St. Cloud State University

theRepository at St. Cloud State

Culminating Projects in Community Psychology, Counseling and Family Therapy

Department of Community Psychology, Counseling and Family Therapy

12-2020

An Evaluation of Estimation Data Collection when Measuring Problem Behavior in a Classroom Setting

Amanda Griffin
amandagriffinap@gmail.com

An Evaluation of Estimation Data Collection when Measuring Problem Behavior in a Classroom Setting

by

Amanda M. Griffin

A Thesis

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in Applied Behavior Analysis

December, 2020

Thesis Committee:
Kimberly Schulze, Chairperson
Benjamin Witts
Odessa Luna

Abstract

The purpose of this study was to extend previous research comparing continuous data measurement to estimation data by comparing the two when measuring levels of problem behaviors in an applied group setting with three individuals diagnosed with autism spectrum disorder. Many previous research studies have evaluated and confirmed the accuracy of momentary time-sampling data collection as compared to partial-interval recording for measuring problem behavior; however, to date, no research studies have evaluated the use of estimation data recording for measuring problem behavior. Estimation data was analyzed further by comparing it to data attained from common (in practice) discontinuous data (i.e., PIR and MTS) intervals. The results indicated that, as compared to continuous measurement, estimation data was moderately accurate and most closely aligned with 5s PIR data.

Keywords: autism spectrum disorder, estimation, data collection, classroom, continuous measurement

Table of Contents

List of Tables

List of Figures

Figure Page

Chapter 1: Introduction and Literature Review

The types of behaviors for which data are collected in a classroom setting varies, but frequently involve the monitoring of problem behaviors (also often referred to as aberrant behaviors). Problem behaviors often create social and educational barriers (Horner, 2000) and interfere with an individual's ability to acquire new skills and demonstrate previously established ones (Gardenier et al., 2004). Therefore, treatment for these individuals often targets the reduction of problem behaviors. For children diagnosed with autism spectrum disorder (ASD), problem behaviors may include aggression, self-injury, tantrums, inattention, repetitive behaviors, stereotypy, and social withdrawal (Horner, 2000; Leekam et al., 2011; McCracken et al., 2002). As with all target behaviors, precise measurement is vital to monitoring treatment efficacy. In terms of accuracy, continuous data collection (e.g., duration, frequency) is generally regarded as the most accurate methodology; however, it is also the most impractical method due to its requirement of a dedicated observer. Because of the disadvantages and cumbersome nature of continuous data collection, more practical data collection methods have been developed, such as time-sampling.

Two commonly used time-sampling methods are partial-interval recording (PIR) and momentary time-sampling (MTS). When using PIR data collection methods, an observer records whether the target behavior occurred at any time during the interval. An observer using MTS, records whether the target behavior is occurring at the moment that each interval ends. For both of these methods, data are generally reported as the percentage of intervals in which the behavior occurred, which is then used to estimate the proportion of the observation period that the target behavior occurred.

Many studies have evaluated the accuracy of MTS and PIR data collection methods.

Research has demonstrated that PIR overestimates behavior compared to the actual duration (Gardenier et al., 2004; Murphy & Goodall, 1980; Powell et al., 1977; Radley et al., 2015) and MTS has been determined to both underestimate and overestimate duration, though the margin of error is smaller as compared to PIR (Gardenier et al., 2004; Meany-Daboul et al., 2007; Powell et al., 1977; Radley et al., 2015). Contrastingly, studies comparing MTS and PIR to frequency recording have found MTS to be either no more (i.e., Harrop & Daniels, 1986) or less accurate (i.e., Meany-Daboul et al., 2007) as compared to PIR. Based on these findings, practitioners and researchers have been urged to use MTS, "as it is more accurate and representative of actual levels of behavior" (Radley et al., 2015, p. 377).

As stated by Cooper et al. (2007), a "major advantage of momentary time sampling is that the observer does not does not have to attend continuously to measurement, whereas interval recording methods demand the undivided attention of the observer" (p. 93). For practitioners working in applied treatment settings, this makes data collection using MTS procedures seem less cumbersome as well as making a dedicated observer unnecessary; however, this statement is somewhat unsound. Though Meany-Daboul and colleagues (2007) mentioned that therapists are often responsible for collecting data on more than one individual or target behavior simultaneously, their study used 10-s MTS intervals. In an applied treatment setting, little can be accomplished in 10 seconds. In the other aforementioned studies, similar short MTS intervals of 10 s (i.e., Gardenier et al., 2004; Murphy & Goodall; 1980; Powell et al., 1977; Radley et al., 2015), 15 s (i.e., Harrop & Daniels, 1986; Radley et al. 2015), 20 s (i.e., Gardenier et al., 2004), or 30 s (i.e., Gardenier et al., 2004) were evaluated. Such short intervals essentially render the role of the data collector as a dedicated observer, especially if s/he is collecting data on multiple target behaviors or individuals simultaneously, as is often necessary in classroom and other group settings. Aware of this issue (e.g., the need for interventionists to be able to teach and take data simultaneously), Gunter, et al. (2003) evaluated the accuracy of MTS data collected using 2-, 4-, and 6-min intervals as compared to continuous recording. The authors found that the 4- and 6- min interval time samples produced data paths that significantly differed from continuous measurement data paths, yet the 2-min interval time sample data paths, though slightly higher, showed a high degree of correspondence. Though these findings seem promising for the practical use of MTS in a group or classroom setting, the authors cautioned that further research is needed.

Recognizing that smaller intervals tend to be more accurate, but are often not used in everyday practice, a later study by LeBlanc et al. (2019) examined the most common intervals used by a sizable group of data collectors (using the data collection tool, Catalyst). They determined that for PIR, 5 min, 2 min, 1 min, and 15 min intervals were the first, second, third, and fourth most commonly used intervals, respectively. They found that 2 min, 5 min, 1 min, and 30 s intervals were the first through fourth most commonly used MTS intervals, respectively. Based on various correspondence measures, they concluded that intervals of 3 minutes or less reflected the greatest correspondence, with MTS generally producing greater correspondence compared to PIR. By converting continuous measures (i.e., duration) to discontinuous measures (i.e., PIR and MTS), they also found that sessions with greater amounts of problem behavior resulted in a higher correlation between the two measures (i.e., continuous and discontinuous) and lower correspondence was associated with lower amounts of problem behavior. Overall, compared to Gunter and colleagues' 2003 study, they found that longer intervals (e.g., 3 min) could produce high correspondence with continuous data, but that very low levels of problem behavior produced inadequate correspondence with continuous measurement, even when intervals as small as 10 s were used. So, while this study showed that greater intervals could be accurately used to reflect rates of problem behavior, time-sampling was not suitable for behaviors that occurred at low rates.

