5-12-2022

Analyzing the validity of self-reports of emotional responding using an electronic device

Tiffany E. Shiflet
James Madison University

Analyzing the Validity of Self-Reports of Emotional Responding Using an Electronic Device

Tiffany E. Shiflet

A thesis submitted to the Graduate Faculty of

JAMES MADISON UNIVERSITY

In

Partial Fulfillment of the Requirements

for the degree of

Master of Arts in Psychological Sciences

Applied Behavior Analysis Concentration

Department of Graduate Psychology

May 2022

FACULTY COMMITTEE:

Committee Chair: Daniel D. Holt, Ph.D.

Committee Members/ Readers:

Bryan K. Saville, Ph.D.

Jeff Dyche, Ph.D.

Acknowledgements

I would like to express my gratitude to my advisor and thesis chair, Dr. Daniel Holt for his guidance and feedback throughout this thesis project. I would also like to thank my committee members, Dr. Jeff Dyche and Dr. Bryan Saville, for their feedback on this thesis document throughout this learning experience. My appreciation also goes to Dr. Louise Freeman Davis for her time and contribution as the research assistant for this thesis project. I would also like to thank the outpatient ABA clinic in Harrisonburg, Virginia, for allowing me to conduct this research study in their facility.

Table of Contents

List of Figures

Abstract

The assessment of internal behaviors such as thoughts, feelings, or physiological symptoms not seen by the naked eye are often assessed with indirect measures such as self-reports and questionnaires given the lack of accessibility and observations by outside observers. The self-management of human behaviors, including internal events, carries socially valid implications for an individual's quality of life, including children and individuals with neurological, developmental, and intellectual disabilities. This study aimed to address the following question: are there valid measurement procedures (e.g., collecting data on physiological responses) to analyze the correspondence between self-reports of emotional states and observable and measurable overt behaviors? To address the limitations in measurement validity of private, covert behaviors, the primary researcher of this study conducted a correspondence check of these three measurements with two participants with autism spectrum disorder in the following manner: collected heart rate data of the participants via a wearable device (Fitbit® Charge 5), conducted direct observation and data recording of overt behaviors, and collected data on the participants' self-report of their emotional or physiological state via a 5-point rating scale across several time periods per session. Results of the study showed a strong positive correlation between observer and participant ratings from the 5-point rating scales; however, physical activity of the participants seemed to consistently mask true heart rate readings of the emotional responses. More research can be done on the correspondence of these three variables and their implications for guiding self-management programs of private events.

Introduction

Historical Views of Self-Management and Internal Events

The history and evolution of self-management and human behavior dates to early psychological theories about an individual changing his or her own behavior by means of internal factors such as "willpower," "drive," or "desire" (Cooper et al., 2020, p. 682). Based on early psychological and cognitive-behavioral theories, causal factors for behavior rest with events preceding behavior, including internal factors such as thoughts and feelings. For example, if *A* happened before *B*, then *A* must be the cause for *B*. The terminology used to describe behavioral processes related to an individual's behavior change, particularly the term 'self-control,' have placed an onus on the individual (the "self") as the sole contributor to behavioral events or behavior change. From early psychological theories of self-control or self-management, causal factors for behavior were assumed to be internal for variables not easily observed in the environment (Brigham, 1983; Cooper et al., 2020). In his early experiments with animals, E. L. Thorndike (1898), an American psychologist known best for laying the foundation of educational psychology and learning theory, introduced the Law of Effect, which stated that behaviors followed by pleasant consequences were more likely to occur again in the future in the presence of the same stimulus, whereas behaviors followed by unpleasant consequences were not likely to occur again in the future (Catania, 2013). Thorndike's experimental analysis of animal behavior and his learning theory paved the way for operant conditioning within behaviorism. B. F. Skinner's early work in behaviorism was based in how animal and human behaviors result from conditioning, mainly that operant contingencies of behavior occur as behavior is influenced and maintained by its consequences (defined as stimuli that are added to the environment and/or occur

immediately following the behavior and produce some effect on the future probability of that behavior occurring again). Skinner (1953) outlined operant conditioning within self-management programs and how understanding the environmental contingencies operating within and around the organism are significant to effective behavior change.

There are some limitations in common language used to describe self-control or self-management. One limitation involves fallacies of circular reasoning, which do not provide a credible understanding or explanation for why a behavior occurs. An example of circular reasoning would include the following: Student A completes their homework on time because they are organized; Student A is organized because they complete their homework on time. Based on this circular statement, there is no explanation for how the individual has come to be "organized" or control their organized behavior. Not only is this rationale circular, but it also relates to another limitation involving the difficulty of explaining how one can achieve self-control. Some other ways in which "self-control" may be interpreted may insinuate that all the onus for behavior and behavior change fall on the individual, leaving out other observable and environmental variables contributing to behavior. Understanding the variables involved in the occurrence of behavior (e.g., the event(s) right before a behavior occurs, setting events, maintaining consequences, etc.), allows one to more effectively engage in desired behavior change and control various parts of their environment long-term.

Self-management Defined

According to Cooper et al. (2020), self-management is defined as "…the personal application of behavior change tactics that produces a desired improvement in behavior" (p. 683). More specifically, Brigham (1983) defined self-management as "…the ability of

the individual to interface his or her behavior with the environment…. [through] the application of behavior analysis principles and procedures to modify the behavior/environment interactions *of* the individual *by* the individual" (p. 49). These definitions derive from the operant behavior approach consistent with Skinner's analysis of the behavior-environment relation. The individual whose behavior is being targeted acts in some way to change their own behavior, and controls one or all aspects of the behavior change program. Hughes and Lloyd (1993) provided another take on self-management (or 'self-control') as the ability of an individual to learn to engage in responses that forego immediate access to reinforcement that often results in later contact with reinforcement that is typically of higher quality or value. Acting against those immediate contingencies allows for one to access reinforcement of a different magnitude, quality, or duration by changing when and how one responds in the present moment. These factors contribute to behavior change for a variety of behaviors, including overt (observable, shown openly) and covert (hidden, not openly acknowledged or seen) behaviors.

The Behavior Change Process

Behavior change cannot occur entirely independent of other individuals; to some degree, other external change agents are involved in the process. Cooper et al. (2020) described various roles that other individuals (external change agents) play in the behavior change process of a person, such as how the agent "...manipulates motivating operations, arranges discriminative stimuli, provides response prompts, delivers differential consequences, and observes and records the occurrence or nonoccurrence of the target behavior" (p. 683). For behavior change to be effective, one must attend to and

analyze the various components of the three-term contingency of behavior and the implications of operant behavior. Operant behavior is explained by Sidman (2008) as occurring when a behavior is controlled by its consequences, or more specifically, the event or stimulus change that occurs right after a behavior occurs. Behavior operates under contingencies in which reinforcement has been delivered in the past contingent upon a certain response. The antecedent-behavior-consequence unit can also be described as the stimulus-response-reinforcer paradigm (Sidman, 2008). Environmental variables play key roles in understanding behavior, specifically as they pertain to antecedent manipulations, motivating operations, and delivery of consequences. Understanding how environmental variables impact responses allows for one to manipulate or arrange the various parts of the behavior chain for purposeful behavior change.

Private Events and Radical Behaviorism

Understanding the environmental conditions impacting behavior includes how the behavior analytic community continues to investigate better ways of addressing the contingencies surrounding self-reports and behaviors that are not overt or directly observable to others. Skinner termed these behaviors as events that occur "within one's skin," also known as "private events" (Skinner, 1953, p. 257), which are those events that are not accessible to or observable by others; the only observer is the individual themself. Examples of internalized behaviors include events such as thinking and the labels used within our verbal behavior communities to describe events occurring internally such as physiological responses or thoughts we label as emotions (Skinner, 1953).

In the field of behavior analysis, there are obvious limitations to analyzing and studying private events. Some of these limitations are attributed to the facts that these types of behaviors encompass a wide range of responses and involve broad topics

surrounding internal events, lack of specificity in describing private behavior, and that private events do not often fit the *behavioral* dimension of applied behavior analysis, meaning that these behaviors are not observable and measurable by outside observers (Baer, Wolf, & Risley, 1968). These challenges with accessibility and observation contradict the integrity of analyzing, understanding, and addressing behavior, which presents issues in the validity and reliability of the behavior change procedures for private events.

Despite the challenges with accessing and addressing behaviors categorized as private events, Skinner still recognized internal events as behaviors, just as those outside the skin are behavioral. Skinner described the "environment" to be "...any event in the universe capable of affecting the organism, but part of the universe is enclosed within the organism's own skin. Some independent variables may, therefore, be related to behavior in a unique way" (1953, p. 257). Skinner seemed to differentiate private events from observable, overt events simply by their accessibility, not by their form or "special properties" (Skinner, 1953, p. 257). Implications for studying private events rests with how we measure, define, and discuss these behaviors, and in conceptually systematic ways that maintain the integrity of the science of behavior analysis. Skinner (1964) mentioned the following:

> No entity or process which has any useful explanatory force is to be rejected on the ground that it is subjective or mental. The data which have made it important must, however, be studied and formulated in effective ways. The assignment is well within the scope of an experimental analysis of behavior, which thus offers a promising alternative to a commitment to pure description on the one hand and an appeal to mentalistic theories on the other. (p. 96)

One must remain conceptually systematic in their approach to studying, analyzing, and addressing behavior change with evidence-based practices for all types of behavior, both overt and covert. As it relates to private events such as thoughts and feelings, some studies have assessed how early children begin to label and understand emotional vocabulary. Studies have shown that the use of emotional language develops with age. One study by Simoës-Perlant et at. (2018) conducted three sets of experiments with children ages seven to eleven to assess their understanding of emotional vocabulary. Results from the study showed that around the age of 7 years, children could identify the valence (attributing pleasant or unpleasant labels) of emotions more prominently, and by the age of about 9 years they could also identify the degree of arousal or activation (excited or calm). While these assessments attribute some understanding related to self-reports and private events, there is more to be studied regarding the application of data collection and treatments regarding private behaviors. As behavior scientists, how do we conduct assessments for independent variables in which behaviors are a function but not directly accessible? Skinner made mention of physiology of sensation (that is, physiology of receptors within the nervous system) and field of psychophysiology as additional ways of addressing behaviors occurring "within one's skin" that may only be experienced or observed by the individual themselves (Skinner, 1953, p. 281). Critchfield et al. (1998) also made a case for the advantages of self-report information in areas where behavior cannot be observed or where there is a limited view/observation of behavioral contingencies; for example, humans carry complex learning histories, and situations such as trauma-informed considerations should be regarded as an example of learning histories that may not be observed simply by a three-term contingency. In these cases, Critchfield

et al. (1998) reiterated the significance of self-reports. In using self-reports for behavior change, some factors involved in the process can include reliable descriptions and measurement of the behavior of interest.

