

Experimental Design

The experimental design of this study was a multiple baseline across participants with the following phases: baseline, independent self-monitoring with the IBRST, and follow-up.

Procedure

Baseline

Baseline was conducted in the same format as Experiment 1. The teacher was told to instruct class as normal and there was no self-monitoring with the IBRST. Only direct observation data was collected in the manner described above. Baseline was considered

complete once data were stable or trending contra-therapeutically for at least three data points in row.

Student Selection of Anchors and Training to Use Timer

After baseline data were completed, student selection of the anchors occurred. Participants were trained to set 15 min on a single-function timer using behavioral skills training. In this experiment, the students were required to select their own anchors for the IBRST. Students were provided with step-by-step instructions (see Appendix G) as well as information from his/her baseline data including minimum, maximum, and average scores for problem behavior and on-task behavior. Participants were, at first, not given any assistance by the researcher beyond the instruction sheet. Participants were given as much time as he/she requested to set the anchors; however, no opportunity was provided for the participant to ask the researcher any questions.

After the participant selected the anchors, the researcher checked the IBRST to ensure that five and only five anchors were selected with a 5 representing the worst day for problem behavior and the best day for appropriate behavior and that the anchors were in the appropriate numerical order and distributed equally among the anchors. If the anchors were not in numerical order (e.g. a 5 on the problem behavior scale means 5-10 instances whereas a 4 means 20-25 instances) or there were too many or not enough anchors or anchors were not distributed equally, the researcher provided a statement about the specific issue only and gave the anchor setting page back to the participant. Each time the anchors needed to be corrected, they were noted anecdotally on a separate data sheet (see Appendix H). After the participant corrected the errors, the researcher reviewed the anchors again for issues. Feedback continued to be provided

and noted as necessary until the anchors were set in the proper order and enough anchors were selected. This occurred in the same manner as described in Experiment 1: first providing a broad question, then a detailed question, and then providing the student with the anchor. Only Danny was able to set the anchors independently after 2 prompts, a broad question and then a detailed question. All other participants required the anchors to be set for them (i.e., after a broad question and a detailed question they were still unable to create the anchors), similar to the results of Experiment 1. Once the anchors were set appropriately, the intervention condition began.

Intervention

The intervention condition was conducted in a similar manner to Experiment 1; however, instead of being prompted by the researcher throughout, the participant was provided with a laminated checklist to use each observation period to score the IBRST independently. Since no reinforcement component was necessary during the first experiment, only self-monitoring using the IBRST was used for this experiment too.

In this condition, the checklist prompted the student to obtain a clean copy of the IBRST and a stopwatch from the researcher. Then, the checklist asked the participant to set the timer for 15 min as trained previously and to begin working on whatever workbook was currently assigned. At the end of 15 min, the timer went off and the participant was prompted by the checklist to circle 2 numbers on his/her IBRST, one for problem behavior and one for on-task behavior. The researcher checked to ensure the participants marked the IBRST and any verbal prompts were provided and noted, as necessary (see Appendix I). No prompts were required for any participant to ask for the data sheet and stop watch, set the timer, or score when appropriate. No feedback was provided on the accuracy of the IBRST. Next, the participant was instructed by

the checklist to place the IBRST to the side and set the timer again for 15 min. At the end of the second 15 min period, the checklist instructed the student to circle two numbers again and then give the timer and the completed IBRST to the researcher. Again, verbal praise was provided if the IBRST was completed correctly, but not regarding the accuracy of the score.

CHAPTER FIVE: EXPERIMENT 2: RESULTS

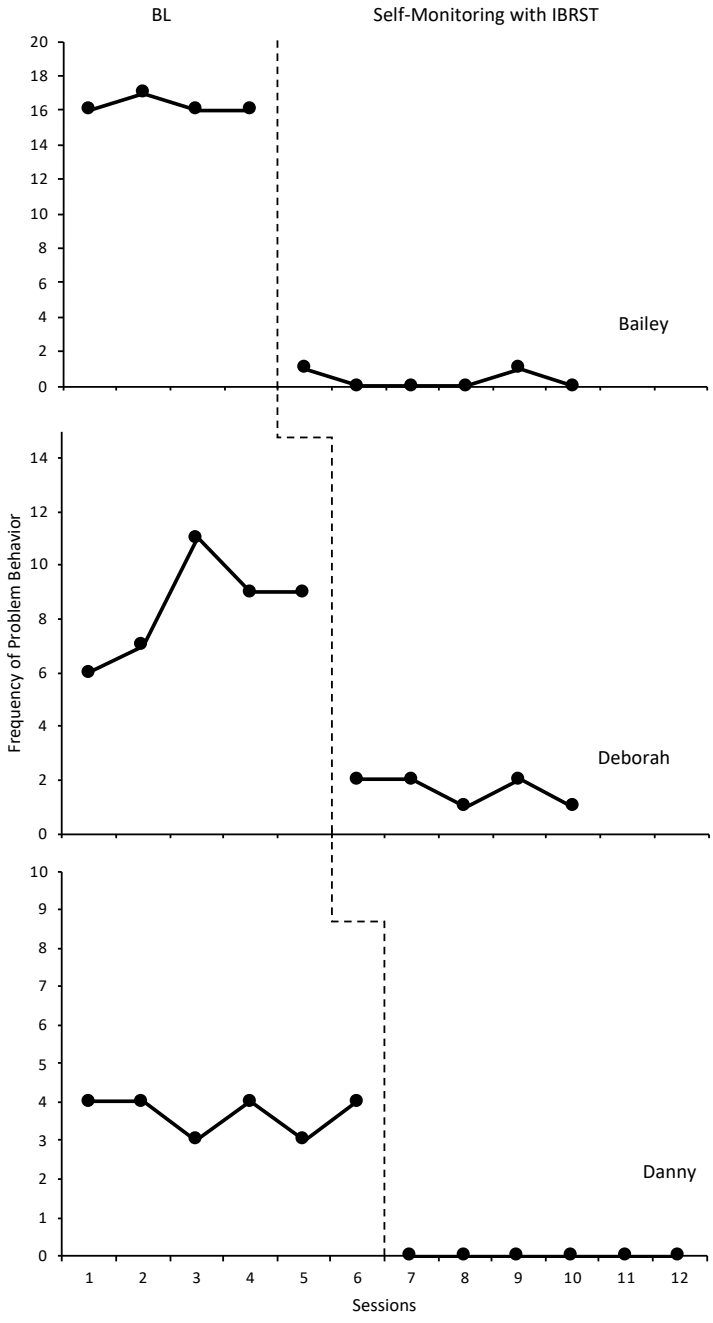

Figure 5. Results of self-monitoring using the IBRST on problem behavior (SDO data).