Estimation is another, yet less commonly used, method of data collection. This data collection method involves the teacher retrospectively estimating the participant's performance at the end of the teaching session without the need to record data after every teaching trial or after set periods of time (Leaf et al., 2008). When using estimation data collection, the participant's behavior is categorized into a numerical scale ranging from 0 to 4, with 0 typically representing 0-20% (e.g., independent correct responding), 1 representing 21-40%, 2 representing 41-60%, 3 representing 61-80%, and 4 representing 81-100% (Ferguson et al., 2020b, Ferguson et al., 2020a). So far, estimation data collection has only been evaluated during the implementation of discrete trial teaching (DTT), a commonly implemented early intervention method of instruction for individuals diagnosed with ASD (Ferguson et al., 2020b; Ferguson et al., 2020a; Taubman et al., 2013). Though, currently there are only three research studies have evaluated estimation data collection, findings show promise for its use, especially in group settings (e.g., social skills group, classroom).

Taubman and colleagues (2013) compared continuous trial by trial data collection, MTS data collection, and estimation data collection when implementing DTT for children diagnosed with ASD. Their study evaluated each data collection method in accuracy, efficiency, and satisfaction measures. The authors found that continuous recording was the most accurate, time sampling less accurate, and estimation the least accurate. In regards to efficiency, findings were the opposite, with estimation data collection the most efficient in terms number of trials delivered by the interventionist during a teaching session and the fewest number of teaching sessions required before meeting mastery criterion compared to MTS and trial-by-trial data collection. Satisfaction measures revealed that continuous recording was rated as the most accurate and preferred method of data collection for the behavior therapists and the therapists also believed that estimation data was likely the most efficient.

Ferguson and colleagues (2020b) sought to replicate and expand on Taubman and colleagues' (2013) research by comparing trial-by-trial data collection to estimation data collection when implementing DTT for children diagnosed with ASD by evaluating each method's accuracy and efficiency. The authors found that estimation data accurately matched the continuous data and that both data collection methods were equally efficient in terms of the number of trials delivered and the number of teaching sessions to reach mastery criteria. Ferguson and colleagues (in print), then extended their previous study by comparing trial-by-trial data to estimation data collection of a group (3 participants with ASD) during a discrete trial teaching format. In this study, they found that the overall accuracy of estimation data across all 3 teaching formats was 77%, estimation data was more accurate in identifying mastery, led to more teaching trials, and was favored by the interventionists. Though the overall accuracy of estimation data in this third study was lower as compared to previous studies, it was the first to evaluate the use of estimation data in a group setting and showed that estimation data adequately identified when a skill was mastered. The authors of all three studies emphasized the need to attempt to replicate their findings across other behaviors and a broader range of behavior analytic procedures.

Research has demonstrated that ABA-based group instructional formats can be an effective method for teaching children with ASD a wide variety of skills, including choral responding during songs and pre-math skills (Taubman, et al., 2001), sight words/phrases (Ledford et al., 2008), emotions (Leaf et al., 2011), observational learning skills (Leaf et al., 2013), and social skills (Leaf et al., 2017). For children diagnosed with ASD, there are many advantages of ABA-based group instruction in a school setting, some of which include exposure to a natural learning environment, social opportunities, observational learning opportunities, and skill and behavior generalization (Leaf et al., 2008). As with all quality ABA-based treatment, data collection is a fundamental component for monitoring progress. However, collecting data on multiple individuals and/or behaviors in a classroom setting can be overwhelming. Though data collection is generally regarded as important in an educational setting, a study by Sandall, et al. (2004) found that teachers have many perceived barriers when it comes to data collection in the classroom. Some of these perceived barriers were the inability to simultaneously collect data without disrupting teaching, difficulty managing data for multiple children, and a lack of time. The characteristics of estimation data collection may mitigate the perceived barriers associated with collecting data in a classroom setting. Research studies have evaluated the use of MTS data collection methods with problem behaviors in a group setting (Saudargas & Zanolli, 1990; Radley et al., 2015), but, to date, no studies have evaluated the use of estimation data collection with problem behaviors nor in a group setting.

The purpose of this study was to expand on the findings of Taubman et al. (2013), Ferguson et al. (2020b), and Ferguson et al. (2020a) by evaluating estimation data recording for measuring problem behavior in a classroom setting for children with ASD. This study evaluated the accuracy of estimation data recording as compared to continuous measurement. It also compared estimation data to data obtained with commonly used, as determined by LeBlanc and colleagues (2019), MTS and PIR intervals.

Chapter II: Method

Child Participants

Three children diagnosed with autism spectrum disorder (ASD) served as participants in this study. Brian was a 7-year-old boy with a Vineland-3 (Sparrow et al., 2016) adaptive behavior composite score of 70 and a Wechsler Intelligence Scale for Children (WISC-V; Wechsler, 2014) IQ score of 119. Brian enjoyed talking and learning about numbers, phonics, and grammar. He would often stare at and read academic material posted around the classroom (e.g., math posters, calendar) throughout the day and sometimes engaged in hand flapping and or body tensing when looking at those materials. Daniel was a 6-year-old boy with a Vineland-3 adaptive behavior composite score of 75 and a WISC-V IQ score of 65. Throughout his school day, sustained attending (e.g., looking, responding) was a key behavioral target for Daniel, as he often gazed and engaged in self-talk during lessons and downtime in the classroom and regularly required additional 1:1 teaching to learn and understand new concepts. Shawn was a 6-year-old boy with a Vineland-3 adaptive behavior composite score of 85 and a WISC-V IQ score of 109. Shawn engaged in a variety of stereotypic behaviors including scripting, body rocking, body tensing, picking, eye gazing, and tapping materials at moderate to high rates throughout the school day that often averted his attention away from lessons. At the time of the study, Brian, Daniel, and Shawn were in a kindergarten/first grade ABA-based classroom for 1 month when the study began.

Problem Behavior

Prior to beginning data collection, a problem behavior that negatively impacted ability to learn in a group setting was identified and operationally defined for each student.

Gazing. Gazing was the problem behavior selected for Brian and Daniel. Gazing was defined as eyes looking away from the speaker (e.g., teacher, student) and/or relevant teaching materials. Examples of gazing included looking at posters on the wall, out the window, etc. while the teacher was presenting material on the document camera, looking at his pencil or fingers while teacher is talking, and staring at the calendar when the teacher was reading a book. This definition also took into account times that would be okay for the student to look away; therefore, non-examples included looking away during downtime (e.g., while the teacher was passing out supplies to other students, while the teacher was providing behavioral feedback not relative to the lesson to another student, while waiting for other students to complete their independent work), looking at a person entering the classroom, looking at a loud noise (e.g., books dropped in the back of the classroom, phone ringing), and looking his own fingers when asked to hold up a certain number.

Hand Stereotypy. The problem behavior selected for Shawn was engaging in hand stereotypy, with or without materials. Hand stereotypy without materials was defined as picking (fingernails contacting skin, other fingernails, clothes, or carpet with fingers curled or in a curving motion), pinching skin or clothing, rubbing skin or clothing, tensing hands/fingers, and non-functional hand movements. Examples of hand stereotypy included rubbing the carpet, picking at fingers or fingernails, quickly moving hand across line of sight in the air, hands squishing face, and pulling on legs or shoes. Non-examples of hand stereotypy with materials included scratching, hands folded on top of lap, hands/fingers rested on top of legs, and hands resting still on the carpet. Hand stereotypy with materials was defined as the non-functional manipulation of objects. Examples included moving an eraser across his line of sight in the air, erasing table or paper when not required, and rubbing a pencil or crayons on face or neck.