Indirect Measurement and Self-Report Data

As defined in Cooper et al., (2020), indirect measurement occurs as the researcher uses some kind of proxy for the behavior of interest and "...provides secondhand or 'filtered' information that require the researcher or practitioner to make inferences about the relationship between the event that was measured and the actual behavior of interest" (p. 103). Some examples of indirect measurement methods include questionnaires and self-reports. Although the field of behavior analysis typically relies more on direct measurement methods than indirect methods, indirect measurements such as self-reports are useful when the researcher/practitioner has no direct or reliable access to the target behavior, therefore relying on some type of indirect measurement to glean information about the behavior of interest (Cooper et al., 2020; Critchfield et al., 1998). Information about private events of an individual are behaviors that are often tracked via indirect measures due to limitations with accessibility and direct observation (Green & Reid, 2006; Kostewicz, Kubina, & Cooper, 2000; Kubina, Haertel, & Cooper, 1994). Critchfield et al. (1998) mentioned that despite some concerns with self-report data, these concerns or limitations "…do not necessarily render them less informative than data collected in other ways…[and] much can be accomplished by combining self-report methods and the traditional methodological rigor that human operant researchers traditionally have applied to the study of nonverbal behavior" (p. 436). In relation to this point and due to lack of accessibility of private events, one could attempt to provide some

relation between the observable behaviors that occur along with the internal behavior in a reliable and meaningful way (Cooper et al., 2020).

Within the field of applied behavior analysis, self-reports alone are not considered a reliable measure in and of themselves for several reasons. One potential limitation of the validity of self-report measures is the reliance on memory. In some cases, individuals may be reporting information based on past events, creating an increased gap between the actual event and memory recall, which may weaken the certainty of data. In a study by Morsbach and Prinz (2006) on self-reporting of parenting, asking others to recall several events at a time (such as frequency of problematic behaviors of their child) presents a challenging task, particularly for high-frequency behaviors. Misunderstandings related to the definition and interpretation of the question being asked or regarding the agreed-upon behavioral definition to report also pose risks to reliable self-report data, as the individual may not be reporting what the study is intending to measure. Other limitations to the reliability of self-reports involve sensitivity of the material to disclose, which can include the potential pressure to provide a response that is socially desirable and the intrusiveness of some topics to the participant (Morsbach & Prinz, 2006).

These flaws of self-report data carry considerable weight in how treatment decisions are informed and in the accuracy of addressing the true issues at hand. Given these limitations, if self-report data will be measured and considered for behavior change programs, other measurement methods with more evidence-base for accurate reporting should be supplemented with the self-reports in the analysis and treatment process.

Direct Measurement

Cooper et al. (2020) defined direct measurement as occurring when "...the behavior measured is exactly the same as the behavior that is the focus of the

investigation or behavior change program" (p. 104). Direct measurement can occur in natural or contrived environments and must involve the direct observation of observable and measurable behaviors, which aligns with the behavioral dimension of behavior analysis (Baer, Wolf, & Risley, 1969). The seven dimensions of behavior analysis include: applied (behaviors or organisms of investigation or treatment are socially important), behavioral (behaviors under study should be observable and measurable in some way), analytic (demonstration of experimental control to produce the occurrence or nonoccurrence of behavior), technological (techniques of the intervention or application are explained thoroughly so that others could replicate it), conceptually systematic (procedures and treatment is founded in evidenced based principles), effective (behavioral technologies are found to produce an effect or change on behavior that is socially significant), and generality (behavior change procedures are effectively implemented over time, across settings, and across a variety of behaviors) (Baer, Wolf, & Risley, 1968).

Given the physicality of overt behaviors, precise measurement procedures must be put into place for higher accuracy of these events. Examples of direct measurement of a behavior or event include, but are not limited to, frequency, duration, latency, fluency, magnitude, and event recording. A common type of data collection surrounding behaviors that are typically targeted for change is "ABC recording"; this type of recording provides information about what occurred prior to the behavior (antecedent: 'A'), the behavior itself ('B'), and what event or stimulus was added to or removed from the environment or person following the behavior (consequence: 'C'). Along with the behavioral dimension, all behaviors measured should have a high degree of technological

considerations, such as an operationalized definition of the target behavior that is easily understood and well defined for each person participating in data collection, including the individual for self-monitoring purposes. It should be ensured that all participants of the study or treatment understand the target behavior and can demonstrate that they can implement the intervention or record data according to the technological definition of the target behavior.

Direct measurement methods hold more validity to the study and treatment of behaviors compared to indirect measures. Observers can see behaviors occurring in real time and collect the needed information on that behavior by observing the physical events happening in the environment, as opposed to relying solely on an unobservable self-report by an individual.

Physiological Measures

Physiological responses can be measured to determine the reliability and validity of self-reporting skills across a wide variety of behaviors. While physiological symptoms can be assessed via indirect measures (such as questionnaires), other objective measures can be analyzed via electronic devices such as heart rate (beats per minute), respiration rate (number of breaths per minute), movement activity, and stress responses (e.g., sympathetic arousal of the nervous system as measured by increased heart rate, blood pressure, and perspiration). Physiological responses that naturally occur across the day can be monitored for their frequency, rate, or other measures appropriate to the variable of interest. In addressing behavior change programs such as self-management, it would be ideal for the individual or outside observers to identify the physiological changes

within the individual's body as they correspond with events happening in their environment.

Validity and Reliability in Measurement

Cooper et al. (2020) suggested that measurement displays validity when "it yields data directly relevant to the phenomena measured and to the reason(s) for measuring it" (p. 102). Data obtained from direct measurement procedures and relevant dimension(s) for measuring the target behavior have more validity than indirect measures, which rely on inferences of the target behavior. There are three significant elements to measurement validity: (a) using a direct measurement of a socially significant behavior, (b) measuring a dimension of the behavior of interest (e.g., frequency, duration, magnitude, etc.) that relates to the research question or concern of the target behavior, and (c) checking that the data represent the conditions and times in which the target behavior occurs as is relevant to the main question or study about the behavior (Cooper et al., 2020). For example, if a person wishes to reduce the number of negative self-talk statements an individual makes across their day, one may first address the question of social significance to the individual, choose a direct measurement dimension (e.g., frequency or rate) to measure how often the behavior is happening, and observe trends within the target settings and conditions in which the behavior occurs throughout the individual's day.

Reliability is referred to as "the consistency of measurement, specifically, the extent to which repeated measurement of the same event yields the same value" (Cooper et al., 2020, p. 102). Essentially, the greater the consistency in which repeated measures of an event produce the same outcomes over time, the greater the degree of reliability for

the study. The validity and reliability of data should be assessed in all treatment interventions and research studies to determine how credible the measurement systems are to addressing the behaviors of interest.

Studies of Inner Behaviors: Examples from the Literature

There are a few research studies that have measured physiological responses as they relate to challenging or problematic behaviors. A study by Freeman et al. (1999) showed that physiological measures such as heart rate and observable measures of challenging behaviors can be assessed together in real time via electronic devices and video recording/in-situ observers. Visual analysis of their data showed that heart rate and problem behaviors can be measured together in natural environments and showed reliably that higher heart rates were correlated with the challenging behaviors. Barrera et al. (2007) studied some physiological forms and functions of self-injurious problematic behavior (SIB) in three adults with developmental disabilities and found reliable and consistent patterns in heart rate (HR) when compared across various conditions; heart rate consistently escalated immediately before SIB occurred and was followed by temporary drops in HR level quickly following engagement of SIB. Another study of heart rate during challenging behavior was studied by Lydon et al.(2013). Heart rate was monitored in three children with ASD who engaged in challenging behavior, SIB, and stereotypy. Results for all three participants showed that increases in HR co-occurred with SIB (and tantrums for Participant 1), and no consistent HR patterns were shown preceding SIB. Other studies have compared and analyzed heart rate with arousal and activity choices in a child with autism spectrum disorder, showing that both environmental and physiological variables can be assessed in natural contexts such as home settings (Hoch et al., 2010).

There are some studies that have addressed the self-management of private events. Kubina et al. (1994) studied the effects of a 1-minute counting procedure to reduce the negative inner behavior of two female senior citizens. In this study, “negative inner behavior” was described as “negative feelings or thoughts" and both terms were further operationally defined so that the participants could adequately identify when they were engaging in that behavior and record its occurrence. Negative feelings were defined in this study as “...an individually perceived emotional state of dissonant or otherwise uncomfortable emotional tone….such as sadness, gloom, despair, and helplessness” (Kubina et al., 1994, p. 29). A negative thought was described as “...an idea or similar consideration that was self-observed as negative” (p. 29) and the participants gave some specific examples of what their thoughts were. While some feelings and thoughts may not be directly observed by another individual, such as rumination of thoughts (as mentioned as one of the limitations of this study), some reliance on self-reports and self-recording procedures were implemented to distinguish when the target behaviors occurred or not and how often throughout the day. Participants recorded the frequency of their negative inner behavior, as well as the frequency of positive thoughts during the one-minute counting period. “Positive thoughts” was also technologically defined for participants to perform and record. Although the social validity of this study was high (i.e., data showed a correlation between the reduction of their negative inner thoughts over time with use of the intervention, and both participants reported “the outcome was helpful” (Kubina et al., 1994, p. 30) and that it helped decrease their negative thoughts), there were limitations to the study. The most significant limitations were the lack of accessibility and direct measurement to observe thoughts and feelings of another person to adequately say the

intervention served as a direct function to the behavior change. One of the areas for future research described in this study is the need to assess or determine overt, observable behaviors correlating with the inner private events. Repeated observations and measurement should occur to determine if the overt behavior reliably occurs with the private behavior.

Kostewicz et al. (2000) implemented a similar procedure using the 1-minute counting method to manage aggressive thoughts and feelings with non-aggressive thoughts and feelings by comparing a once-a-day intervention using the one-minute method versus a dispersed method via six, 10-second counts of non-aggressive thoughts. Like Kubina et al.'s (1994) study, the participant's frequency of aggressive thoughts per day reduced dramatically, but similar limitations were discussed in this study regarding the accessibility of measuring private events and reliance on indirect measures, such as self-report, to make treatment decisions. Despite these continued limitations, the authors of this study stated the following: "We believe that the limitations of assessments for accuracy and believability should not be viewed as a major reason for not experimenting with inner behaviors" (Kostewicz et al., 2000, p. 185). This aligns with early views of Skinner regarding the study of private events, who stated that despite the challenges of directly accessing private internal behaviors, behavioral scientists should continue to study the environmental contingencies surrounding "the contingencies that govern emotional talk" (Friman et al., 1998, p. 140) to better understand human behavior and address socially significant matters.

In more modern attempts to understand intrinsic behavioral events via assessments, observations, and self-report measurement, a study by Moskowitz et al.