As shown in Figure 5, problem behavior decreases after the introduction of the intervention and for each participant. The average frequency of talking back for the last three sessions of baseline for Danny was 3.7 (range 3-4 instances). In treatment, talking back decreased to an average of 0 for the last three sessions. Bailey's average frequency of non-compliance for the last three sessions in baseline was 16.3 (range 16-17 instances). After implementing the self-monitoring intervention, non-compliance decreased to .33 instances (range 0-1 instance) for the last three sessions of intervention. Finally, for Deborah, property destruction occurred on average 9.7 instances (range 9-11) for the last three sessions of baseline. Following intervention, property destruction decreased to 1.33 instances (range 1-2 instances).

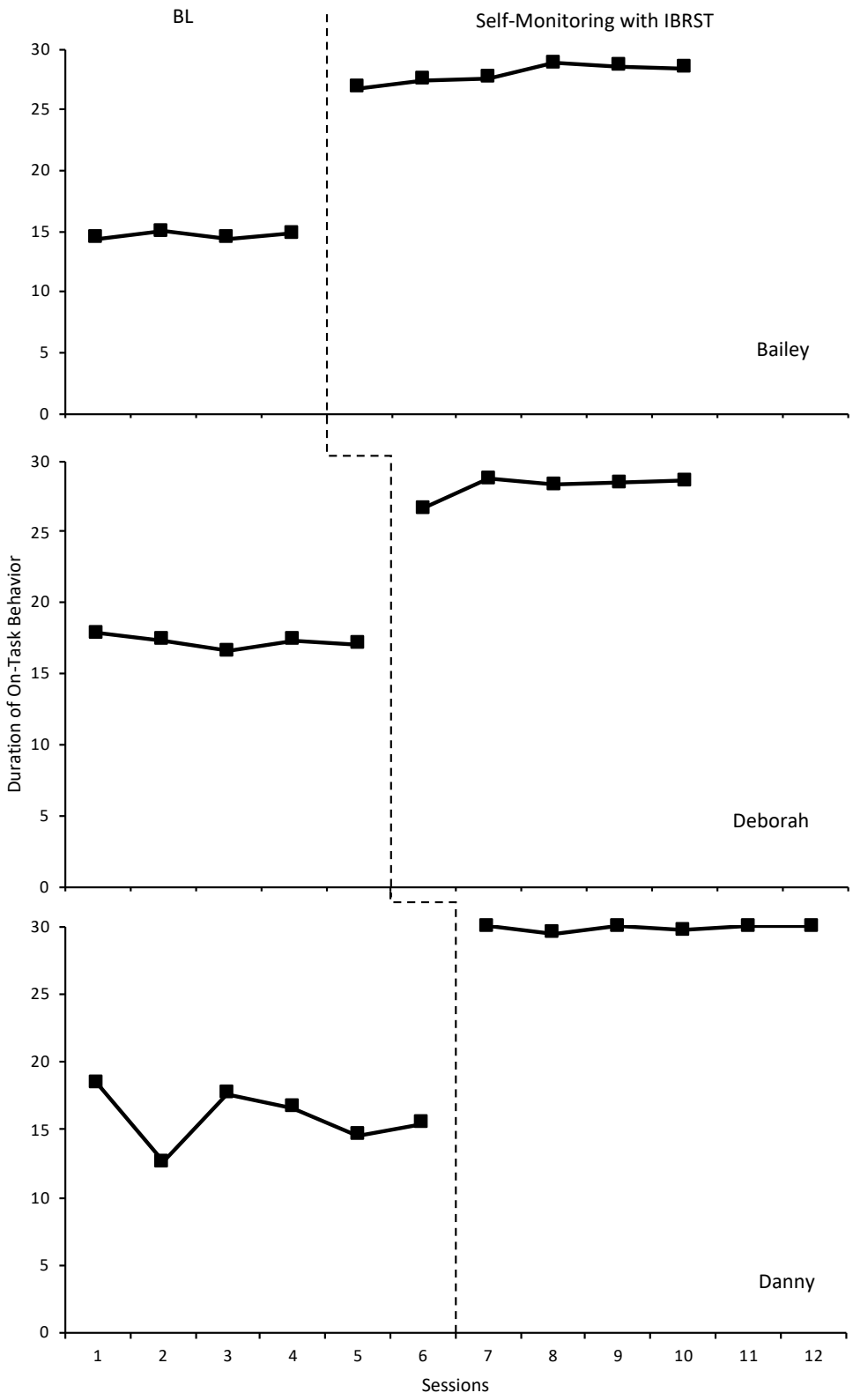

Figure 6. Results of self-monitoring using the IBRST on problem behavior (SDO data).

As shown in Figure 6, there is an increase in on-task behavior only after sequential introduction of the intervention. The average number of minutes of on-task behavior for the last three sessions of baseline for Danny was 15.6 min (range 14.6-16.6 min). In treatment, this increased to an average of 29.9 min (range 29.7-30 min) for the last three sessions. Bailey's average number of minutes of on-task for the last three sessions in baseline was 17 min (range 16.6-17.3 min). After implementing the self-monitoring intervention, on-task increased to an average of 28.4 min (range 28.3-28.6 min) for the last three sessions of intervention. Deborah was on-task on average 14.7 min (range 14.4-15 min) for the last three sessions of baseline. Following intervention, on-task increased to 28.6 min (range 28.4-28.8 min).