Teachers

Three teachers, Jane, Beverly, and Alice, served as lead teachers during the research sessions. Additionally, two teachers, Jack and Martha, served as support staff during the research sessions. Each session consisted of one teacher serving as the lead and one teacher serving as support staff to assist the children and lead teacher, as needed (e.g., prompt, put token on, provide feedback). The teacher that participated in each session was determined based on his or her schedule (i.e., who taught in the classroom that day). Jane was a 30-year-old female with a master's degree in Education and 7 years of experience providing intervention based on the principles of behavior analysis for individuals diagnosed with ASD. Beverly was a 33-year-old female with a bachelor's degree in Sociology and 9 years of experience providing intervention based on the principles of behavior analysis for individuals diagnosed with ASD. Alice was a 27-year-old female with a bachelor's degree in Child and Adolescent Studies and 5 years of experience providing intervention based on the principles of behavior analysis for individuals diagnosed with ASD. Jack was a 34-year-old male with a bachelor's degree in Psychology and 11 years of experience providing intervention based on the principles of behavior analysis for individuals diagnosed with ASD. Martha was a 25-year-old female with a bachelor's degree in Psychology and 2 years of experience providing intervention based on the principles of behavior analysis for individuals diagnosed with ASD. Four of the five interventionists (Martha excluded) had extensive training and experience leading social skills groups and group teaching for individuals with ASD. While all of the teachers also had extensive training on various data collection procedures, including estimation, they did not collect any data nor give any input on data recordings throughout this study.

Setting and Materials

All research sessions took place on a private school's campus in Southern California inside a classroom that provides group and individual intervention based on the principles of behavior analysis for children with ASD. The classroom had multiple areas including a large group carpet area, individual student desk area, small group center, an indoor play area, and a teacher's desk. The classroom consisted of 5 children diagnosed with ASD. Only 3 of the 5 children served as participants in this study, but all 5 children participated in the lessons that took place during sessions.

In the classroom, lessons provided by the lead teacher covered topics such as reading, writing, social skills, math, etc. throughout the day. The locations in which these lessons were taught varied (e.g., while children are seated at their desks, while children are seated on a carpet on the floor). Lesson materials also varied and consisted of a document projector, individual white boards, worksheets, pencils, crayons, dry erase markers, math manipulatives, etc.

A variety of individualized and group behavior management systems were used in the classroom. The teachers predominantly used a level system (Cihon et al., 2019) with Brian and Daniel. With this system, at any point throughout the day (e.g., at the teacher's discretion), the teacher could tell the student(s) that it was time to "check in," which meant that they would look at which of the seven colors on the chart that student was on and, depending on his color, the student was granted a reinforcement break (differential and increasingly preferred the higher on the chart he was) or some sort of differential negative consequence (e.g., back to work, miss out on a fun activity). The teachers predominantly used a response-cost token board with Shawn in which he earned a differential consequence depending on what color (same as the level chart) he had left on his board at the end of completing 10 tokens. Shawn earned tokens for target behaviors such as sitting with his hands on his lap, looking at the teacher, having a quiet mouth, and responding to choral instructions. He had a color removed (cost) for engaging in stereotypic behavior. During the brief research sessions, the participants were not sent away to access tangible reinforcers (e.g., to the play area) in order to maintain a consistent and uninterrupted observation interval, but may have been presented with social reinforcement (e.g., high five, group cheer) throughout at the lead and/or support teacher's discretion.

Two cameras were set up to record the entirety of each session. One camera was placed directly next to the observer and the other was placed off to the right of the students to allow for an alternate angle, if needed, when reviewing recordings for continuous, PIR, and MTS data collection. Additionally, a digital timer was placed on the side of a desk that was within clear view of the observer, the lead teacher, the support teacher, and both cameras.

Data Collection

Data Collector

The author served as the data collector in this study. At the time of the study, she had a bachelor's degree in Neuroscience and over 7 years of experience implementing behavior analytic interventions with individuals diagnosed with ASD. She also had extensive experience and training on various data collection procedures, including PIR, MTS, and estimation data collection.

Estimation Data

Estimation data were collected using a numerical scale ranging from 0 to 4, which reflected the observer's judgement of the percentage of the observation in which that individual child engaged in the problem behavior. The number 0 represented the presence of the problem behavior 0-20% of the observation period, the number 1 represented the presence of the problem behavior 21-40% of the observation period, the number 2 represented the presence of the problem behavior 41-60% of the observation period, the number 3 represented the presence of the problem behavior 61-80% of the observation period, and the number 4 represented the presence of the problem behavior 80-100% of the observation period. The observer recorded her estimation data (i.e., circled the corresponding number) for each problem behavior across all 3 participants after the session concluded; therefore, no data were collected during the session.

Dependent Variables and Measurement

Accuracy of Estimation Data

Compared to Continuous Data. The accuracy of estimation data were measured by comparing the observer's estimation measurement to the observer's continuous (duration) measurement collected post hoc. The observer collected continuous data using the stopwatch feature on an iPhone while watching the video recordings of the research sessions. The observer started the stopwatch anytime the participant began engaging in their corresponding problem behavior (i.e., gazing or hand stereotypy) and stopped the stopwatch when the participant stopped engaging in their corresponding problem behavior. The total amount of time that the participant engaged in problem behavior was then converted to seconds (to the tenth of a second) and then converted to a percentage of time by dividing it by the total time of the observation period and multiplying by 100%. For example, if the participant engaged in problem behavior 323.6s, and the total observation time was 596.4 s, the recorded percentage of time engaged in problem behavior was 55%.

The estimation data collected for each student during each session was then categorized as accurate or inaccurate, as compared to the continuous data. Estimation data was considered accurate if the continuous data percentage was within the estimation rating range (e.g., continuous data from videotapes was 45% and the estimation rating was a 2). Estimation data was considered inaccurate if the continuous data percentage was not within the estimation rating range (e.g., continuous data from videotapes was 45% and the estimation rating was a 3).

Compared to PIR and MTS Data. Additionally, 5s PIR data and 5s MTS data were collected post hoc for further analysis of the accuracy of estimation data. The observer collected PIR data by watching the videotaped sessions and recorded a "+" if the student engaged in the problem behavior at any point during the 5s interval (according to the timer that was visible in the video) or a "-" if the student did not engage in problem behavior at any point during the 5s interval. The observer collected MTS data by watching the videotaped sessions and recorded a "+" if the student was engaging in the problem behavior at the end of the 5s interval (according to the timer that was visible in the video) or a "-" if the student was not engaging in the problem behavior at the end of the 5s interval. The observer was permitted to pause, rewind, and replay the video, as needed, for each interval scored for each student. The 5s interval PIR and MTS data was then used to obtain 30s, 60s, and 120s PIR and MTS intervals, as these are the intervals most commonly used in practice according to LeBlanc et al.'s (2019) study (with the exception of 5-minute intervals, as this study would only reflect two 5-minute intervals).