(2013) shed light on these attempts in scientific research and provided a foundation for the researcher's current study. In their study, the researchers used multimethod strategies for assessing anxiety in three children with diagnoses of autism spectrum disorder and intellectual disability. The researchers created an operational definition of anxiety through three measures: observational data of anxious behaviors (frequency of problematic behaviors), caregiver reports consisting of contextual data and reports of the child's affect (via a rating scale), and physiological data such as heart rate (HR) and respiratory sinus arrhythmia (RSA) from a highly sophisticated electronic device called the Alive Heart and Activity Monitor by Alive Technologies (Moskowitz et al., 2013). Together with the children's caregivers, the researchers identified situations in which the participants engaged in challenging behaviors and exhibited anxious behavioral responses which were individualized and operationally defined for each participant. Some examples of the operational definitions included the following: clinging to (holding onto) a caregiver, crying or tearful responses in the eyes, cowering one's body into a corner, vocalizations such as moaning or whimpering, repetitively questioning where a caregiver was at, running to and checking the door frequently for the caregiver, covering one's ears, rapid movement of the eyes darting back and forth, and mouthing one's fingers (Moskowitz et al., 2013). As mentioned before, the context in which these behaviors occurred were identified and used as the "high-anxiety" setting within the functional assessment of high and low-anxiety conditions, as well as the target setting for intervention (Moskowitz et al., 2017). Results indicated that the participants engaged in higher levels of challenging behaviors, displayed higher heart rates, and lower RSA in the high-anxiety conditions compared to the low-anxiety conditions.

Contextual and affective measures are significant to the analysis of anxious behaviors and physiological data, given that similar physiological responses can occur across a variety of contexts and be attributed to different psychological or emotional attributes; this is described by the example of increased HR and RSA, as well as caregiver report of high affective arousal, while a child is engaged in physical exercise compared to the same reported affect and physiological symptoms while one is taking an exam; both physiological responses occur during different contextual events (Moskowitz et al., 2013). The researchers were able to display an adequate portrayal of how behaviors attributed to private/internal events (such as anxiety) can be reliably assessed, measured, and analyzed for correspondence with one another. The researchers successfully followed up the study with treatment interventions to address anxious behaviors with the same participants using multicomponent treatment packages (Moskowitz et al., 2017), implying that these types of behaviors can be analyzed and treated using evidence-based practices.

While there were many successful strengths to the two studies by Moskowitz et al., there were several limitations to their studies and areas for future research. One limitation is that of the reliability of caregiver reports of anxious behaviors displayed from their children. The caregivers' reports may not have had the accuracy to determine what behaviors were attributed to anxiety or to other emotional states, such as fear. The study also lacked in conducting assessment measures to discriminate anxiety from other emotional states such as anger, frustration, fear, and sadness. Another limitation lies with the device used to track the physiological data. Although the device was highly accurate and reliable in terms of data collection and research purposes, the external validity of

others replicating this study in applied practice settings is not very high given the size of the device pack and how salient the pack and other measurement mechanisms may have been to others outside the study (the device was located in a pack the children wore strapped to their back while electrodes were strapped to their chest for the electrocardiogram). Another area for future study lies with self-report data from individuals with autism spectrum disorder or intellectual disabilities. Moskowitz et al. (2013) used self-report data from the participants' caregivers, which leaves room for more investigations on self-report data from the participants themselves. All of these variables account for areas of focus for future research. In the current study, the researcher plans to analyze self-report data from child participants who have a diagnosis of autism spectrum disorder or developmental or intellectual disabilities.

Current Study

Skinner, along with many other researchers in the field recognized the limitations to accessing and measuring private events and relying solely on self-reports as true values for measuring behavior within the field of behavior analysis; despite these limitations, they have also encouraged the curiosity and study of private events and the environmental variables surrounding these events. The current study aimed to address the following question: are there valid measurement procedures (e.g., analyzing physiological responses such as heart rate) to analyze the correspondence between self-reports of emotional states and observable and measurable behaviors (such as direct observations of overt behaviors)? That is, are there physiological data to support the behavioral observations and verbal reports occurring in the environment? In attempt to address these limitations in measurement validity, this study compared these three variables in the following manner: the researcher collected physiological data via a

wearable electronic device (Fitbit® Charge 5 wristband), collected data on the participants' self-report of their own private/emotional states via a 5-point rating scale, and direct observation and data collection of the participants' overt behaviors occurring in the environment using a similar 5-point scale. Visual and statistical analyses were conducted to determine the correspondence of these three variables.

Method

Participants

Two participants participated in this study. The data analysis per participant involved the comparison of the three variables to analyze measurement validity and correspondence of the variables. Jared (Participant 1) was an 8-year-old male and had a diagnosis of autism spectrum disorder (ASD). Jared took 1 milligram of Tenex daily to help treat symptoms of attention-deficit/hyperactivity disorder. Charlie (Participant 2) was a 9-year-old male who also had a diagnosis of ASD. At the time of the study, both participants were receiving ABA services at an outpatient center in Harrisonburg, Virginia, to address behavioral challenges and adaptive behaviors related to their diagnosis. Both participants displayed strong cognitive, receptive, expressive communication skills to communicate their emotional or physiological state; for example, Jared was observed to say things like "I'm really frustrated right now" or " "I was mad when ____," even outside of the self-report check-ins of this study. Jared could also verbally state rules and describe appropriate social behaviors and explain why a behavior is inappropriate. Charlie could also verbally articulate his emotions and explain events that happened to him related to his emotions. These participants would often vocally label how they were feeling aloud at the times of their check-ins when presented with their

rating scales, even though this was not required for their study. These skills were also assured via review of their current developmental assessments on file that measure various components of verbal behavior and social/emotional skills (i.e., via scores from the Verbal Behavior Milestones Assessment and Placement Program, and the Promoting Emergence of Advanced Knowledge assessment). The primary researcher described the study and procedure to both participants and their assent was obtained before implementing the study via a signed document between the participants and the primary researcher. Self-reports were collected via a 5-point rating scale on paper which included the typed numbers 1 through 5 with an emoji face assigned to each number to represent emotional states increasing in intensity (see Appendices B and D).

Setting

Data collection occurred within an outpatient ABA agency that both participants attend on a weekly basis. It was reported by the participants' BCBA and caregivers that their target behaviors occur in this outpatient setting during the contexts identified in their interviews at the start of the study. The specific settings in which their observations and self-report check-ins occurred included the center's sensory gym, therapy rooms (where individual and group activities are held), and an open classroom that holds a variety of toys and leisure activities. These settings allowed the researcher and research assistant to obtain information on the participants' target behaviors during their typical ABA sessions as they naturally occur on a regular basis.

Caregiver Interview

Prior to implementation of the study, the participants' caregivers (one parent per participant) and the participants' supervising BCBA (who was also the research assistant in this study) met with the primary researcher to answer questions from an interview that

helped identify the target behaviors for each participant. This interview included questions that were taken and adapted from the Functional Activities Interview for Teachers (FAI) consisting of a series of open-ended questions about the participants' target behavior(s). Some questions within this interview addressed the following: the topography of the behavior(s) (e.g., what the behavior(s) look like in form), the environmental conditions (e.g., settings, people, time of day, or activities) in which the target behavior(s) occur most and least often, any precursors to the target behavior(s), any safety concerns the behavior(s) may pose to the participant or others, and identified preferred items, settings, and activities of the participants. These questions also helped guide decisions about where the study took place with each participant.

Behaviors of Interest for the Study

Jared

Jared's target behavior(s) included perseverative speech, inappropriate language, and verbal and physical aggression, all of which typically occurred when denied access to preferred items/activities. *Perseverative speech* involved the act of repeating a question, request, or comment more than once after being denied a request or told not to speak of that topic until he completed a task; examples include asking a third time to play with a toy after being told the toy is not available, saying "I want to tell you my nickname" more than twice during a group game, or commenting about the batteries of a toy more than once after being told the batteries for the toy are dead. Jared's father reported that perseverative speech tends to occur as an anxious response when Jared is unable to access something preferred. *Inappropriate language* involved cursing or swearing at others, name-calling, or negative hyperbolic statements (e.g., *You're the worst mom ever*; *I wish you weren't my mother*). *Verbal aggression* consisted of statements involving threat of

physical harm toward another person (e.g., *I'm going to punch you*), and *physical aggression* consisted of physical contact with others involving force from some part of his body such as pushing, pulling, or hitting others, or pressing his body on another person. It was reported that physical aggression did not occur at a high rate prior to the study being implemented and that this behavior had reduced some within the clinic setting. It was reported that Jared's target behaviors also occur when another child is using something he wants access to, occasionally when other children are being loud, and when instructed to do something he doesn't want to do.

Charlie

Charlie's target behavior involved *frustration*, which was defined in several ways. Precursors to frustration involved Charlie quieting his voice and not responding to others' attempts to talk to him when asked a question, as well as physically "shutting down" or drawing inward toward his physical self (e.g., lowering of his head and shoulders). Frustration escalated behaviorally when Charlie whined or repeated questions or statements about contextual events that occurred, such as repeating that he made a mistake or a statement about a change that disrupted his typical routine. Frustration escalated most when Charlie engaged in yelling (raising his voice above a typical volume one would use when talking indoors) or vocal refusal to engage in a task, crying, property destruction (ripping materials or throwing them away), and using exaggerated and/or catastrophic descriptions of a situation where that response did not match the actual level of difficulty or contextual events (e.g., such as "This is the worst thing ever!" or "Call the fire department!" after making a mistake on a worksheet). It was reported that Charlie's target behaviors occur most often when there is a disruption to his typical daily routine or

he cannot access something he typically has access to in his routine, as well as when he makes a mistake, mostly academic in nature, such as writing a word or answer incorrectly with an ink pen. Charlie often wrote with a pencil to correct mistakes.

Materials

The Fitbit® Charge 5 was the electronic device used to monitor and track the participants' heart rate (beats per minute), and both participants had a separate device for their individual use. The researcher allowed each participant to take their Fitbit® devices home with them and to wear them for as long as they desired throughout the duration of the study (while ensuring that they wore it during data collection/observation times in the outpatient center). The participants were given the charger to maintain adequate battery capacity each day to maintain useability of the device and optimized functioning per session. Their Fitbit® Charge 5 wristbands were adjusted on their wrist by the primary researcher and their caregivers to accommodate comfortable use while wearing the device. Data collection sheets included the Behavioral Indicators Data Sheet for Observers and writing utensils were used to document behavioral observations within the sessions via direct observation (Appendices A and C), as well as the adapted Behavioral Indicators Scale for Participants data sheet for the participant's self-report data (Appendices B and D). Both the primary researcher's and the research assistant's password-protected cell phones were utilized to download the Fitbit® app so that the heart rate data from the Fitbit® Charge 5 devices could be synched, seen, and monitored frequently. Heart rate data were downloaded from the online Fitbit® account linked to each device and analyzed via Excel programming. Other materials that were utilized within this study included materials within the participant's(s') natural environments and treatment sessions where this study took place (e.g., toys and preferred items/activities,

worksheets, or crafts that the participants engaged in with their behavior therapist and peers, writing utensils, etc.).

Variables

Three main variables were studied and compared together: the participants' heart rate (bpm), direct measures of behavioral observations recorded on the behavioral indicators scale for each participant, and the participants' self-reports via the same behavioral indicators scale adapted for each participant. The main physiological symptom that was tracked was the participants' heart rates (bpm); other data that were reviewed included the participants' sleep activity, however, this was not a main measure of the study. Data regarding sleep scores (i.e., hours of sleep per night as well as the amount of time the participant was in light sleep and REM sleep) were automatically recorded via the Fitbit® Charge 5 device and was monitored if these data added any supplemental information to the participant's(s') data analysis (such as setting events that may be related with the occurrence of specific behaviors).