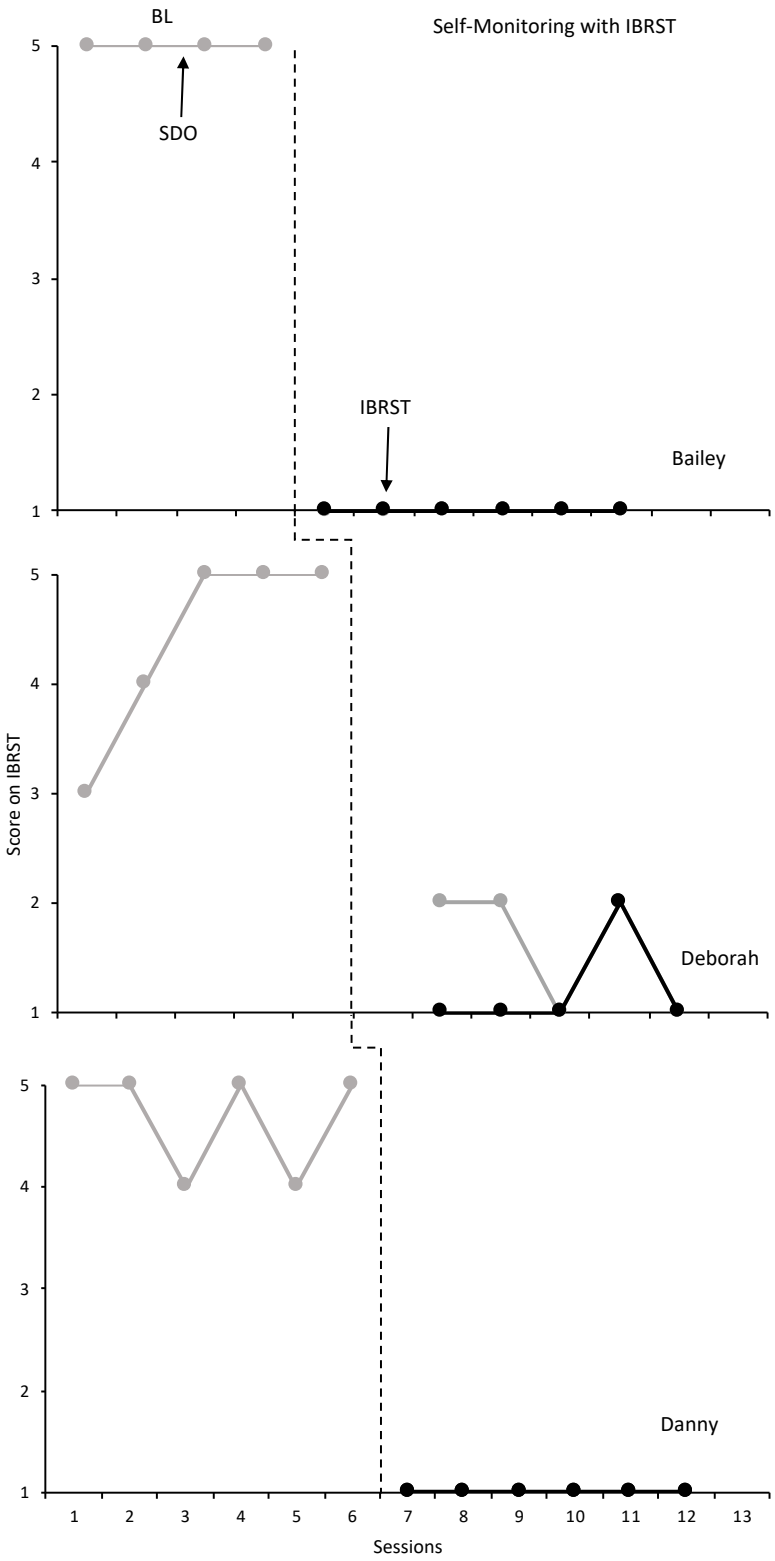

Figure 7. Correlation between converted direct observation scores (SDO) and student-scored IBRST scores for problem behavior.

As indicated in Figure 7, only two data points did not correlate precisely with direct observation scores. The only participant in which there were differences between direct observation data and IBRST scores was Deborah. There was perfect agreement between IBRST and SDO data on 15/17 opportunities or 88% of the time. When differences occurred, it was only by 1 point on the IBRST scale. At no point did scores differ from SDO data for Danny or for Bailey, indicating high levels of agreement between the IBRST and direct observation data. Cohen's kappa was calculated for these data and indicated the correlation was k = .45, or a moderate correlation.

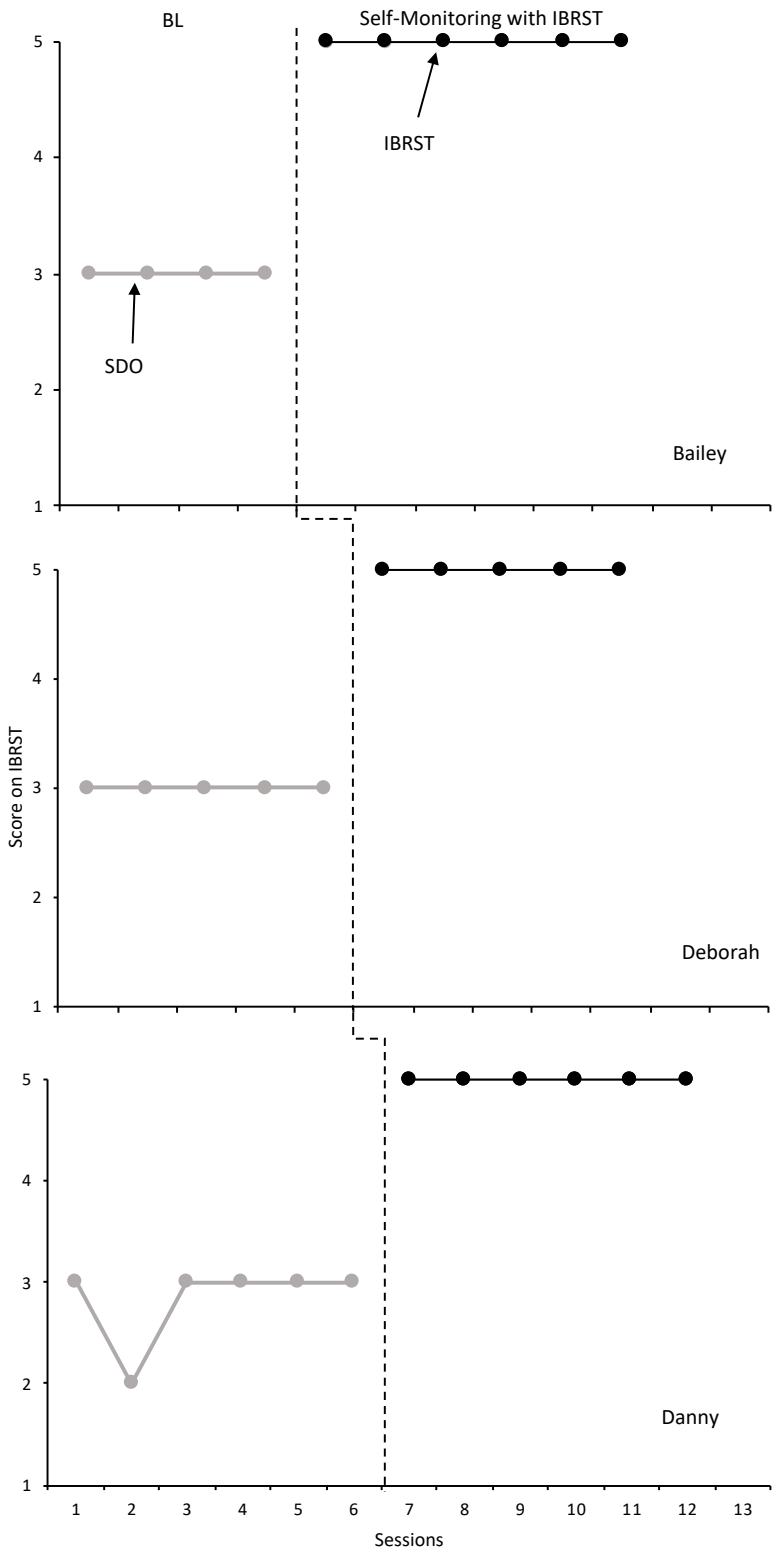

Figure 8. Correlation between converted direct observation scores (SDO) and student-scored IBRST scores for on-task behavior.

As indicated in Figure 8, all data points overlapped between researcher-scored and converted SDO scores and student-scored IBRST data, again indicating strong correlation between SDO data and IBRST scores. Cohen's kappa coefficient was calculated and indicated k = 1.0, indicating perfect correlation.

Table 3. Student social validity results for Experiment 2.