The number of intervals that the participant engaged in problem behaviors was then converted to an estimated percentage of time by taking the total number of "+" intervals divided by the total number of intervals (120 5-second intervals, 20 30-second intervals, 10 60-second intervals, and 5 120-second intervals) and multiplying by 100%. For example, if the participant was recorded to engage in the problem behavior 67 out of a possible 120 intervals, then the recorded percentage of time engaged in problem behavior was 56%.

The estimation data collected for each student during each session was then categorized as accurate or inaccurate, as compared to each time-sample data. Estimation data was considered accurate if the time-sample data percentage was within the estimation rating range (e.g., PIR data from videotapes was 45% and the estimation rating was a 2). Estimation data was considered inaccurate if the time-sample data percentage was not within the estimation rating range (e.g., PIR data from videotapes was 45% and the estimation rating was a 3).

The correspondence between the discontinuous measures (PIR and MTS) and continuous measures were calculated as correlation coefficients. Correlation coefficients were calculated using the CORREL function of Excel, comparing the percentage duration measure to the percentage of intervals measure. Scores closer to 1.0 indicate a greater correspondence between the measures. These coefficients were used to assess the validity of using the time-sample data to evaluate the accuracy of the estimation data (measures with a low correspondence score are not a valid comparisons).

Intervention

General Procedure. As previously mentioned, research sessions consisted of one lead teacher and one support staff. In order to increase the likelihood that there would be variance in the amount of problem behavior that each student engaged in during research sessions (i.e., in an attempt to minimize the chance that students engaged in the same amount of problem behavior across sessions and therefore possibly make their estimation scores predictable for the observer), half of the research sessions took place during the morning (math) and half took place during the afternoon (phonics). This study consisted of a total of 10 sessions. The website Random.org was used prior to beginning the study to determine the random order in which the research sessions took place either during math or phonics and to ensure that 5 sessions occurred during math and 5 sessions occurred during phonics. Each session took place during the first 10 minutes of each subject's lesson, approximately 9:00-9:10 a.m. for math and approximately 1:00-1:10 p.m. for phonics. The lessons were teacher-led, involved independent work, or a combination of both. Math lessons took place while students were seated on the large group carpet at the front of the classroom and phonics lessons took place while the students were seated at their desks.

Throughout the observation period, the observer sat in a chair at the front of the classroom with a clipboard, data sheet, and pen. The observer's location allowed for a clear, unobstructed view of all 3 participants, but she was permitted to move, as needed, to ensure that all participants were within sight. She observed the students for the entirety of the session without any distractions (e.g., grading papers, preparing materials) and was not permitted to take any written notes on the students' behaviors at any point during the observation. The observer did not engage with the students nor participate in the lesson at any point throughout the observation period. The students were unaware of what the observer was doing during sessions (i.e., they were not informed that the observer was watching or measuring their behavior).

Each session started with the support teacher pressing start on the timer, which the timer then began to count up like a stopwatch, displaying the number of minutes and seconds (to the tenth of a second) as they passed. The teachers conducted their lesson, as usual, and the session ended once 10 minutes on the timer had passed. Upon close review of the videotaped sessions, it was discovered that the timer used for the research sessions counted slightly faster than other timers (e.g., the one on iPhones that were used for obtaining post hoc continuous data), so though sessions were intended to last 600 s, they actually lasted 596.4 s, which was the number used when calculating the percentage of time each participant engaged in problem behavior for continuous measurement.

At the end of the 10-minute session, the observer recorded her estimation data (i.e., circled the corresponding number) for each problem behavior across all 3 participants after the session (see Appendix) without any discussion with the lead or support teachers prior to recording her score (i.e., it was based solely on her judgement based on her observations).

Interobserver Agreement

A second independent observer recorded post hoc continuous (duration) data, 5s PIR data, and 5s MTS data on participant behavior for 40% of videotaped research sessions. Interobserver agreement (IOA) was calculated for 5s time-sample intervals, and this data were used to then extract 30s, 60s, and 120s time-sample data.

For continuous data, IOA was reported as the mean percentage of difference. Mean percentage of difference scores were calculated as the absolute value of the difference between the observer's duration percentage and a second observer's duration percentage. Scores closer to 0% indicate greater correspondence between the two measures. The mean percentage difference of continuous data for Brian, Daniel, and Shawn was 2.6% (range, 0.7-4.4%), 5.4% (range, 1.3-10.1%), and 7.6% (range, 4.2-10.8%), respectively.

For PIR and MTS data, IOA was calculated by totaling the number of intervals of agreements divided by the number of agreements plus disagreements and multiplying by 100. IOA for PIR data for Brian, Daniel, and Shawn was 80.4% (range, 70.8-93.3%), 81.7% (range, 73.3-93.3%), and 77.1% (range, 71.7-90.8%), respectively. IOA for MTS data for Brian, Daniel, and Shawn was 90.2% (range, 85.0-93.3%), 85.8% (range, 78.3-90.8%), and 79.0% (range, 72.5-84.2%), respectively.

Chapter III: Results

Accuracy of Estimation Data

Compared to Continuous Data

Figure 1 displays the accuracy of the observer's estimation data collected in the classroom as compared to continuous data collected from video recordings of each session. In the figure, sessions denoted with an asterisk indicate that the estimation datum for that session was inaccurate when compared to the continuous (duration) datum and sessions denoted with a double asterisk indicate that the estimation datum for that session was inaccurate, but within 5% of the continuous datum. Of the 30 estimation data recordings across participants, 60% were accurate. Individually, estimation data recordings were accurate 90%, 50%, and 40% of instances for Brian, Daniel, and Shawn, respectively.

Compared to PIR and MTS Data

Table 1 depicts the accuracy of the observer's estimation data collected in the classroom as compared to PIR and MTS data collected from video recordings of each session. In the table, non-highlighted (i.e., white) cells are problem behavior percentages that were accurately estimated, cells highlighted in gray are problem behavior percentages that were inaccurately estimated, but within 5% difference, cells highlighted in black are problem behaviors that were inaccurately estimated by greater than a 5% difference, and an asterisk represents that the estimation data overestimated the time interval percentage. Estimation data were accurate as compared to 5s PIR data for 80% of sessions. Individually, estimation data recordings were accurate 100%, 90%, and 50% of instances for Brian, Daniel, and Shawn, respectively. Estimation data were accurate as compared to 5s MTS data for 60% of sessions. Individually, estimation data recordings were accurate 90%, 50%, and 40% of instances for Brian, Daniel, and

Shawn, respectively. Figure 2 (PIR) and Figure 3 (MTS) display these data in graph form, where sessions denoted with an asterisk indicate that the estimation datum for that session was inaccurate when compared to the interval datum and sessions denoted with a double asterisk indicate that the estimation datum for that session was inaccurate, but within 5% of the interval datum.