Direct measurements involved in-person observations of the participant's overt behaviors and technological definitions of the topography of their target behavior(s) included on the data sheets to reference. The behavioral indicators scale was used for the observers to rate the intensity of the participant's behavior(s) and physical affect at the time of each check-in. This data sheet was organized to gather information similar to anecdotal observation data, or "ABC recording", per check-in time to analyze the environmental contingencies surrounding the target behavior (i.e., antecedent and consequent events), for the purpose of providing detailed information pertaining to the events surrounding the target behavior and any potential patterns or trends gleaned from the data (Cooper et al., 2020). The indirect measurements included the participants' self-

reports which were tracked when each participant circled a number on their rating scale during each check-in time per session.

Data Analysis

The researcher reviewed the data obtained from the study via visual analysis of the graphs. Physiological data obtained from the Fitbit® Charge 5 were downloaded from the researcher's password-protected Fitbit® account and organized and analyzed via Excel Workbook, and these graphs displaying indicators of heart rate (beats per minute) by the Fitbit® Charge 5 were viewed in comparison to the direct and indirect measures in this study. The researcher manually graphed data from the participants' self-report data sheets and the researchers' direct observation data sheets on the same equal-interval line graph for closer visual analysis and comparison.

The graphs for this study had ordinal scales corresponding to the rating scales on the Behavioral Indicators data sheet for both direct measurement (observations) and indirect measurement (participant self-report). Although the numbers are scaled with intervals of one, these ratings are placeholders corresponding with the Behavioral Indicators rating scale from the data sheets, meaning that the difference in scale does not indicate a specified difference in value (e.g., a '2' is not twice the value of a score of '1').

Interobserver Agreement Reliability

Interobserver Agreement (IOA) is the process in which two or more observers collect data on a behavior or event simultaneously (or asynchronous if watching a recorded video) but separately to measure the believability of a study and measurement system (Johnston et al., 2020). IOA is a significant measurement tool to determine the degree that the observers report the same data (observed values). In this study, IOA was collected on 37% of sessions between the primary researcher and the research assistant,

which involved two sessions with Charlie and one session with Jared. IOA was collected on a total of 16 check-ins across these three sessions. Between the primary researcher and the research assistant, disagreements between their ratings of the participants using the behavioral indicators scale for observers were off by one point on the scale. Much like the correspondence between the participant and researcher's scores, IOA also showed differentiation mostly by one rating point except for a two-point differentiation during one of the check-ins for Charlie. Due to the nature of data collection for this study based on rating scales and ranges of 1-5 per check-in, the researcher calculated IOA by adding the number of times the researchers met complete agreement (circled the same number on their rating scales) and dividing by the total number of check-ins measured during IOA (16 total), as well as the number of times the researchers disagreed in rating by one point divided by the number 16.

IOA met exact agreement in rating scores between the researchers on 50% of check-ins, disagreement by one point on 44% of check-ins, and a difference of two on 6% of check-ins. In reviewing the observer data sheets, both observers had recorded the same data for antecedent events and consequence events for each check in, however, there were some differences in how the primary researcher and research assistant made notes of facial features of the participant. For example, the primary researcher rated Charlie as a '1' when he was bouncing on a ball in the sensory gym and made note of him smiling and giggling as the researcher approached him for a check-in, whereas the research assistant recorded "bouncing on a ball" without the observable emotional affect.

Procedural fidelity

Procedural fidelity is a measure of the degree to which the researcher/program provider implements the procedure with the individual as intended and outlined by the

treatment plan or data collection protocol (Feely et al., 2017; Johnston et al., 2020). For the current study, the researchers assessed the fidelity of the data collection tools via a procedural checklist created by the researcher, and procedural fidelity was assessed in a few ways. First, the researcher verified that the Fitbit® Charge 5 was fully charged and functioning properly and was positioned on the participant's wrist at the scheduled session times where both direct measurement data from observers and self-report data from the participants were collected. The researcher also collected a permanent product measure via a timestamp to verify that the participant self-reported within a designated session period and that the researcher collected data at the same time as the participants as indicated on the participants' and researchers' data sheets during the chosen check-in time. Finally, the research assistant and primary researcher coordinated procedural fidelity checks of direct measurement data to check that direct measures via the Behavioral Indicators Data Sheet for Observers were taken at appropriate times and according to the technological definitions of the target behavior (e.g., comparing data sheets and time stamps). Procedural fidelity was collected on 36% of the total check-in times for this study in which data collection was performed with both the primary researcher and the research assistant. The researchers met 99.5% fidelity across all sessions in which treatment fidelity was measured.

Procedure

A behavioral indicator rating scale (Likert scale ranging 1-5) was created by the primary researcher based off information from the FAI-adapted interview to determine how to code the target behavior. These ratings addressed the intensity or magnitude of the behavior based on technological descriptions of each number on the scale as a way for

the participant to rate their overall well-being during each self-report check-in, as well as for the researchers to rate the same well-being or behavioral intensity of the participants via direct observations. The descriptors for each point on the scale for the participants' check-in sheet were individualized per participant and represented with visual emojis (e.g., facial expressions resonating with the intensity of their emotional response) and written description of what each point on the scale represented related to the target behavior. An example of the rating scale created for Jared included: "1 = I feel happy or excited; 2 = I feel calm, relaxed, and engaged in what I am doing; 3 = I feel confused or anxious as to why I cannot have what I want, 4 = I feel like I might lose control and say something mean to someone else, 5 = I feel out of control of my emotions, like I want to hit someone or yell/curse at somebody". See Appendices A-D to view each of the participants' rating scales for self-report and the observer's corresponding rating scales. The primary researcher discussed the rating scale and indicators with the participants before implementing the study to obtain the participants' input of their own behaviors and emotions that they may experience during the contexts/activities in which their target behaviors typically occur (as identified during the parent interview), and the primary researcher edited the descriptions based on feedback from the participants. These descriptions were defined in detail so that the participant and researchers had clear definitions of how to take this type of data. Once the data sheets were created and reviewed with the participants, implementation began.

The researcher or research assistant prompted the participant to complete a self-report check-in at various times of the session, which were guided by the contexts outlined in the FAI-adapted interview; these check-in periods occurred during times in

which the target behavior was reported to occur most and least often for each participant in an attempt to capture variability in self and observer-report ratings and heart rate data. Five to six check-ins were performed during each 2-hour session with the participants. When it was time for a check-in, the researcher checked the participant's Fitbit® wristband to ensure that it was secured and positioned properly on the participant's wrist and charged optimally (this also occurred at the beginning of the session to ensure adequate heart rate readings across the session). The researcher(s) and participants completed their ratings at the same time, and the researcher ensured that they positioned themselves and their data sheet away from the participant so that they could not see the participants' rating and vice versa. A treatment integrity checklist was created to ensure that the primary researcher and research assistant were implementing the steps of this study with fidelity (see Appendix C).

Results

Physiological Data

Data from each participant's Fitbit® Charge 5 device were downloaded from their Fitbit® online account and graphed in an Excel document. Heart rate is displayed as beats per minute (bpm) and the data that are shown in Figures 1 through 4 display heart rate variability across half-hour periods of time within the participants' sessions for two different session dates per participant. Jared's average resting heart rate averaged between 66-75 bpm, and Charlie's resting heart rate averaged slightly higher, between 90-105 bpm. The participants' data show the Fitbit® device captured heart rate readings as the participants wore the wristband and reflects variability in heart rate across different motor activities the participants engaged in. In Figure 1, for instance, Jared's heart rate

displayed some variability in range between 94 and 116 beats per minute in time segment ‘A’ while he was engaged in a preferred activity where he alternated often between sitting and standing; this activity involved flying pieces of paper over a fan and catching them when they fell to the ground. In segment ‘B,’ his heart rate increased to 113 bpm while he was running in the sensory gym chasing his peers. This graph shows his heart rate declined between segments ‘B’ and ‘C’ when he was instructed to leave the gym and transition to a table in the therapy room to complete some worksheets, where his heart rate ranged between 76 and 105 bpm while transitioning to the room and sitting at the table. In segment ‘D,’ Jared’s heart rate quickly increased again, with his highest heart rate reaching 127 bpm while he was in the sensory gym again running and playing with an exercise ball.

Charlie displayed a slightly higher resting heart rate than Jared throughout the study, ranging between 95-110 bpm during sedentary activities. Snapshots of Charlie’s heart rate data within two of his sessions are displayed in Figures 3 and 4.

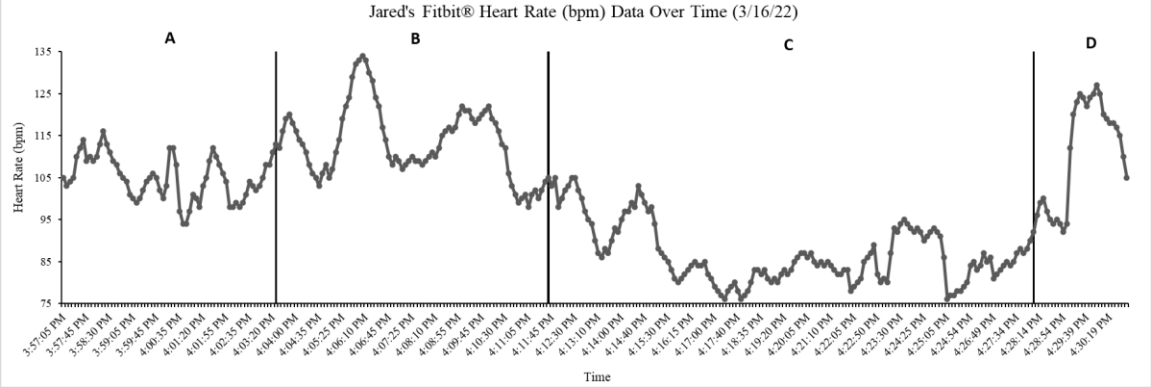

Figure 1. Jared’s heart rate readings over approximately a half-hour period. Segment ‘A’ shows some fluctuations in his heart rate as he engaged in a preferred activity where he alternated between sitting, standing, and slight movement to catch pieces of paper that were blown around by a fan. Segment ‘B,’ Jared was running with peers in the sensory gym and then instructed to leave the gym at 4:06 PM. ‘C’

displayed heart rate data after Jared transitioned to the therapy room to sit at a table and complete worksheets. Heart rate elevated once more in segment ‘D’ when Jared was running and playing with a therapy ball in the gym.