	The IBRST helps me stay on-task during class.	The IBRST is easy to use.	The IBRST helps my behavior during class.	The IBRST would help other kids during class.	I liked using the IBRST during class.	I think I can do the IBRST all by myself.	I think using the IBRST helps me stay focused.	I think I could set goals using the IBRST by myself.
Danny	4	4	5	5	4	3	4	3
Deborah	5	4	5	4	5	4	5	4
Bailey	4	5	5	4	5	2	4	2

Student social validity results indicated high-likeability and perceived effectiveness of the intervention. The average scores were as follows: 1. *4.3* (range 4-5), 2. *4.3* (range 4-5), 3. *5*, 4. *4.3* (range 4-5), 5. *4.7* (range 4-5), 6. *2.7* (range 2-4), 7. *4.3* (range 4-5), 8. *3* (range 2-4). Students scored the intervention as easy to use, that they liked using it, that the IBRST helped them stay on task, and that it would help other children stay on-task during class. However, students rated two questions, that they could set the anchors on the IBRST by themselves (Question 6) and that they could use the IBRST themselves (Question 8) lower than other questions. This indicates that the students did not believe they could run the entire IBRST self-monitoring program as written without assistance. This is consistent with implementation issues which occurred during the study when only Danny was able to create his own anchors, but still

required some assistance. Deborah and Bailey were not able to set their own anchors and required the researcher to set them and scored these items lower than Danny did.

Table 4. Teacher social validity results for Experiment 2.

	Self-monitoring with IBRST was acceptable to decrease problem behavior.	Self-monitoring with IBRST was acceptable for increasing on-task behavior.	Self-monitoring was effective at decreasing problem behavior.	Self-monitoring was effective at increasing on-task behavior.	Suggest IBRST to other teachers.	Student's behavior warranted need for treatment.	I would use IBRST during other instructional times.	The student appeared to enjoy self-monitoring using IBRST.
Teacher 1	5	5	5	4	4	5	5	5
Teacher 2	5	5	5	5	4	5	5	5

Teacher social validity results also indicated high-likeability and perceived effectiveness of self-monitoring with the IBRST. The teacher scored the intervention a 5 as an acceptable option for decreasing problem behavior, a 5 as an acceptable option for increasing on-task behavior, a 5 as an acceptable option for increasing on-task and decreasing problem behavior, a 4 that she would suggest self-monitoring using IBRST to other teachers, and a 5 that the student appeared to enjoy using the IBRST.

CHAPTER SIX: DISCUSSION

The purpose of Experiment 1 was two-fold. First, it was to examine if there was concurrent validity between direct observation data and student self-monitoring data collected using the IBRST and second, to what extent would student self-monitoring using the IBRST result in reduced problem behavior and increased appropriate behavior in a classroom setting. In Experiment 1, all participants greatly increased on-task behavior and decreased problem behavior to almost zero after self-monitoring was sequentially introduced for each participant. Concurrent validity results indicated a large degree of agreement between direct observation data with perfect agreement for on-task behavior.

Experiment 2 also had two purposes: 1) to what extent would students be able to independently set their own anchors and use the IBRST to self-monitor their own behavior and 2) would there be concurrent validity with direct observation data? Unfortunately, two of the students were not able to set any of their anchors independently and one student, Danny, needed help with the first anchor before being able to set the rest of the anchors by himself. Bailey and Deborah were prompted via a broad, then detailed question regarding each anchor. The participants attempted to create the anchor; however, the anchor did not make sense based on their data (i.e. an anchor that was extremely high or low) nor was it in the range of data. After the researcher asked a broad and follow-up detailed question and saw that the anchors were still not appropriate, the anchors were set for the two participants. The students were able to use the IBRST independently after the anchors were set and did not require any prompts to set the timer

or circle data appropriately when the timer went off every 15 min. Concurrent validity for this study again indicated perfect agreement for on-task behavior and moderate agreement for problem behavior (as found by Cohen's kappa).

The results of these two experiments confirm the results found by several researchers who evaluated the effects of self-monitoring on increasing appropriate behaviors in classroom settings (Graham-Day et al., 2010; Kartal & Ozkan, 2015; Rosenbloom et al., 2016). In contrast to Stasolla and colleages (2017), this intervention did not require prompting every 20s to record behavior nor used an additional reinforcement component yet remained effective for increasing on-task behavior. This suggests that an additional reinforcement component may not be necessary to improve certain behaviors for some students. It might be considered that the timer itself was functioning as a strong discriminative stimulus for on-task behaviors with the students in this study. Although during the delivery of the timer, attention was minimized (i.e., the researcher just set the timer and placed it on the participants desk (Exp. 1) or the timer was just handed to the participant when requested (Exp. 2), there is a chance that any attention perceived on behalf of that participant functioned as attention. The classroom environment, as created by the school, set up a strong establishing operation for attention as this was minimized via the use of cubicles. In addition, timers are a common ABA intervention strategy in a classroom setting and these clients may have had several timer-based interventions in their histories. Timers could have been paired with reinforcement before the study began; however, timers were never used to increase on-task behavior for any of the participants in this study before the study began. In addition, Danny had never received timer-based interventions and his results were similar to that of Bailey and Deborah.

It is also important to note that students were only required to score their own behaviors every 15 min. In comparison to the study completed by Stasolla et al. (2017), the response effort required of the student was much less. In the Stasolla study, participants were required to score their own behavior every 20 s as indicated by the application, Clicker 5. In this study, participants were only asked to score themselves one time every 15 min. As indicated in the social validity results, the students believed the IBRST was easy to use and did not require too much of their time. This might increase the likelihood that the intervention would remain in place after the researcher was no longer present. The teacher also indicated she would be likely to use the IBRST with other students and recommend it to other teachers.

Similar to other self-monitoring studies, the most significant response effort for implementing the intervention was placed on the student (Moore et al., 2013). The student was responsible for setting the timer and scoring themselves and the teacher was not using valuable class time to record data. The participants were also able to see what their respective best day should be for both problem behavior and on-task behavior. This may have served as an additional discriminative stimulus to remain on-task. This may have also provided additional "buy-in" on behalf of the student as they had to work with the researcher to set their own anchors and score themselves. Attention to this factor is critical as interventions that are viewed as impractical or ineffective are less likely to be adopted once the researcher is no longer present (Leko, 2014; Lloyd & Heubusch, 1996).

The findings of this study are in contrast to many studies that evaluated the effects of self-monitoring on decreasing problem behavior. Previously self-monitoring has been shown to be more effective in increasing on-task behavior but having varied or no result on decreasing problem behavior (Crawley et al., 2006; Crutchfield et al., 2014; & Davis et al., 2016). The

current study demonstrated that self-monitoring was effective in decreasing problem behavior via visual analysis of the data. This study evaluated a variety of problem behaviors including verbal outbursts, talking back, non-compliance, and property destruction. This suggests self-monitoring may be a simple, yet effective intervention for a variety of problem behaviors seen in the classroom.