The correlations between continuous measurement (i.e., percentage duration) and each discontinuous measurement (i.e., PIR and MTS) are shown in Figure 4. The correlation between continuous data and 5s PIR data was .95, then steadily decreases to .45 with 120s intervals. The correlation between continuous data and 5s MTS data across participants was .97 and decreases to .79 with 120s intervals.

Figure 1

Accuracy of Estimation Data Compared to Continuous Data

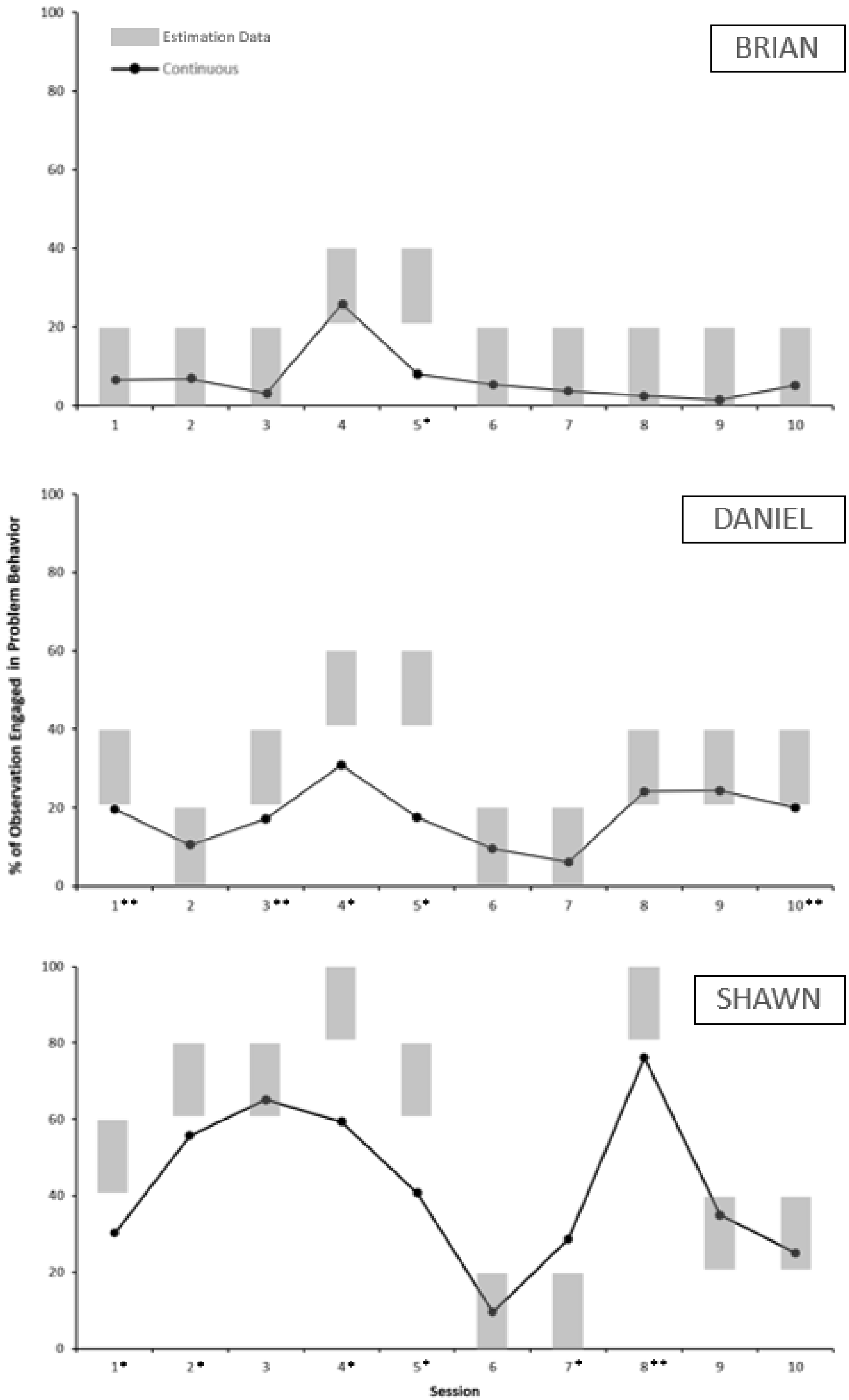


Table 1

Accuracy of Estimation Data as Compared to PIR and MTS

	Session	Continuous	Est.	5s PIR	30s PIR	60s PIR	120s PIR	5s MTS	30s MTS	60s MTS	120s MTS
Brian	1	7%	0-20%	9%	25%	50%	80%	5%	0%	0%	0%
	2	7%	0-20%	9%	25%	40%	60%	6%	5%	0%	0%
	3	3%	0-20%	3%	20%	10%	20%	3%	0%	0%	0%
	4	26%	21-40%	33%	75%	90%	100%	22%	20%	10%*	20%
	5	8%	21-40%	23%	45%	70%	100%	15%*	5%*	10%*	20%
	6	5%	0-20%	8%	25%	50%	80%	5%	5%	0%	0%
	7	4%	0-20%	2%	5%	10%	20%	2%	0%	0%	0%
	8	3%	0-20%	16%	65%	90%	100%	4%	15%	20%	20%
	9	1%	0-20%	12%	40%	50%	80%	7%	0%	0%	0%
	10	5%	0-20%	13%	40%	60%	100%	8%	10%	0%	0%
Daniel	1	20%	21-40%	27%	70%	80%	100%	13%*	20%	10%*	0%*
	2	11%	0-20%	9%	40%	60%	80%	3%	5%	0%	0%
	3	17%	21-40%	22%	53%	78%	100%	8%*	11%*	11%*	25%
	4	31%	41-60%	48%	90%	100%	100%	32%	35%*	40%	40%
	5	18%	41-60%	43%	90%	100%	100%	23%	20%*	10%*	20%*
	6	10%	0-20%	8%	25%	50%	80%	5%	5%	10%	20%
	7	6%	0-20%	7%	25%	40%	80%	6%	0%	0%	0%
	8	24%	21-40%	35%	90%	100%	100%	23%	25%	40%	60%
	9	24%	21-40%	41%	85%	100%	100%	27%	30%	30%	20%
	10	20%	21-40%	28%	75%	100%	100%	18%	20%	10%*	20%
Shawn	1	30%	41-60%	47%	100%	100%	100%	30%*	35%*	10%*	20%*
	2	56%	61-80%	67%	100%	100%	100%	53%*	55%*	60%	80%
	3	65%	61-80%	78%	100%	100%	100%	69%	70%	60%	60%
	4	60%	81-100%	79%	100%	100%	100%	62%*	60%*	70%*	100%
	5	41%	61-80%	73%	100%	100%	100%	53%*	60%	70%	60%
	6	10%	0-20%	23%	65%	90%	100%	18%	15%	10%	20%
	7	29%	0-20%	46%	85%	100%	100%	41%	50%	40%	40%
	8	76%	81-100%	86%	100%	100%	100%	74%*	85%	70%*	40%*
	9	35%	21-40%	46%	80%	100%	100%	32%	35%	40%	20%
	10	25%	21-40%	49%	90%	90%	100%	37%	35%	50%	40%