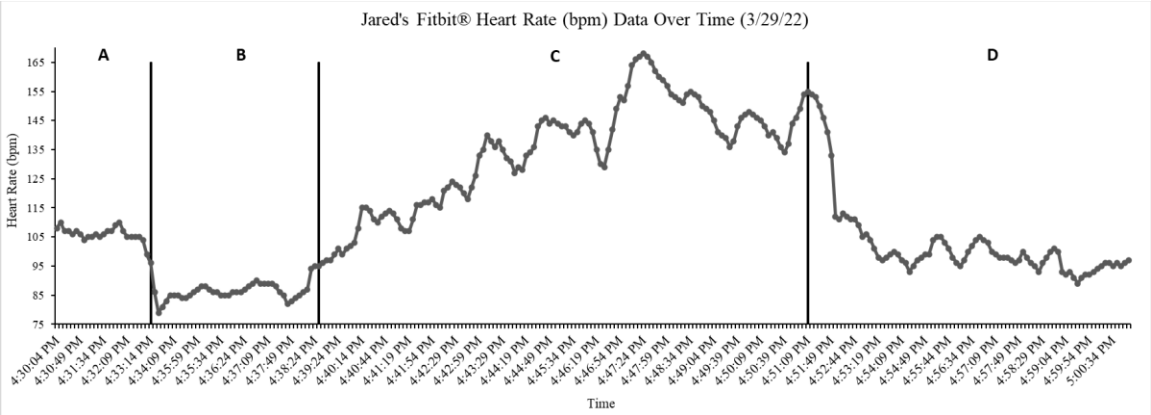

Figure 2. Segment ‘A’ displays a slightly elevated heart rate for Jared as he was transitioning from free play to the therapy room and spinning his body around while making the transition. In segment ‘B,’ Jared was sitting at the table coloring a worksheet. ‘C’ shows Jared’s heart rate increasing as he transitioned to free play where he engaged in a game of tag with a peer, with his heart rate reaching a high of 167 bpm. Segment ‘D’ displays Jared’s heart rate as he transitioned from the gym to the therapy room and sat at the table to engage in a board game with peers.

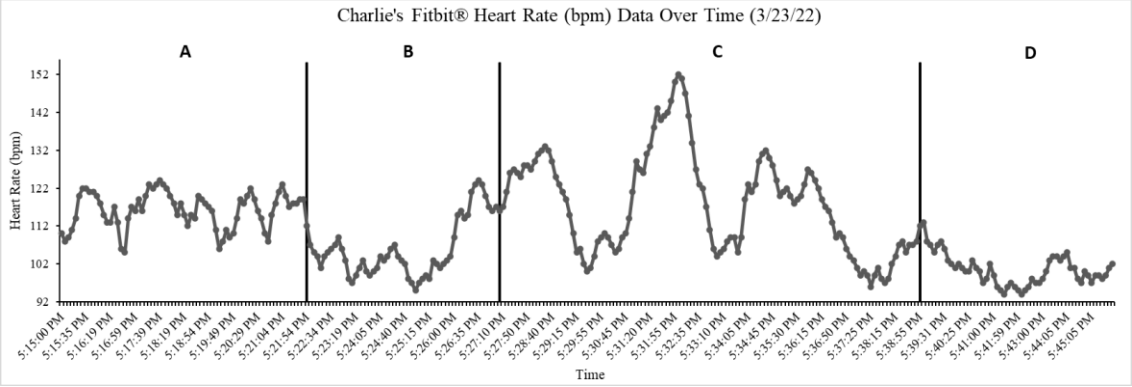

Figure 3. Charlie’s heart rate readings over approximately a half-hour period. Segment ‘A’ shows heart rate data while Charlie was walking around the open classroom and asking peer’s questions or to join him in games. In segment ‘B,’ Charlie was wandering around the classroom and stopped to ask for blocks; he perseverated on this request and was denied access to the blocks. Segment ‘C’ shows data from Charlie’s

motor play in the sensory gym where he ran and bounced on a therapy ball, with his heart rate reaching 151 bpm at the highest data point. Segment 'D' displays data from when Charlie was seated at the table playing a game with peers; he sat quietly and participated.

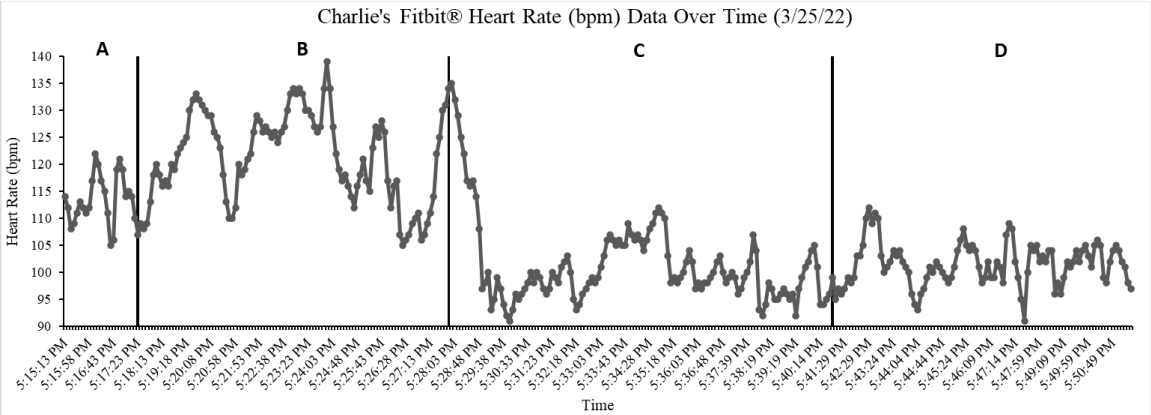

Figure 4. Segment 'A' displays data from when Charlie was at the table completing a math worksheet (a highly preferred activity). In segment 'B,' Charlie was engaged in free play where he alternated between bouncing on a ball, running, or walking in the sensory gym or open classroom. Segment 'C,' Charlie was seated at a table playing a card game with peers. In segment 'D,' Charlie was engaged in independent play where he was standing and drawing on paper.

These data show that the Fitbit® device was able to capture accurate readings of heart rate variability across various motor movement activities, as there are increases in both of the participants' heart rate during high energy motor activities (e.g., running in the gym or around the open classroom, bouncing on a therapy ball), and lower bpm during more sedentary activities (e.g., sitting at the table completing worksheets, sitting and playing a game of "Go fish" with peers, etc.).

Correspondence Between Participant and Observer Ratings

Participant and observer ratings were collected each session at the same time per "check-in" period to report the participants' behavioral/emotional affect during that time using the 5-point scales. Each participant participated in a total of 22 check-ins across

their four sessions, making a total of 44 possible comparisons between observer and participant ratings. Figure 5 is a scatter plot of these ratings.

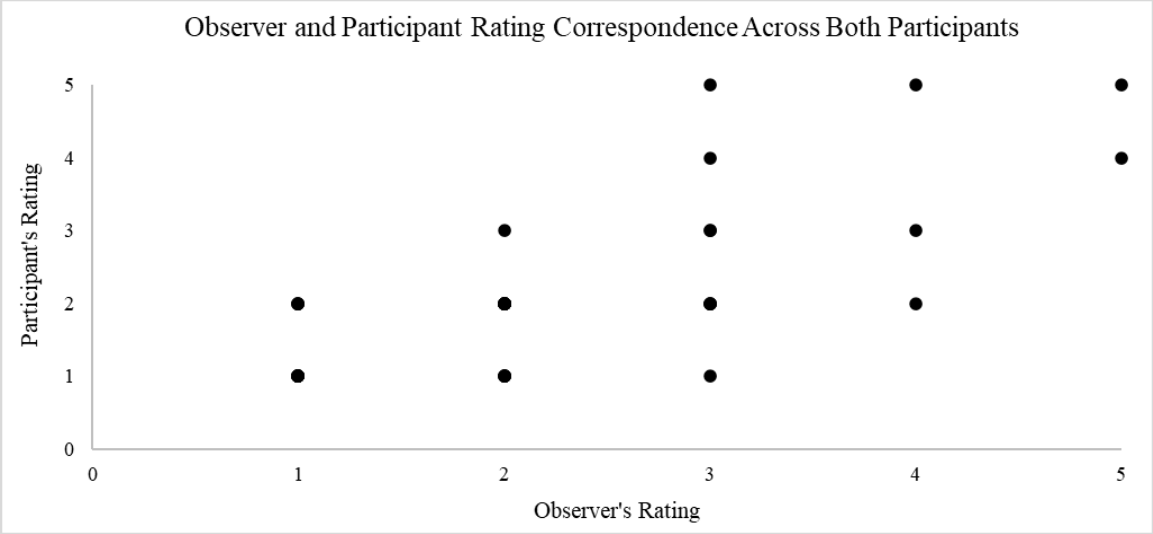

Figure 5. Scatter Plot data of observer and participant ratings for both participants.

A correlation analysis was conducted in order to examine the question of whether there was a relation between the observer's ratings and the participants' ratings. A strong positive correlation of .72 ($p < .01$) was calculated between the observer's ratings and the participants' ratings. The observers and participants often rated closely on the scale during each check-in period; for instance, 23 total ratings were in perfect agreement, and the observer and participant ratings were off by one point on 18 occasions. The remaining 3 check-ins were off by two points, however, this was the greatest amount of disagreement in scores overall. These findings indicate that there was good agreement across the ratings - where lower participant ratings also had lower observer ratings and higher participant ratings also had higher observer ratings.

Physiological, Observer, and Self-Report Data Correspondence

Further analysis compares the correspondence between the Fitbit® Charge 5 device's heart rate data (in beats per minute), the participant's self-report data from their

five-point scale, and the observer's direct measurement data from their five-point scale. For both Jared and Charlie, elevated heart rates and their observer ratings and self-report ratings of '1' consistently corresponded with highly preferred motor activities in which the client was engaged in something of their choosing. Some examples of this particular correspondence can be seen in Figure 6, Panels 4 and 8 (Jared), as well as Figure 7, Panels 3 and 4 (Charlie). Higher heart rates in these conditions were attributed to more vigorous physical activity.

Some differences in ratings occurred most often when the researcher scored the participant as a '2', indicating in their observations that "the child's body is relaxed, calm, and he is engaged in the activity at hand without refusals" (as written on the observer's scale for both Jared and Charlie). While this researcher observed behaviors such as staying seated, looking at and completing their assigned tasks at the table, answering questions during social skills time, and quietly engaging in the task at hand (receiving a rating of '2'), the participants rated themselves occasionally as a '1' ("I feel happy or excited.") . A score of '1' on the observer's data sheet states that the participant "shows signs of being happy, or excited as evidenced by smiling, laughing, or engaging in physical motor activities or positive interactions with other". This could be that some internalized behaviors such as the participant's thoughts or feelings of being happy or excited may not be directly observed by others in the way of apparent physical dimensions described in the observer's rating scale (e.g., smiling or laughing, engaging in physical motor activities, or positive social interactions with others); this could have been a limitation to using emojis to represent facial expressions corresponding with specific emotions Heart rate data in some of these scenarios, however, show some increases in

heart rate even though behavioral responses remain the same; an example of this is in Figure 7, Panel 1 with Charlie. During this check-in period, Charlie was engaged in a worksheet at the table where he was coloring and drawing during an adult-led activity that included a highly preferred theme (Minecraft). While the researcher rated him as a '2' for being calm and engaged in the task at hand, Charlie rated himself as a '1' that he was happy and excited. There was an increase in heart rate during this check-in from 117 bpm to 124 bpm. In Panel 6 of Figure 7, however, similar observer and participant ratings occurred in this scenario as well but with a decreasing heart rate as Charlie was sitting on the floor playing with a toy independently. Much of the variation in heart rate for these participants in this study were attributed to physical activity levels compared to emotional responses.