It might be considered that the IBRST also served as a break from the participant's routine curriculum provision. The students were required to work independently on workbooks in individual cubicles from the beginning of the school day until they went home with the exception of lunch and a short break for recess. Therefore, novel stimuli were introduced into an otherwise austere environment. Furthermore, the use of the timer signaled a 2-min break from academic tasks when they were given the opportunity to score their responses to the IBRST. The students are typically not allowed breaks from work so this might have functioned as a small but powerful reinforcer.

The results of this study suggest that the IBRST may be an accurate and reliable measure that is practical for students to use during academic times and an appropriate measure to use in assessing student outcomes. Across both experiments, there was perfect agreement 56/60 times or 93.3% of the time. When there were differences in scoring, there was only a difference by 1 point. This means that when asked, students were able to accurately score their own behavior and that this data may be useful to the teacher providing the intervention if the teacher is unwilling or unable to take data. Although it might be suggested that students may be untruthful in scoring their behavior, a simple, additional reinforcement component could be added by providing reinforcement for accurate data collection.

Finally, this study addressed the limitations of previous studies regarding a lack of social validity data (Bruhn et al., 2017; Crawley et al., 2006; Davis et al., 2016). Both teachers and students evaluated the use of the intervention across both experiments and evaluated perceived effectiveness, feasibility, and likeability of the self-monitoring intervention. Students and teachers appeared to like the intervention and believe it to be effective in decreasing problem behavior and increasing on-task behavior. Again, it is important that both the participants and the implementers of the intervention enjoy using the intervention and believe it to be effective and efficient as these are key indicators of whether the intervention will be used once the researcher is no longer present (Kennedy, 2002).

Although this study suggests promising results, there are several limitations. The first limitation is the short period of time participants were involved in the study. Data collection only occurred over a period of 1-2 months with follow-up data occurring shortly after that. It may be that the participants were reacting in a certain manner to the intervention because it was a novel event. If the intervention was conducted for an extended period of time, this effect, if it occurred, may have decreased over time and changes in level may not have been as large as was seen. Future research should evaluate this intervention over an extended period of time, perhaps an entire semester or school year, to evaluate the effects of the intervention long-term. Future research could also examine the effects of additional follow-up probes.

Similarly, a second limitation of this study may have been the private school setting in which the study was conducted. The classroom was set up to be as sterile as possible. As previously mentioned, the students sat in silence in individual cubicles and worked on individual workbooks the entire school day. The teacher did not interact at all with the students unless they had an academic question, and these often resulted in aversive vocal remarks from the

supervisor. The only breaks during which the students got to interact with other children or objects was during a 30-min lunch or recess after lunch. Therefore, the classroom setting may have functioned as an establishing operation for any novel stimuli presented during work. This may mean that results seen in this study may not generalize to other classroom settings where more interaction occurs. Future research could evaluate the results of self-monitoring using the IBRST in a public school setting to determine generalization of intervention effects.

A third limitation of this study is the number of participants included in this study. There were only three participants for both experiments. Although generalization across routines was demonstrated across these students, there was not a true test of external validity to determine the effects of self-monitoring using the IBRST for many students. Future research should examine this intervention as it has the potential to be an effective and efficient classroom management strategy.

Another limitation is the teacher was not involved in the study except for answering the social validity questionnaire. It is unclear the feasibility of this intervention if the teacher had to help students create the anchors which is more representative of what would occur in a typical classroom setting. Future research should evaluate the extent to which school teachers can independently develop and use self-monitoring with the IBRST in their classrooms. Future research should also evaluate best practices to help the students create their own anchors. As the participants in this study had such stable data in baseline, math was required to set the anchors which may have been difficult for the participants. This deviated from the methodology used by Iovannone et al. (2014) in which data were simply selected from baseline data instead of calculated. It might be important to develop a manualized system for teachers and then students to use for selecting anchors to create a more independent strategy for classroom use.

Despite the aforementioned limitations, the results of the study are encouraging. All three students showed large decreases in problem behavior and large increases in on-task behavior across routines. This study extends the current research as to the possible uses of self-monitoring in a classroom setting for not only appropriate behaviors, but for problem behaviors as well. This study also suggests the IBRST may be used as an accurate and reliable tool by other, less-trained data collectors. On over 90% of all opportunities, there was perfect agreement between converted direct observation data and student-scored IBRST data. Overall, self-monitoring using the IBRST was a valid and reliable data-collection method that served a dual purpose of an effective classroom management tool.

REFERENCES

Barnes, S. (2015). An evaluation of the prevent-teach-reinforce model within a multi-tiered intervention system. (Unpublished thesis).

Bruhn, A., McDaniel, S., Kreigh, C. (2015). Self-monitoring interventions for students with behavior problems: A systematic review of current research. *Behavioral Disorders, 40,* 102-121.

Bruhn, A. L., Woods-Groves, S., Fernando, J., Choi, T., Troughton, L. (2017). Evaluating technology-based self-monitoring as a Tier 2 intervention across middle school settings. *Behavioral Disorders, 42,* 119-131.

Carr, E. G., Dunlap, G., Horner, R. H., Koegel, R. L., Turnbull, A., Sailor, W., … Fox, L. (2002). Positive behavior support: Evolution of an applied science. *Journal of Positive Behavior Interventions, 4,* 4-16.

Cariveau, T., & Kodak, T. (2017). Programming a randomized dependent group contingency and common stimuli to promote durable behavior change. *Journal of Applied Behavior Analysis, 50,* 121-133.

Chaffee, R. K., Briesch, A. M, Johnson, A. H., & Volpe, R. J. (2017). A meta-analysis of class-wide interventions for supporting student behavior. *School Psychology Review, 46,* 149-164.

Chafouleas, S. M., Riley-Tillman, T. C., McDougal J. L. (2002). Good, bad, or in-between: How does the daily behavior report card rate? *Psychology in the Schools, 39,* 157-169. doi: 10.1002/pits.10027

Chafouleas, S. M., McDougal, J. L., Riley-Tillman, T. C., Panahon, C. J., & Hilt, A. M. (2005). What do daily behavior report cards (DBRCs) measure? An initial comparison of DBRCs with direct observation for off-task behavior. *Psychology in the Schools, 42,* 669-676.

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

Crawley, S. H., Lynch, P. L., & Vannest, K. (2006). The use of self-monitoring to reduce off-task behavior and cross-correlation examination of weekends and absences as an antecedent to off-task behavior. *Child and Family Behavior Therapy, 28,* 29-48.

Crutchfield, S. A., Mason, R. A., Chambers, A., Wills, H. P., & Mason, B. (2015). Use of a self-monitoring application to reduce stereotypic behavior in adolescents with autism: A preliminary investigation of I-Connect. *Journal of Autism and Developmental Disorders, 45,* 1146-1155.

Davis, J. L., Mason, B. A., Davis, H. S., Mason, R. A., & Crutchfield, S. A. (2016). Self-monitoring interventions for students with ASD: A meta-analysis of school-based research. *Journal of Autism and Developmental Disorders, 3,* 196-208.