Figure 2

Accuracy of Estimation Data Compared to PIR

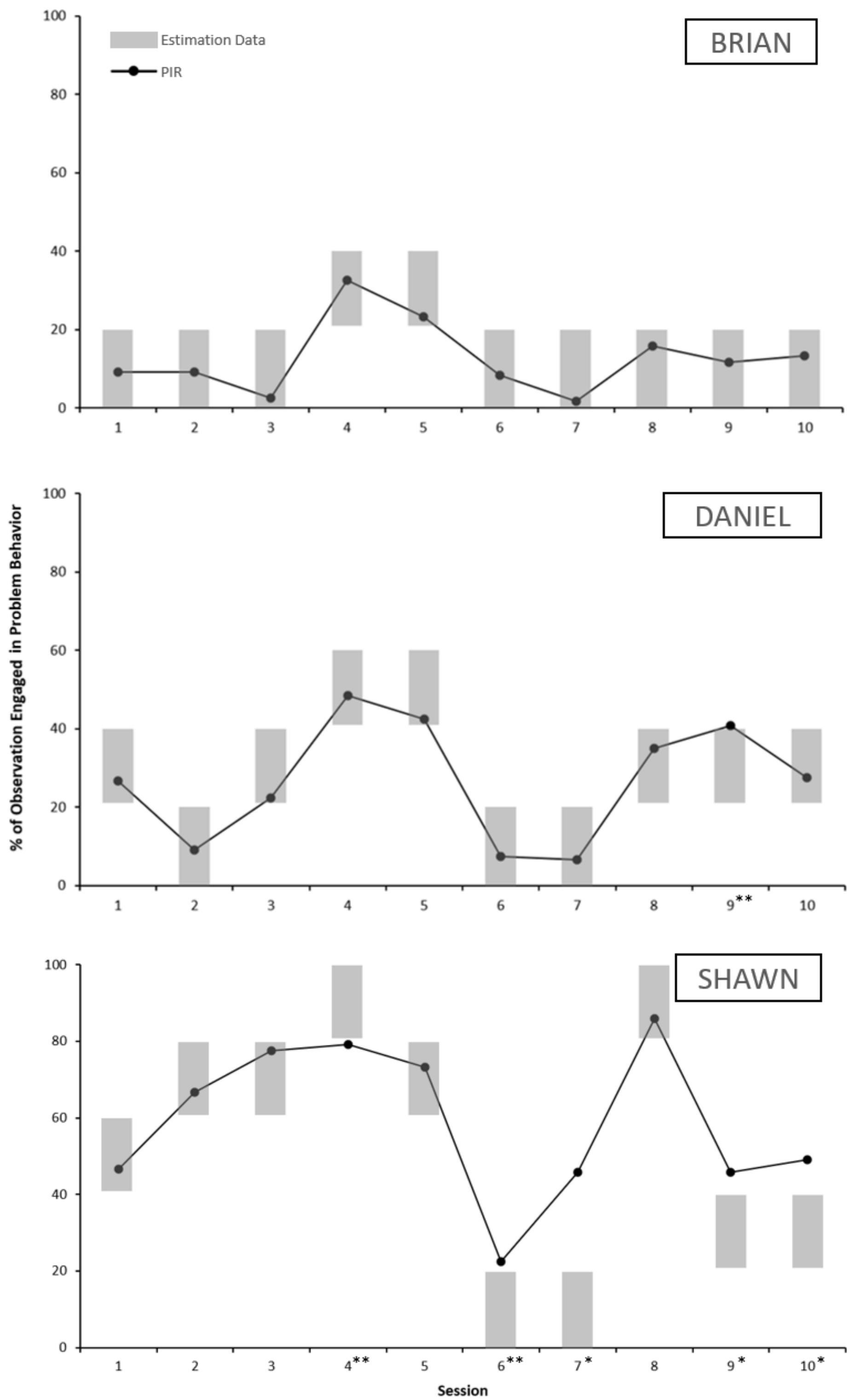


Figure 3

Accuracy of Estimation Data Compared to MTS

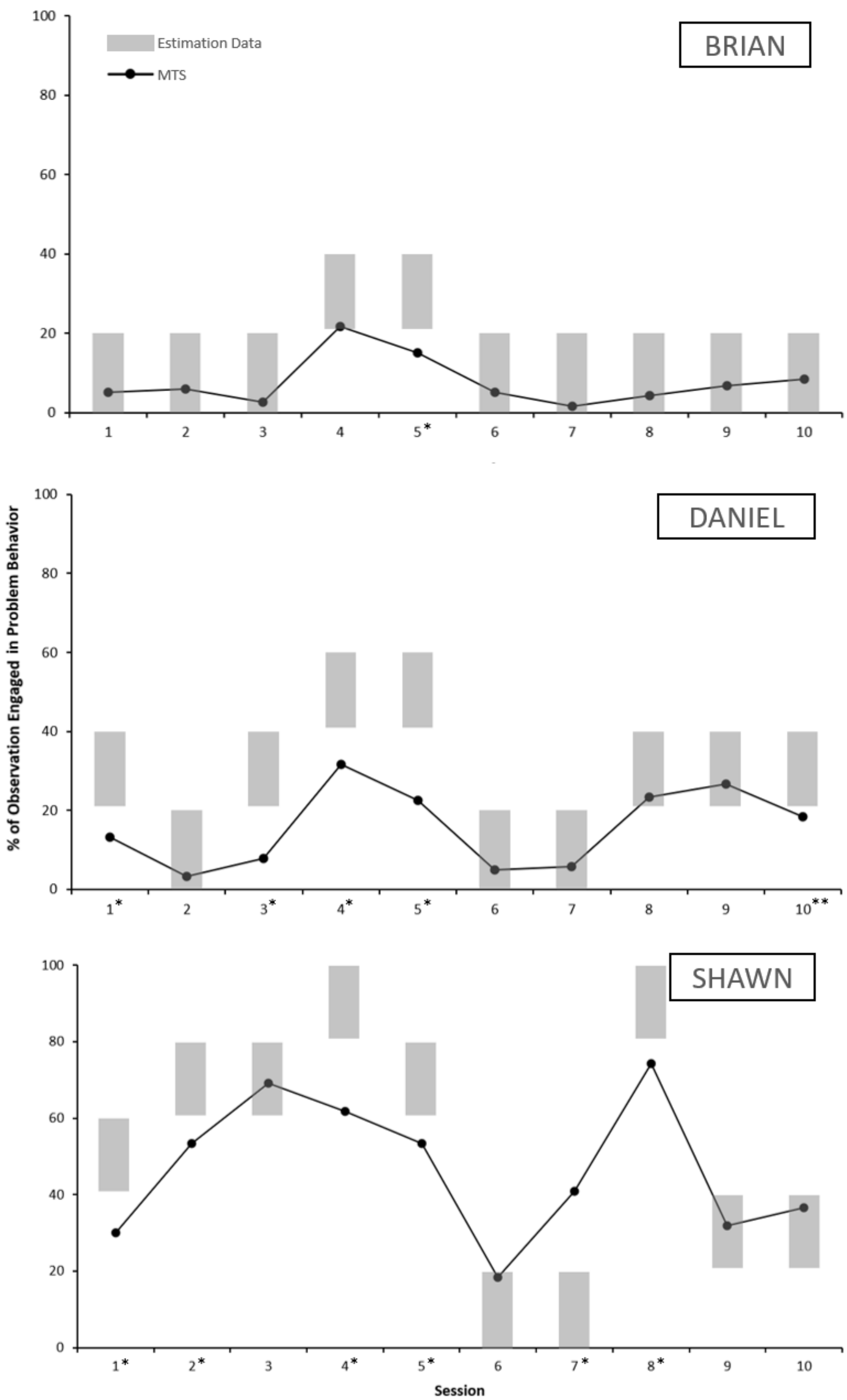
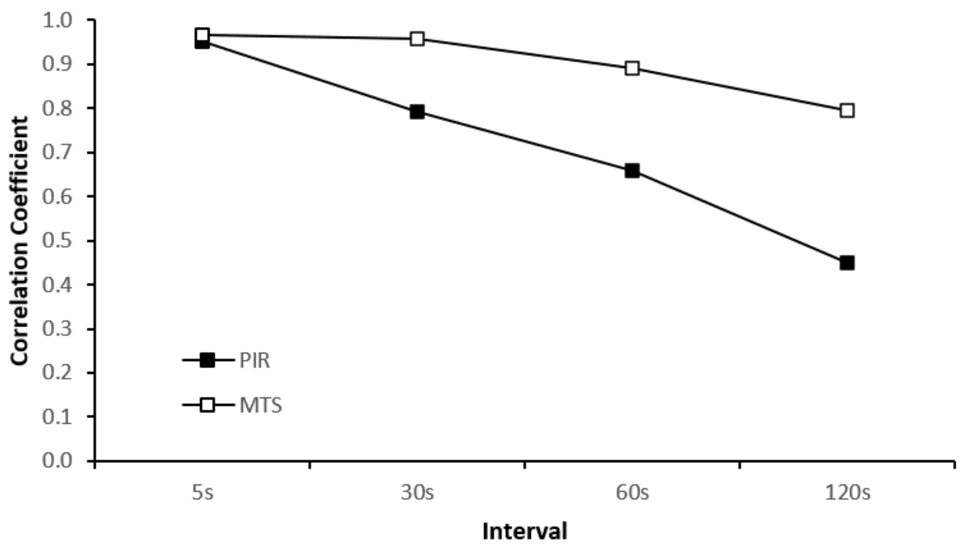


Figure 4

Correlation Coefficients



Chapter IV: Discussion

The purpose of this study was to evaluate the accuracy of estimation data collection as compared to continuous measurement, MTS data, and PIR data, for measuring problem behavior in a classroom setting for children with ASD. Overall, the accuracy of estimation data as compared to continuous measurement was moderate at 60%, with the estimations of an additional 13% of sessions inaccurately capturing the continuous data, but were within 5% of continuous data. When compared to 5s PIR data, which demonstrated a high correlation to the continuous data ($r = .95$), estimation data aligned with 80% of instances. However, when compared to 5s MTS data, which also demonstrated a high correlation to the continuous data ($r = .97$), estimation data aligned with 60% of opportunities. These results differ from previous research in which MTS data has been found to be a more accurate representation of behavior levels when compared to PIR (Radley et al., 2015).

It may also be noteworthy to examine the data based on each individual participant. Estimation data was most accurate for Brian, who displayed relatively low rates of behavior. With the exception of Session 5 (in which the observer overestimated levels of behavior for all participants), the estimation rating increased when Brian's behavior increased. These relative increases and decreases were also observed for Daniel, who displayed relatively moderate levels of behavior. This suggests that estimation data may be beneficial for tracking relative increases and decreases in an individual's level of problem behavior on a day-to-day (or in school even subject-to-subject) basis for individuals with low to moderate levels of behavior.

Shawn's estimation data showed the lowest accuracy. While it is possible that estimation data may not be suitable for high and/or variable levels of problem behavior, other differences may have contributed to this inaccuracy. Compared to the other two participants, not only was the problem behavior tracked differently, but the operational definition of Shawn's problem behavior was complex and encompassed multiple topographies of hand stereotypy (e.g., with and without materials). Future studies should examine whether or not the type of problem behavior recorded affects the accuracy of the estimation data. Additionally, also compared to the other two participants, Shawn displayed a wide variety of other types of problem behavior including body rocking, vocal stereotypy (e.g., non-sensical noises, scripting), and gazing. Therefore, the observer may have been estimating the rate of problem behaviors overall, as her estimations were often greater compared to the continuous data.