There were some graphs that seemed to display an increase in heart rate during a more intense behavioral/emotional response such as those displayed by Jared while he was engaged in a sedentary task at the table. Figure 11, Panel 2 shows data from Jared's session where he was seated at a table and started to show an intensity in behavioral responses after he was told to wait to talk about his preferred topic until he completed the activity. Jared's heart rate peaked right before the check-in was conducted, from 96 bpm to 106 bpm; this likely shows his heart rate increasing, around the time he yelled, hit the table, and muttered "stupid" in a whispered tone. A similar correlation occurs in Figure 12, Panel 3 in which Jared was seated at a table for several minutes stacking blocks. Although sedentary, Jared's heart rate was increasing from 98-111 bpm right before the check-in when his blocks kept falling down, even after his behavior counselor attempted to help him. He threw the blocks once, yelled, and hit the clipboard. Albeit quick, two of

these observable behaviors involved some kind of motor response, and the increase in heart rate could be due to these slight motor movements as well. Following the check-in, his behavior counselor presented a different task at the table in which there was a pause in play with the blocks.

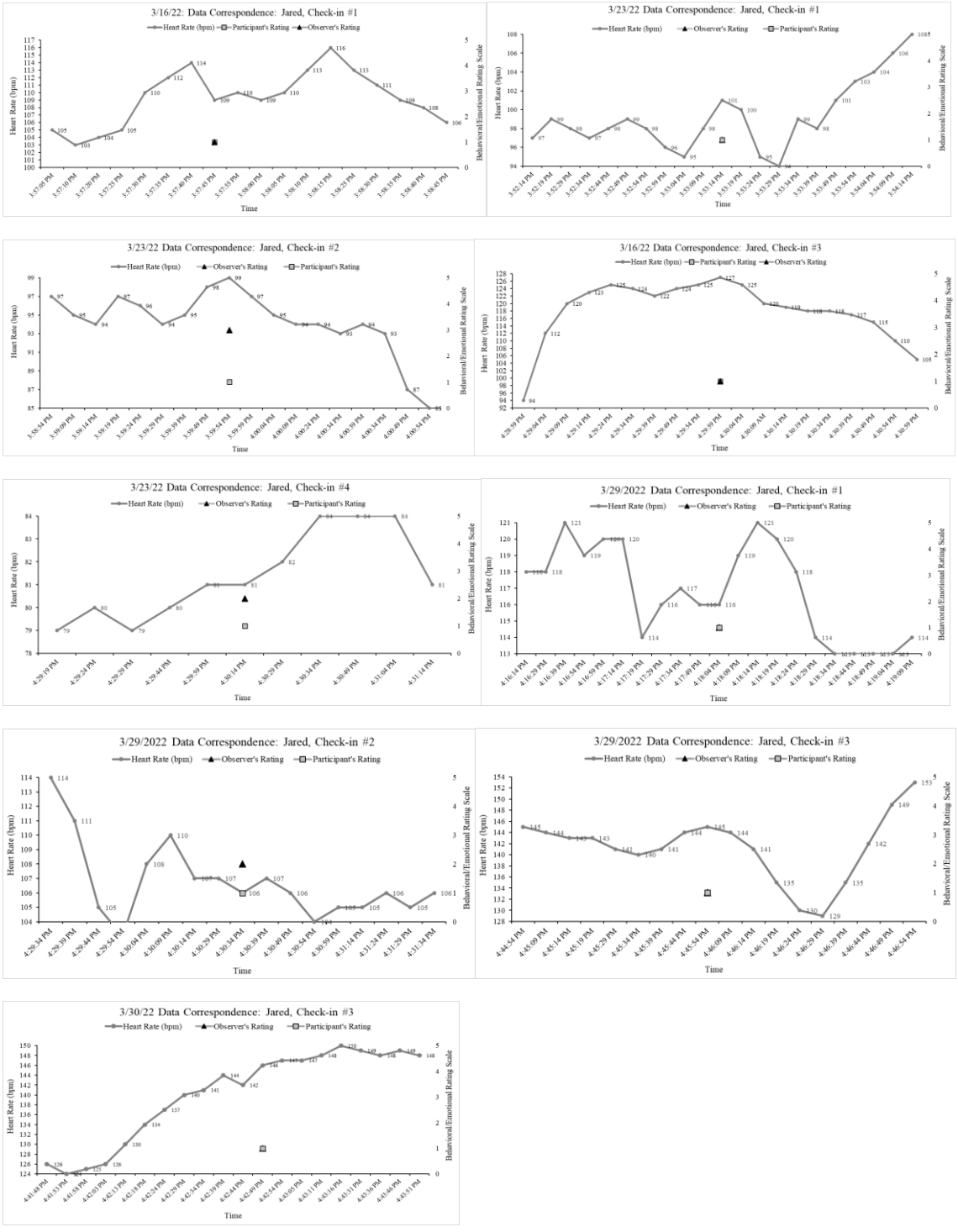

Figure 6. Level 1 participant ratings.

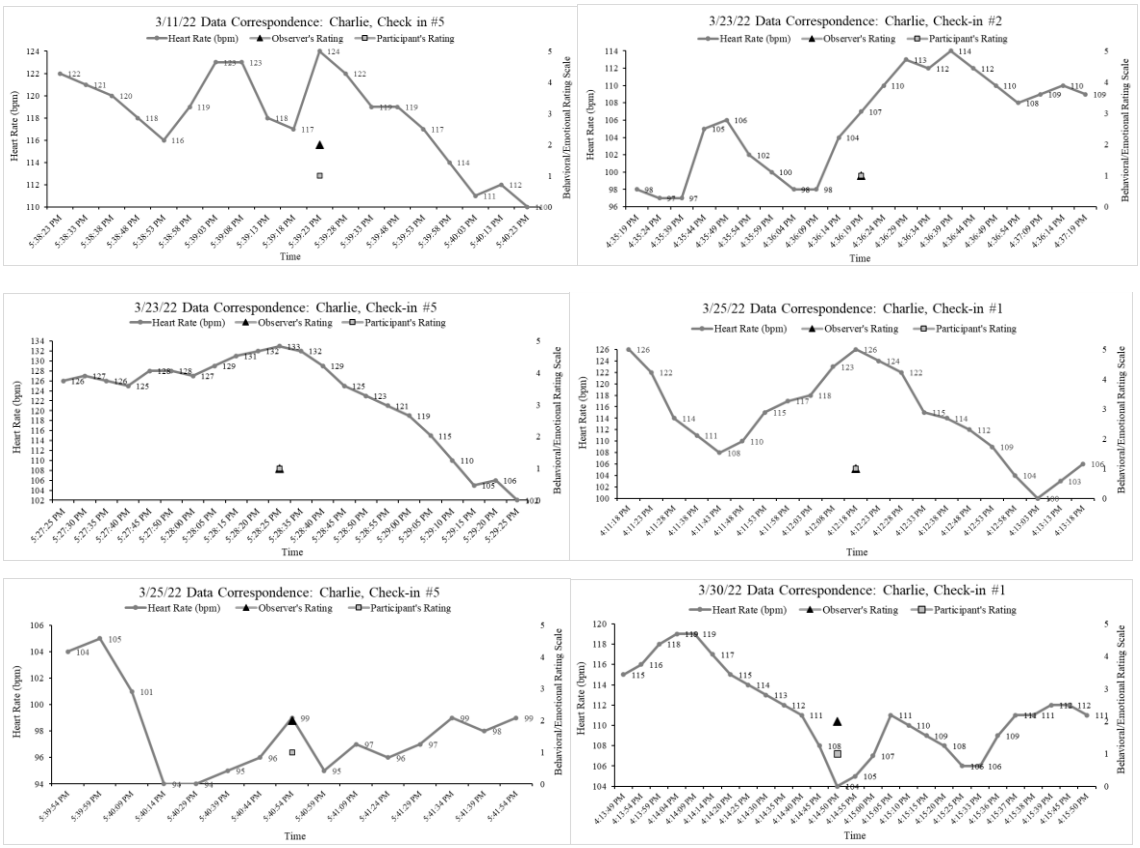

Figure 7. Level 1 participant ratings (continued).

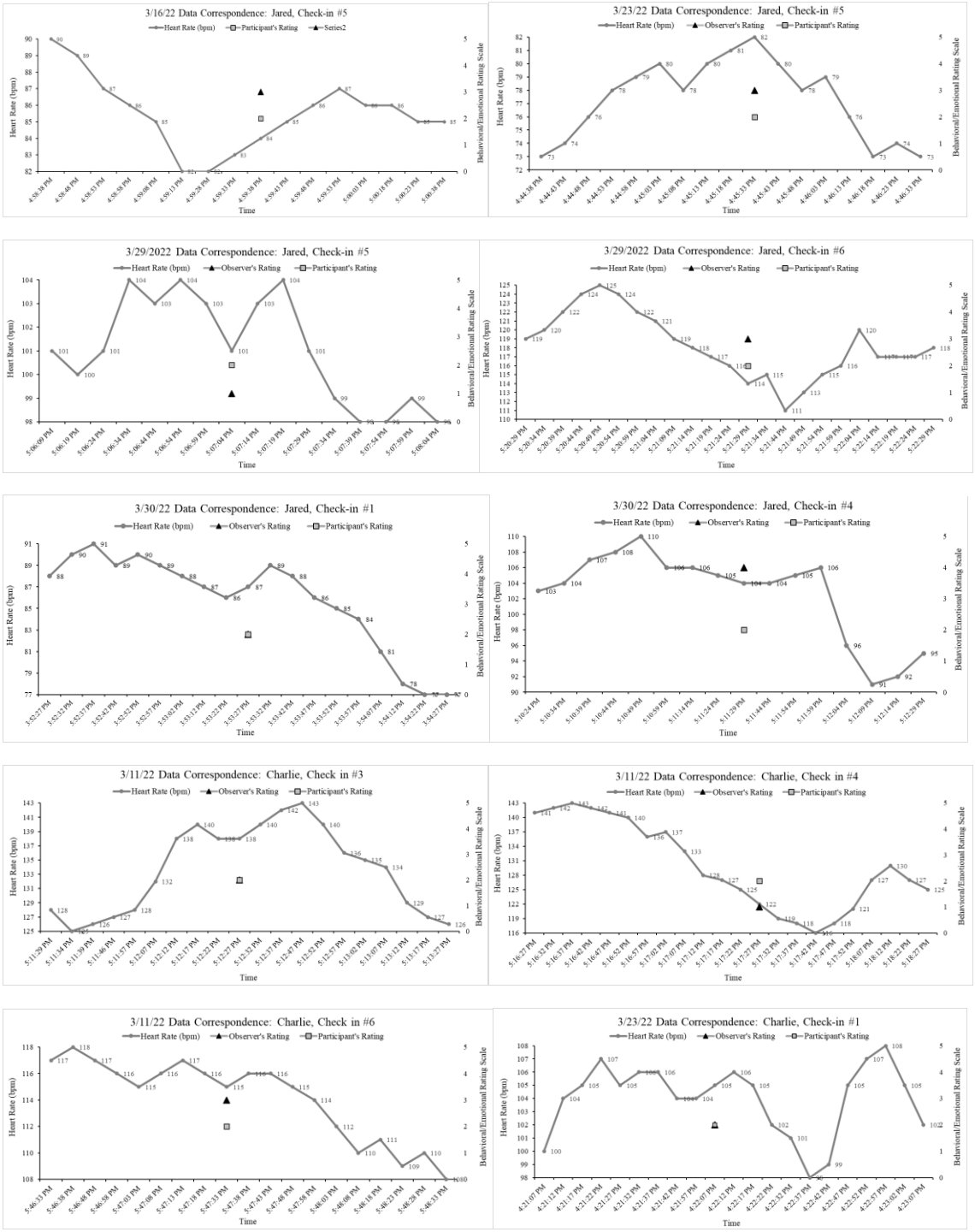

Figure 8. Level 2 participant ratings.