Debnam, K. J., Pas, E. T., Bradshaw, C. P. (2012). Secondary and tertiary support systems in schools implementing school-wide positive behavioral interventions and supports: A preliminary descriptive analysis. *Journal of Positive Behavior Interventions, 14,* 142-152.

Deno, S. L. (1995). School psychologist as problem solver. In A. Thomas & J. Grimes (Eds.). *Best practices in school psychology* (Vol. 3, pp. 471-484). Washington, DC: National

Diegelman, K. M., & Test, D. W. (2018). Effects of a self-monitoring checklist as a component of the self-directed IEP. *Education and Training in Autism and Developmental Disabilities, 53,* 73-83.

Dooley, P., Wilczenski, F. L., & Torem, C. (2001). Using an activity schedule to smooth school transitions. *Journal of Positive Behavior Interventions, 3,* 57-61.

Dunlap, G., Iovannone, R., Kincaid, D., Wilson, K., Christiansen, K., … English, C. (2010). Prevent-teach-reinforce: A school-based model of individualized behavior support. Baltimore, MD: Paul H. Brookes.

Emerson, E., & Einfeld, S. (2010). Emotional and behavioral difficulties in young children with and without developmental delay: A bi-national perspective. *Journal of Child Psychology and Psychiatry, 51,* 583-593.

Fauth, R. C., Platt, L., Parsons, S. (2017). The development of behavior problems among disabled and non-disabled children in England. *Journal of Applied Developmental Psychology, 52,* 46-58.

Fragale, C., Rojeski, L., O'Reilly, M., & Gevarter, C. (2016). Evaluation of functional communication training as a satiation procedure to reduce challenging behavior in instructional environments for children with autism. *International Journal of Developmental Disabilities, 62,* 139-146.

Fronapfel, B., Dunlap, G., Flagtvedt, K., Strain, P. & Lee, J. (2018). Prevent-Teach-Reinforce for young children: A program description and demonstration of implementation in an early childhood setting. *Education and Treatment of Children, 41,* 233-248.

Ganz, J. (2008). Self-monitoring across age and ability levels: Teaching students to implement their own positive behavioral interventions. *Preventing School Failure: Alternative Education for Children and Youth, 53,* 39-48.

Graham,-Day, K., Gardner, III R., & Hsin, Y. (2010). Increasing on-task behaviors of high school students with Attention-Deficit Hyperactivity Disorder: Is it enough? *Education and Treatment of Children, 28,* 328-344.

Greenberg, J., Putnam, H., & Walsh, K. (2014). *Training our future teachers: Classroom management*. Washington, DC: National Council on Teacher Quality. (pp. 1-43).

Gulchak, D. (2008). Using a mobile handheld computer to teach a student with an emotional and behavioral disorder to self-monitor attention. *Education and Treatment of Children, 31,* 567-581.

Gureasko-Moore, S., DuPaul, G., & White, G. (2007). Self-management of classroom prepraredness and homework: Effects on school functioning of adolescent with attention deficit hyperactivity disorder. *School Psychology Review, 36,* 647-664.

Hoge, R. D., & Andrews, D. A. (1992). Assessing conduct problems in the classroom. *Clinical Psychology Review, 12,* 1-20.

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

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

Ingram, K., Lewis-Palmer, T., Sugai, G. (2005). Function-based intervention planning: Comparing the effectiveness of FBA function-based and non-function-based intervention plans. *Journal of Positive Behavior Interventions, 7,* 224-236.

Iovannone, R., Greenbaum P., Wang, W., Dunlap, G., & Kincaid, D. (2014). Inter-rater agreement of the individualized behavior rating scale tool. *Assessment for Effective Interventions, 39,* 195-207.

Iovannone, R., Greenbaum, P., Wang, W., Kincaid, D., Dunlap, G., & Strain, P. (2009). Randomized controlled trial of the Prevent-Teach-Reinforce (PTR) tertiary intervention for students with problem behaviors. *Journal of Emotional and Behavioral Disorders, 17,* 213-225.

Joseph, L. M., & Eveleigh, E. L. (2011). A review of the effects of self-monitoring on reading performance of students with disabilities. *The Journal of Special Education, 45,* 43-53.

Kartal, M. S., Ozkan, S. Y. (2015). Effects of class-wide self-monitoring on on-task behaviors of preschoolers with developmental disabilities. *Education and Training in Autism and Developmental Disabilities, 50,* 418-432.

Kennedy, C. H. (2002). The maintenance of behavior change as an indicator of social validity. *Behavior Modification, 26,* 594-604.

Kennedy, C., Meyer, K., Knowles, T., & Shukla, S. (2000). Analyzing the multiple functions of stereotypical behavior for students with autism: Implications for assessment and treatment. *Journal of Applied Behavior Analysis, 33,* 559-571.

Kim, M., Blair, K. S., & Lim, K. (2014). Using tablet assisted Social Stories to improve classroom behavior for adolescents with intellectual disabilities. *Research in Developmental Disabilities, 35,* 2241-2251.

Koyama, T, & Wang, H. (2011). Use of an activity schedule to promote independent performance of individuals with autism and other intellectual disabilities: A review. *Research in Developmental Disabilities, 32,* 2235-2242.

Landa, R. J., Gross, A. L., Stuart, E. & Faherty, A. (2013). Developmental trajectories in children with and without autism spectrum disorders: The first 3 years. *Child Development, 84,* 429-442.

Leko, M. M. (2014). The value of qualitative methods in social validity research. *Remedial and Special Education, 35,* 275-286.

Lloyd, J. W., & Heubusch, J. D. (1996). Issues of social validation in research on serving individuals with emotional or behavioral disorders. *Behavioral Disorders, 22,* 8-14.

Maggin, D. M., Pustejovsky, J. E., & Johnson, A. (2017). A meta-analysis of school-based group contingency interventions for students with challenging behavior: An update. *Remedial and Special Education, 38,* 353-370.

Martinez, D. F. (2016). Use of the Individualized Behavior Rating Scale Tool (IBRST) as a self-monitoring tool to improve classroom behavior. (Unpublished thesis).

Massey, N., Wheeler, J. (2000). Acquisition and generalization of activity schedules and their effects on task engagement in a young child with autism in an inclusive pre-school classroom. *Education and Training in Mental Retardation and Developmental Disabilities, 35,* 326-333.

Miller, B., & Taber-Doughty, T. (2014). Self-monitoring checklists for inquiry problem-solving: Functional problem-solving methods for students with intellectual disability. *Education and Training in Autism and Developmental Disabilities, 49,* 555-567.

Moore, D. W., Anderson, A., Glassenbury, M., Lang, R., & Didden, R. (2013). Increasing on-task behavior in students in a regular classroom: Effectiveness of a self-management procedure using a tactile prompt. *Journal of Behavioral Education, 22,* 302-311.

Plavnick, J., Ferrir, S. J., & Maupin, A. N. (2010). The effects of self-monitoring on the procedural integrity of a behavioral intervention for young children with developmental disabilities. *Journal of Applied Behavior Analysis, 43,* 315-320.

Public Agenda (2004).