There are several limitations and areas for future research. First, this study used a dedicated observer as the recorder of estimation data, which may not be practical in a classroom or applied treatment setting. Future studies should evaluate the accuracy of the lead teacher or therapist's estimation data when measuring problem behavior. If a lead teacher or therapist is able to record accurate estimation data on problem behaviors and teach simultaneously, then data collection would minimally disrupt lessons and would be easy to manage.

Second, the general composition of the participants and classroom create further limitations. The observer in this study was a highly-trained and well-experienced ABA therapist, including with regard to estimation data collection. Additionally, no interobserver agreement data was collected for the observer's estimation rating. Future studies should assess the agreement between multiple individuals recording evaluation data and evaluate the clinical skills and experience level required to make accurate estimation ratings. Furthermore, the classroom size was relatively small and only 3 participants were evaluated at a time. Subsequent studies should look into how group size and the number participants evaluated affects the accuracy of estimation ratings.

Thirdly, only two types of problem behaviors (i.e., eye gazing and hand stereotypy) were measured in this study. Future studies should evaluate the accuracy of estimation data for measuring a wide variety of problem behaviors. Additionally, it should also be determined if the number of problem behaviors measured (e.g., all the same, all different) using estimation data affects its overall accuracy.

Lastly, within this study, treatment decisions were not made based on the data collected. Future research should assess if estimation data collection procedures influence treatment decisions differently than those made based on the data obtained from other collection procedures (e.g., time-sampling).

Continuous data collections have been repeatedly shown to be the most accurate way to measure a learner's behaviors, but are not always feasible in clinical and classroom settings. This study extended previous research on the use of estimation data collection to potentially minimize the difficulties of continuous data collection within clinical practice. Due to the possibility of estimation data relieving many of the constraints imposed on therapists and teachers when using other data collection methods in clinical settings to measure problem behavior, estimation data collection of problem behavior warrants further investigation, including its possibility to be used in conjunction with other data collection methods.