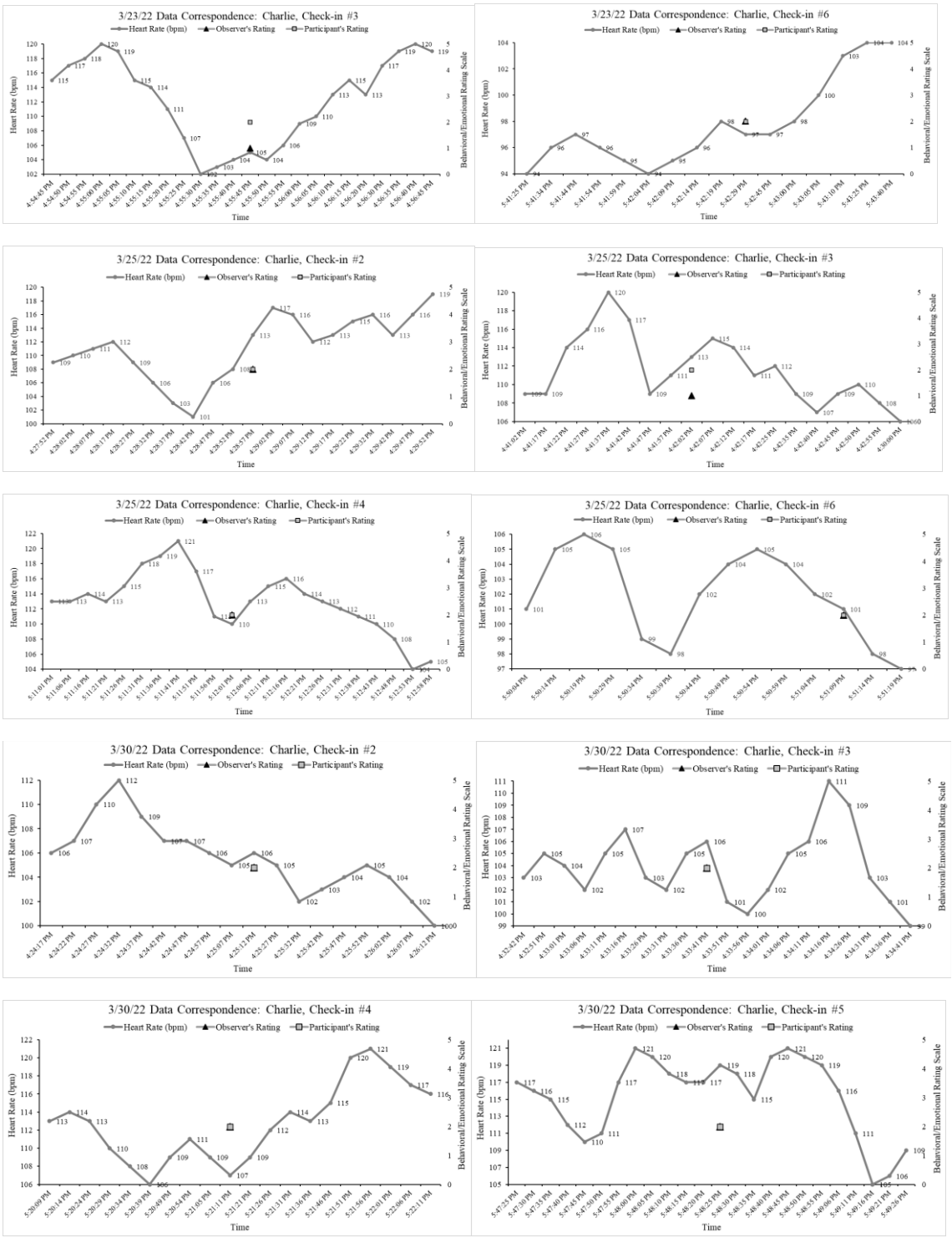

Figure 9. Level 2 participant ratings (continued).

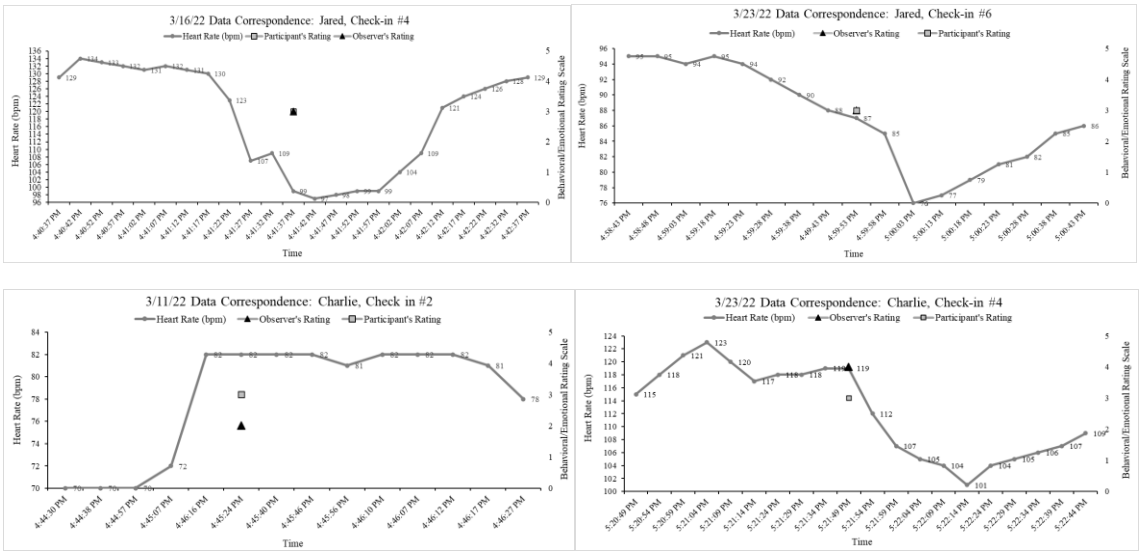

Figure 10. Level 3 participant ratings.

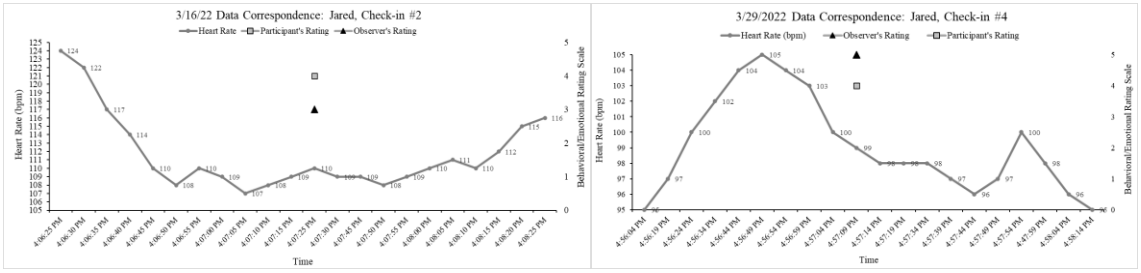

Figure 11. Level 4 participant ratings.

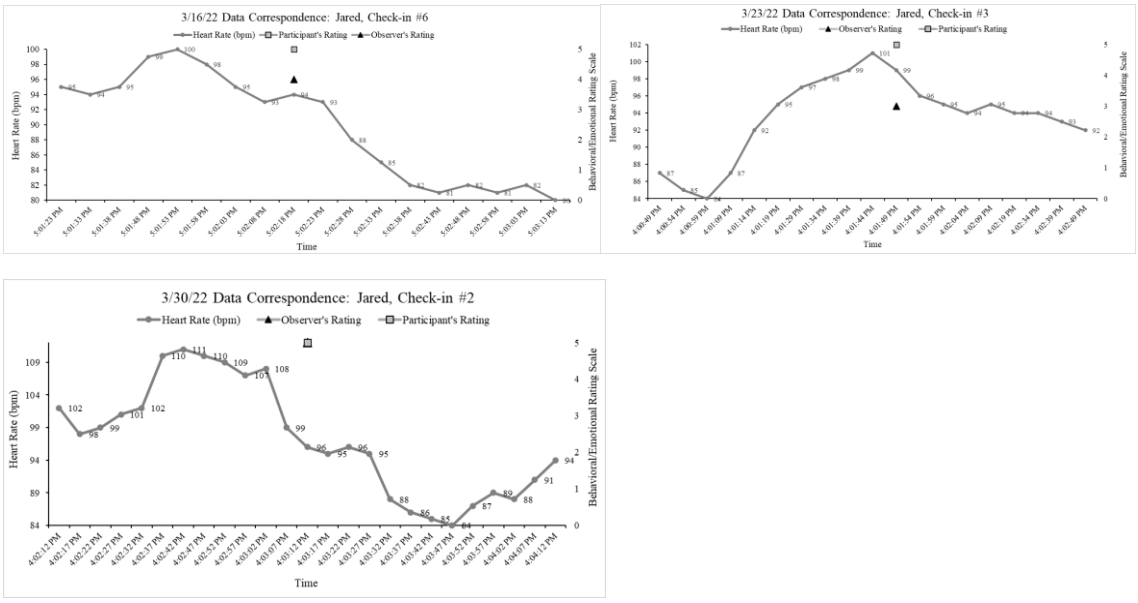

Figure 12. Level 5 participant ratings.

Discussion

The current study worked with two young individuals with autism spectrum disorder to discover the correspondence between the participants' self-report of their emotional states, behavioral observations from an outside observer, and the participants' heart rate across data collection sessions. This study involved similar procedures as Moskowitz et al. (2013) by assessing the correspondence of physiological measures (i.e., heart rate) to the behavioral measures of specific target behavior(s) of the children from the researchers, and in-vivo emotional/behavioral reports from the participants themselves during "check-ins" throughout the sessions. As opposed to Moskowitz et al. (2013), this study analyzed self-reports from the participants themselves (children on the autism spectrum) rather than the participants' caregivers in order to identify the validity of the participants' own self-reports of their emotional states. This study focused on three variables to address a question of the validity of self-report measures of emotional states. More specifically, this study attempted to address the role that physiological data, such as heart rate, plays in supporting emotional self-report scales and the behavioral observations of an outside observer.