Riley-Tillman, T. C., Chafouleas, S. M., Sassu, K. A., Chanese, J. A. M., Glazer, A. D. (2008). Examining the agreement of direct behavior ratings and systematic direct observation data for on-task and disruptive behavior. *Journal of Positive Behavior Interventions, 10,* 136-143.

Rosenbloom, R., Mason, R., Wills, H., Mason, B. (2016). Technology delivered self-monitoring application to promote successful inclusion of an elementary student with autism. *Assistive Technology, 28,* 9-16.

Schindler, H., & Horner, R. (2005). Generalized reduction of problem behavior of young children with autism: Building trans-situational interventions. *American Journal on Mental Retardation, 110,* 36-47.

Smith, D. J., Young, K., West, R., Morgan, D., & Rhode, G. (1988). Reducing the disruptive behavior of junior high school students: A classroom self-management procedure. *Behavioral Disorders,* 231-239.

Stasolla, F., Caffo, A. O., Perilli, V., Boccasini, A., Damiani, R., … Albano, A. (2017). Comparing self-monitoring and differential reinforcement of an alternative behavior to promote on-task behavior by three children with cerebral palsy: A pilot study. *Life Span and Disability, 20,* 63-92.

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

Sugai, G., & Horner, R. H. (1999). Discipline and behavioral support: Practices, pitfalls, and promises. *Effective School Practices, 17,* 10-22.

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C., … Ruef, M. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions, 2,* 131-143.

Sugai, G., Sprague, J. R., Horner, R. H., Walker, H. M. (2000). Preventing school violence: The use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of Emotional and Behavioral Disorders, 8,* 94-101.

The New Teacher Project. (2013). *Perspectives of irreplaceable teachers.* Retrieved from http://tntp.org/assets/documents/TNTP_Perspectives_2013.pdf

Walker, H. M., Horner, R. H., Sugai, G., Bullis, M., Sprague, J. R., Bricker, D., & Kaufman, M. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional and Behavioral Disorders, 4,* 193-256.

Walker, V., Lyon, K. J., Loman, S., Sennott, S. (2018). A systematic review of Functional Communication Training (FCT) interventions involving augmentative and alternative communication in school settings. *Augmentative and Alternative Communication, 34,* 118-129.

Vandermeer, J., Beamish, W., Milford, T., & Lang, W. (2013). iPad-presented Social Stories for young children with autism. *Developmental Neurorehabilitation, 18,* 75-81.

US Department of Education (2017).

APPENDICES

Appendix A: Direct Observation Recording Sheet

Participant _____ Observer _____ BL/SM Session #_____

Routine/Activity _____ Target Behavior: _____

DIRECTIONS: Tally each occurrence of problem behavior which meets the definition above.

The count total occurrences and mark this number at the bottom of the box.

Total Frequency: _____

DIRECTIONS: Using a stopwatch, mark the on-set and off-set of each instance of on-task behavior. Count the number of total seconds and write below. Then divide by 1800 for % on task.

On-Set	Off-set	Duration in s

On-Set	Off-Set	Duration in s

Total Duration in Seconds_____/1800 s =

_____% total on-task

Appendix B: Sample Individualized Behavior Rating Scale Tool

Student:_____ Date: _____

Target Behavior	IBRST Scoring Key	Session		Session		Session ____		Session ____		Session ____	
		Ob. 1	Ob. 2	Ob. 1	Ob. 2	Ob. 1	Ob. 2	Ob. 1	Ob. 2	Ob. 1	Ob. 2
Problem Behavior *Definition:*	Frequency:										
	11+	5	5	5	5	5	5	5	5	5	5
	8-10	4	4	4	4	4	4	4	4	4	4
	5-7	3	3	3	3	3	3	3	3	3	3
	2-4	2	2	2	2	2	2	2	2	2	2
	0-2	1	1	1	1	1	1	1	1	1	1
On-Task Behavior *Definition:*	Duration:										
	25-30 min	5	5	5	5	5	5	5	5	5	5
	19-24 min	4	4	4	4	4	4	4	4	4	4
	13-18 min	3	3	3	3	3	3	3	3	3	3
	7-12 min	2	2	2	2	2	2	2	2	2	2
	0-6 min	1	1	1	1	1	1	1	1	1	1

Appendix C: Treatment Fidelity Checklist

Treatment Fidelity Checklist – Study 1	
1. Passed out IBRST prior to academic period beginning	Yes/No/NA
2. State academic period has begun	Yes/No/NA
3. Set timer for 15 min	Yes/No/NA
4. After 15 min, instruct students to score IBRST in Ob. 1	Yes/No/NA
5. Waited 2 min for scoring	Yes/No/NA
6. Provide praise for scoring IBRST (not accuracy)	Yes/No/NA
7. Instruct students to place IBRST to side of desk	Yes/No/NA
8. Set 2nd timer for 15 min	Yes/No/NA
9. After 15 min, instruct students to score IBRST in Ob. 2	Yes/No/NA
10. Provide 2 min for scoring	Yes/No/NA
11. Provide praise for scoring IBRST (not accuracy)	Yes/No/NA
12. Collect completed IBRST	Yes/No/NA
13. Provide praise for participating appropriately	Yes/No/NA
Total "Yes" = _____/13 total steps	Percent Fidelity:____

Treatment Fidelity Checklist – Study 2	
1. Passed out IBRST prior to academic period beginning	Yes/No/NA
2. State academic period has begun	Yes/No/NA
3. Do not provide prompts to start timer – start personal timer	Yes/No/NA
4. After 15 min has elapsed, determine if student has scored IBRST	Yes/No/NA
5. If not scored within 2 min, prompt student to complete and mark if prompts are needed	Yes/No/NA
6. Set 2nd timer once student has placed IBRST away	Yes/No/NA
7. Do not provide prompts to start 2nd timer – start personal timer	Yes/No/NA
8. After 15 min has elapsed, determine if student has scored IBRST	Yes/No/NA
9. If not scored within 2 min, prompt student to complete and mark if prompts are needed	Yes/No/NA
10. Collect completed IBRST	Yes/No/NA
11. Provide praise for participating appropriately	Yes/No/NA
Total "Yes" = _____/11 total steps	Percent Fidelity:____

Appendix D: Social Validity (Teachers)

DIRECTIONS: Please circle the number that most closely corresponds to your opinion.

1. Self-monitoring using the IBRST was an acceptable option for decreasing student's problem behavior.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

2. Self-monitoring using the IBRST was an acceptable option for increasing student's on-task behavior.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

3. Self-monitoring using the IBRST was effective at decreasing the student's problem behavior.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

4. Self-monitoring using the IBRST was effective at increasing the student's on-task behavior.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

5. I would suggest the IBRST to other teachers.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

6. The student's behavior warranted the need for intervention.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

7. I would be willing to use the IBRST during other instructional times.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

8. The student appeared to enjoy self-monitoring their behavior using the IBRST.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

Comments/suggestions for improvement:

Appendix E: Student Social Validity: Study 1

DIRECTIONS: Please circle the number that most closely corresponds to your opinion.