References

Cihon, J.H., Ferguson, J.L., Leaf, J.B., Leaf, R., McEachin, J., & Taubman, M. (2019). Use of a level system with flexible shaping to improve synchronous engagement. *Behavior analysis in practice*, *12*(1), 44-51.

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

Ferguson, J., Griffin, A., Terhune, T., Milne, C., Cihon, J., Majeski, M., ... & Leaf, R. (2020a). *Comparing two data collection procedures when using discrete trial teaching to teach academic targets within a group instructional format* [Unpublished manuscript].

Ferguson, J., Milne, C., Cihon, J., Dotson, A., Leaf, J., McEachin, J., & Leaf, R. (2020b). Comparison of estimation data collection to trial-by-trial data collection during discrete trial teaching. *Behavior Analysis in Practice.*

Gardenier, N. C., MacDonald, R., & Green, G. (2004). Comparison of direct observational methods for measuring stereotypic behavior in children with autism spectrum disorders. *Research in Developmental Disabilities*, *25*(2), 99-118.

Gunter, P. L., Venn, M. L., Patrick, J., Miller, K. A., & Kelly, L. (2003). Efficacy of using momentary time samples to determine on-task behavior of students with emotional/behavioral disorders. *Education and Treatment of Children*, 400-412.

Harrop, A. & Daniels, M. (1986). Methods of time sampling: A reappraisal of momentary time sampling and partial interval recording. *Journal of Applied Behavior Analysis*, *19*(1), 73-77.

Horner, R. H. (2000). Positive behavior supports. *Focus on Autism and other Developmental Disabilities*, *15*(2), 97-105.

Leaf, J. B., Leaf, J. A., Milne, C., Taubman, M., Oppenheim-Leaf, M., Torres, N., ... & Yoder, P. (2017). An evaluation of a behaviorally based social skills group for individuals diagnosed with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *47*(2), 243-259.

Leaf, J. B., Oppenheim-Leaf, M. L., Dotson, W. H., Johnson, V. A., Courtemanche, A. B., Sheldon, J. B., & Sherman, J. A. (2011). Effects of no-no prompting on teaching expressive labeling of facial expressions to children with and without a pervasive developmental disorder. *Education and Training in Autism and Developmental Disabilities*, 186-203.

Leaf, R. B., Taubman, M. T., McEachin, J., & Driscoll, M. (2008). *It's time for school!: Building quality ABA educational programs for students with autism spectrum disorders*. New York, NY: DRL Books.

Leaf, J. B., Tsuji, K. H., Lentell, A. E., Dale, S. E., Kassardjian, A., Taubman, M., ... & Oppenheim-Leaf, M. L. (2013). A comparison of discrete trial teaching implemented in a one-to-one instructional format and in a group instructional format. *Behavioral Interventions*, *28*(1), 82-106.

LeBlanc, L. A., Lund, C., Kooken, C., Lund, J. B., & Fisher, W. W. (2019). Procedures and accuracy of discontinuous measurement of problem behavior in common practice of applied behavior analysis. *Behavior Analysis in Practice*, 1-10.

Ledford, J. R., Gast, D. L., Luscre, D., & Ayres, K. M. (2008). Observational and incidental learning by children with autism during small group instruction. *Journal of Autism and Developmental Disorders*, *38*(1), 86.

Leekam, S. R., Prior, M. R., & Uljarevic, M. (2011). Restricted and repetitive behaviors in autism spectrum disorders: a review of research in the last decade. *Psychological Bulletin*, *137*(4), 562.

McCracken, J. T., McGough, J., Shah, B., Cronin, P., Hong, D., Aman, M. G., ... & McDougle, C. J. (2002). Risperidone in children with autism and serious behavioral problems. *New England Journal of Medicine*, *347*(5), 314-321.

Meany-Daboul, M. G., Roscoe, E. M., Bourret, J. C., & Ahearn, W. H. (2007). A comparison of momentary time sampling and partial-interval recording for evaluating functional relations. *Journal of Applied Behavior Analysis*, *40*(3), 501-514.

Murphy, G. & Goodall, E. (1980). Measurement error in direct observations: A comparison of common recording methods. *Behaviour Research and Therapy*, *18*(2), 147-150.

Powell, J., Martindale, B., Kulp, S., Martindale, A., & Bauman, R. (1977). Taking a closer look: Time sampling and measurement error. *Journal of Applied Behavior Analysis*, *10*(2), 325–332.

Radley, K. C., O'Handley, R. D., & Labrot, Z. C. (2015). A comparison of momentary time sampling and partial-interval recording for assessment of effects of social skills training. *Psychology in the Schools*, *52*(4), 363-378.

Sandall, S. R., Schwartz, I. S., & Lacroix, B. (2004). Interventionists' perspectives about data collection in integrated early childhood classrooms. *Journal of Early Intervention*, *26*(3), 161-174.

Saudargas, R. A., & Zanolli, K. (1990). Momentary time sampling as an estimate of percentage time: A field validation. *Journal of Applied Behavior Analysis*, *23*(4), 533-537.

Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). *Vineland-3: Vineland adaptive behavior scales*. PsychCorp.

Taubman, M., Brierley, S., Wishner, J., Baker, D., McEachin, J., & Leaf, R. B. (2001). The effectiveness of a group discrete trial instructional approach for preschoolers with developmental disabilities. *Research in Developmental Disabilities*, *22*(3), 205-219.

Taubman, M., Leaf, R., McEachin, J., Papovich, S., & Leaf, J. (2013). A comparison of data collection techniques used with discrete trial teaching. *Research in Autism Spectrum Disorders*, *7*(9), 1026-1034.

Wechsler, D. (2014). Wechsler intelligence scale for children–Fifth Edition (WISC-V). *Bloomington, MN: Pearson*.

Appendix A

Estimation Data Collection Form

Staff: ______________________ CARPET/DESKS Date: ______________

Instructions: Upon the conclusion of the teaching session, estimate the percentage of the observation period that that the participant's problem behavior occurred throughout the session. No other notes or data should be collected during the observation time.

0 = Behavior occurred 0-20% of the observation period
1 = Behavior occurred 21-40% of the observation period
2 = Behavior occurred 41-60% of the observation period
3 = Behavior occurred 61-80% of the observation period
4 = Behavior occurred 81-100% of the observation period

Participant:	Brian				
Behavior:	GAZING: Eyes looking away from the speaker (e.g., teacher or student) and/or relevant teaching materials/events				
Estimation:	0	1	2	3	4

Participant:	Daniel				
Behavior:	GAZING: Eyes looking away from the speaker (e.g., teacher or student) and/or relevant teaching materials/events				
Estimation:	0	1	2	3	4

Participant:	Shawn				
Behavior:	Hand SSB (with or without materials): picking (fingernails contacting skin, other fingernails, clothes or carpet with fingers curled or curving motion), pinching, tensing hands/fingers, hands buried/hidden in lap, non-functional hand movements, non-functional manipulation of materials				
Estimation:	0	1	2	3	4