1. The IBRST helps me stay on-task during class.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

2. The IBRST is easy to use.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

3. The IBRST helps my behavior during class.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

4. The IBRST would help other kids during class.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

5. I liked using the IBRST during class.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

Comments?

Appendix E: Student Social Validity: Study 2

DIRECTIONS: Please circle the number that most closely corresponds to your opinion.

1. The IBRST helps me stay on-task during class.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

2. The IBRST is easy to use.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

3. The IBRST helps my behavior during class.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

4. The IBRST would help other kids during class.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

5. I liked using the IBRST during class.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

6. I think I can do the IBRST by myself.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

7. I think using the IBRST helps me stay focused.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

8. I think I could set goals using the IBRST by myself.

Strongly Disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly Agree 5

Comments?

Appendix F: IOA for Direct Observation Data

Participant: _______________ Session #:__________________

Problem Behavior

_______ (Lower frequency)/_________(Higher frequency) = _______________

On-Task

_________(Lower %)/_________(Higher %) = _____________

Participant: _______________ Session #:__________________

Problem Behavior

_______ (Lower frequency)/_________(Higher frequency) = _______________

On-Task

_________(Lower %)/_________(Higher %) = _____________

Participant: _______________ Session #:__________________

Problem Behavior

_______ (Lower frequency)/_________(Higher frequency) = _______________

On-Task

_________(Lower %)/_________(Higher %) = _____________

Participant: _______________ Session #:__________________

Problem Behavior

_______ (Lower frequency)/_________(Higher frequency) = _______________

On-Task

_________(Lower %)/_________(Higher %) = _____________

Appendix G: Student Instruction Sheet

Lowest Score _____ Average Score _____ Highest Score _____

Problem Behavior Anchor Checklist	
1. Look at the scores above.	Yes/No/NA
2. Start with setting anchor 4. This should be higher than average but lower than highest score. Write the score below.	Yes/No/NA
3. Set anchor 5. This should be your highest score. Write the score below.	Yes/No/NA
4. Set anchor 1. This should be lower than your lowest score. Write the score below.	Yes/No/NA
5. Set anchor 2. This should be your lowest score. Write the score below.	Yes/No/NA
6. Set anchor 3. This should be your average score. Write the score below.	Yes/No/NA
7. Double-check to make sure you have all 5 anchors.	Yes/No/NA
8. The number you wrote down is the highest number of the anchor. Make each anchor a range of numbers. *For example, if your anchor 2 is 6 and anchor 1 is 3, make Anchor 2 (4-6) and make Anchor 1 (0-3).*	Yes/No/NA

Anchor 1: _____	Anchor 2: _____	Anchor 3: _____	Anchor 4: _____	Anchor 5: _____

Lowest Score _____ Average Score _____ Highest Score _____

On-Task Behavior Anchor Checklist	
1. Look at the scores above.	Yes/No/NA
2. Start with setting anchor 4. This should be higher than average but lower than highest score. Write the score below.	Yes/No/NA
3. Set anchor 5. This should be your highest score. Write the score below.	Yes/No/NA
4. Set anchor 1. This should be lower than your lowest score. Write the score below.	Yes/No/NA
5. Set anchor 2. This should be your lowest score. Write the score below.	Yes/No/NA
6. Set anchor 3. This should be your average score. Write the score below.	Yes/No/NA
7. Double-check to make sure you have all 5 anchors.	Yes/No/NA
8. The number you wrote down is the highest number of the anchor. Make each anchor a range of numbers. *For example, if your anchor 2 is 6 and anchor 1 is 3, make Anchor 2 (4-6) and make Anchor 1 (0-3).*	Yes/No/NA

Anchor 1: _____	Anchor 2: _____	Anchor 3: _____	Anchor 4: _____	Anchor 5: _____

Appendix H: Corrections for Anchors/Participant

Participant Number	Total Number of Corrections

Participant Number	Total Number of Corrections

Participant Number	Total Number of Corrections

Participant Number	Total Number of Corrections

Appendix I: Number of Prompts Needed to Complete IBRST

Participant Number	Session Number	Number of Prompts

Participant Number	Session Number	Number of Prompts

Participant Number	Session Number	Number of Prompts

Participant Number	Session Number	Number of Prompts

Participant Number	Session Number	Number of Prompts

Participant Number	Session Number	Number of Prompts

Participant Number	Session Number	Number of Prompts

Participant Number	Session Number	Number of Prompts

Participant Number	Session Number	Number of Prompts

Participant Number	Session Number	Number of Prompts

Participant Number	Session Number	Number of Prompts

Appendix J: IRB Approval Letter

USF
UNIVERSITY OF
SOUTH FLORIDA

RESEARCH INTEGRITY AND COMPLIANCE
Institutional Review Boards, FWA No. 00001669
12901 Bruce B. Downs Blvd., MDC035 • Tampa, FL 33612-4799
(813) 974-5638 • FAX (813) 974-7091

November 19, 2018

Jessica Moore, M. A.
ABA-Applied Behavior Analysis
Tampa, FL 33612

RE: **Expedited Approval for Initial Review**
IRB#: Pro00037097
Title: An Evaluation of the Individualized Behavior Rating Scale Tool (IBRST) in Inclusive Classroom Settings

Study Approval Period: 11/15/2018 to 11/15/2019

Dear Ms. Moore:

On 11/15/2018, the Institutional Review Board (IRB) reviewed and **APPROVED** the above application and all documents contained within, including those outlined below.

Approved Item(s):
Protocol Document(s):
IRBProtocol102318b.docx

Consent/Assent Document(s)*:
Child Written Assent V1 111518.pdf
Parental Permission V1 111518.pdf
ChildVerbalAssentVersion1.docx

*Please use only the official IRB stamped informed consent/assent document(s) found under the "Attachments" tab. Please note, these consent/assent documents are valid until the consent document is amended and approved. Verbal forms are not stamped.

It was the determination of the IRB that your study qualified for expedited review which includes activities that (1) present no more than minimal risk to human subjects, and (2) involve only procedures listed in one or more of the categories outlined below. The IRB may review research through the expedited review procedure authorized by 45CFR46.110. The research

proposed in this study is categorized under the following expedited review category:

(7) Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

This research involving children as participants was approved under 45 CFR 46.404: Research not involving greater than minimal risk to children is presented.

Requirements for Assent and/or Permission by Parents or Guardians: 45 CFR 46.408: Permission of one parent is sufficient.

Assent is required of all children.

As the principal investigator of this study, it is your responsibility to conduct this study in accordance with IRB policies and procedures and as approved by the IRB. Any changes to the approved research must be submitted to the IRB for review and approval via an amendment. Additionally, all unanticipated problems must be reported to the USF IRB within five (5) business days.

We appreciate your dedication to the ethical conduct of human subject research at the University of South Florida and your continued commitment to human research protections. If you have any questions regarding this matter, please call 813-974-5638.

Sincerely,

Kristen Salomon, Ph.D., Chairperson
USF Institutional Review Board