This study was able to address measurement validity by using direct measurement procedures of the participants' socially significant target behaviors and measuring the variability of their heart rate and emotional responses communicated via a Likert scale of one through five indicating a range of emotional states. Like a few other studies (Hoch et al., 2010; Freeman et al., 1999), these variables were able to be assessed efficiently in an applied setting with both participants. The data in this study also captured antecedent and consequence conditions for each self-report/observational check-in, including contexts in which the participants' target behaviors typically occur. Jared engaged in his target

behavior(s) more often than Charlie, as Charlie did not engage in his target behaviors very often or display much variation in behavior during some conditions originally identified to evoke the behaviors.

As shown in the results, a strength of this study lies in the strong positive correlation between participant and observer ratings using their 5-point scales. These scales were created to compare the behavioral indicators from the observers' observations and the participants' self-report of their emotional state. These scales were created to assign each emotional response the participants may engage in to a corresponding behavioral/topographical description. In further analysis, the participant and observer ratings scales aligned more closely with categorical expressions of emotions as opposed to linear progression of one emotion or behavior. For example, the facial expressions representing the numbers on the participants' rating scales seems to be showing some dimension of rating their emotional expressions as "positive-to-negative" based on the verbal behavior of our communities and societal expectations of how we label and categorize emotions; that is, attributing the feelings "happy and excited" as positive emotions and "frustrated" or "anxious" as a negative emotions. Although this was not the intent of the study, this scale is more categorical, while the data suggests some element of ordinal. Despite the differences between these types of rating scales, there was still a strong positive correlation between the participant and observer ratings, indicating close agreement in labeling the emotional and behavioral affect of the participants. One limitation to this study related to measurement of the participant's self-report of emotional responding is the confound of physical activity. Heart rate was seen to vary more depending on the participant's activity level than their reported emotional

responses. The Fitbit® Charge 5 device displayed credibility in capturing variation in heart rate across a wide variety of activities that the participants engaged in throughout their sessions. While heart rate was seen to fluctuate for each participant based on physical activities they engaged in, there were limited situations in which predicted levels of heart rate were captured based solely on the participants' emotional affect as communicated via their self-report. Jared displayed some level 4 and 5 ratings, however, some of these instances where his behavioral/emotional affect intensified occurred during sedentary activities but may have included some physical motor activities such as hitting or kicking something that could have impacted the increase in heart rate. These results may align more with Freeman et al. (1999) where they were able to see more consistent patterns in heart rate readings corresponding with self-injurious behaviors like hitting, biting, or other forms of physical aggression rather than capturing heart rate readings of emotional responses where physical aggression or motor activity did not occur. Future research could conduct a similar study with individuals with limited mobility or with individuals engaged in more sedentary activities to see if there is a closer measure of physiological responses based solely on emotional responses.

A limitation with Charlie's data is that he did not display much variability in behavioral or emotional affect within this study. Charlie displayed more flexibility and adaptive behaviors during antecedent conditions that would typically trigger target behaviors, such as when he is told that a specific material he typically works with is unavailable (e.g., using liquid glue versus stick glue, using a pen instead of a pencil) or when a change is presented, such as new social skills programming in which he was to initiate conversation and play with peers rather than adults. Most of Charlie's observer

ratings and self-report ratings remained at levels ‘1’ and ‘2’, with some variability in level ‘3’ on three occasions across all 44 check-ins. The lack of variability in Charlie’s behavioral and emotional data within this study presents some limitations in not being able to see the differences in emotional, behavioral, and physiological affect as emotions intensify and change. It is possible that Charlie could have displayed more adaptive responses due to awareness of the outside observers of this study or given his progress with increasing adaptive skills as part of treatment.

Another limitation regarding Charlie’s rating scale was found in some difficulty with ranking level three responses from an observational perspective. Charlie’s father and behavior analyst reported that he occasionally becomes “closed off” when his target behavior of frustration escalates; topographically, this may look like quieting his voice and not responding to others who are talking to him, withdrawing from a space he was previously at, or tilting his head downward. Although there were few level three ratings during this study with Charlie, this behavior became difficult at times to discern given the nature of its description. For example, on check-in number 6 on March 11th (Figure 8, Panel 9), the researcher rated Charlie as a ‘3’ on the behavioral indicators data sheet (“The child is showing precursors to frustration by quieting his voice or not responding to other people; may withdraw briefly from the space”) when he was building a block tower and they kept falling down, in which he mentioned to an adult that the blocks weren’t “working”, knocked them down, and was not responsive to questions others asked him. During this check-in, Charlie rated himself as a level ‘2’ which is listed on his data sheet as, “My body feels calm and relaxed and I feel engaged in what I’m doing.” Given that these descriptions share similarities topographically, there may have been some

discrepancies between how the observer rated him and how Charlie rated himself. Heart rate data during that check-in was slightly higher than his average resting heart rate, but gradually declined after that check-in period.

In regard to Jared's self-report rating scale, the primary researcher read the full scale to Jared during several check-ins to remind him what each number on the rating scale represented due to some inattentiveness to the rating scale and difficulty with reading. In this case, Jared needed this assistance at times in addition to the facial emojis corresponding with each number in order to adequately attend to the rating scale and rate himself during some check-ins. As part of treatment integrity, the researcher read the full scale to him to avoid suggesting a particular number for him to circle so as not to compromise the integrity of the participant's true self-report information.

Another limitation of this study resides with the setting and time frames in which the study was conducted. While these settings allowed the researcher to observe various behaviors, the setting for this study was limited to an outpatient center during the same time period in the afternoon each session, mostly due to limited scheduling availability between the researcher and the participants' families to conduct the study elsewhere. This may have shown limited variability in behavioral and emotional responses for the participants, particularly with Charlie given that he did not engage in high-intensity behaviors or variability during the study, whereas it was reported in the interview that the target behavior(s) do occasionally occur at home. Conducting these check-ins across various times of the day, in different settings (e.g., home, school, community), and with different individuals (e.g., teachers, parents, siblings) may provide more information about the correspondence between these variables, as limiting the study to one setting and

time frame may limit the types of results. Future research could further investigate data that is obtained across these different settings and contexts.

In terms of physiological data, this study only reviewed heart rate as the physiological variable for comparing self-report data and behavioral observations. Moskowitz et al. (2013) used respiration rate to show the physiological changes within the participants. The researcher reviewed sleep data collected from the Fitbit® device, however, both participants in this study do not typically experience sleep issues that may contribute to variability in behavior. During the last session with Jared, however, he displayed more challenging behaviors in his session and his mother reported at the end of the session that he got less sleep than usual that night before; the researcher reviewed his sleep data for the previous night and noticed a decrease in his typical hours of sleep, which could potentially be a setting event for more intense or challenging behavior that day. Future studies could look at sleep data, respiration rate, and/or electrodermal activity as additional physiological measures of biofeedback to compare to self-report data of behavioral and emotional responding.

In looking at the correspondence between physiological symptoms, observer data, and self-report data of a person's emotional responding and/or verbal reports about their emotional state, this could carry some implications for future research on self-management programs related to private events. Although the correspondence of observer and participant ratings showed a positive correlation, the participants' heightened physical activities seemed to mask true heart rate readings of emotional responses. More research is needed on the correspondence of heart rate and additional physiological factors and self-report data of individuals where one can look solely at the

correspondence between a person's reported or observed emotional state during more sedentary physical responses to better understand correlations between the physiological symptoms and heightened or neutral emotional states.

Despite the confound of physically rigorous activity in the heart rate data, the Fitbit® wristband was found to be an efficient device to track heart rate data and observe changes in the participants' heart rate in real time. The participants reported that they liked wearing the devices and used some other features of the device at times, such as Charlie using the alarm setting to signal when it was time for him to transition from free play to the next adult-led activity. This device could be easy for an individual to use to track physiological changes within their body in real time such as heart rate or other functions like sleep data and stress responses using the electrodermal activity function the Fitbit® device offers. As part of a self-management program, one could check their heart rate readings throughout their day as it aligns with other physiological or emotional symptoms they may be experiencing (e.g., anxiety, stress, etc.) and see how these responses correlate with one another. If a self-management program is tailored toward functional ways of coping with stress, anxiety, or other emotional states, one could track their physiological symptoms or use these physiological markers to indicate when they have achieved a neutral state, such as when their heart rate has returned to their average resting heart rate reading.

This study also shows some implications for understanding the language used to describe and report internal events. Some of the differences in correspondence between the observer's and participants' rating could also be attributed to each individuals' unique learning histories with recognizing and talking about emotions. For this reason, it may

not be surprising to have found differences in correspondence between the ratings. Like Critchfield et al. (1998) discussed, we must investigate the operant contingencies surrounding how individuals learn to self-report; this is particularly true for private events in which only the individual has access to the internal events, and we must therefore rely on public behaviors and surrounding environmental events as correlations.

In conclusion, this study showed ways of comparing physiological data with individuals' self-reports of their emotional state and an outside observer's report of their corresponding behaviors during the same "check-in" times. This study revealed a strong positive correlation between the participants' self-report of their emotional state during the check-ins across their afternoon and another observer's direct measurement of behaviors corresponding with the participant's emotional indicators. With the third variable of physiological data, it was shown that the Fitbit Charge 5 device was reliable technology to capture the participants' heart rate over time and across physical activity, however, the heart rate data mostly captured changes within the participants' activity level as opposed to changes in heart rate correlating with their emotional ratings outside of active motor movements. With more research on the correspondence of these three measures, more information relating to these events together could be beneficial for helping individuals observe and manage their own behavioral and physiological responses as they occur across their day, such as recognizing changes in their heart rate when experiencing a variety of symptoms related to their emotional states (e.g., stress, anxiety, excitedness, a neutral state or affect, etc.). This information could be useful in the creation of self-management programs where one would analyze more than one

variable (such as overt behavior) changing over time to add more validity to their personal goals or treatment program.

Appendix A

Behavioral Indicators Data Sheet (for Observers)

Jared

Use the descriptive scale below to rate the behavioral affect of the student/client at the specified check-in time. Circle only one number per check-in. Write the time of the check-in and circle either "AM" or "PM" corresponding to the time of day. In the "Activity/Context" section, briefly write what was going on in the student's/client's environment or what s/he was engaged in at the time of the check-in. Additional information can be provided in the 'Notes' section if desirable.

Target Behavior(s): Perseverative speech; Verbal profanity/aggression; Physical Aggression

The topographical definitions of each of these behaviors were included on the researchers' data sheet within the rating scale

1	the child shows signs of being happy or excited as evidenced by smiling and laughing or engaging in physical motor activities or positive social interactions with others (e.g., talking about preferred topics, playing a game of chase, etc.).
2	the child's body is relaxed, calm, and he is engaged in the activity at hand without refusals.
3	the child is showing signs of agitation, discomfort, or anxiety as evidenced by repeatedly asking for something that he was told is unavailable or perseverating on a topic that he was told not to perseverate on.
4	the child is showing signs that he is upset as evidenced by speaking in hyperbolic negative statements. (Ex: You are the worst mom. I wish you were not my mom, etc.) or vocal refusals of the task at hand.
5	the child is showing signs of distress by swearing, displaying physical aggression toward another person, or yelling.

Date: _____

Time	Activity/Context	Circle Rating	Notes
AM/PM		**1 2 3 4 5**	
AM/PM		**1 2 3 4 5**	
AM/PM		**1 2 3 4 5**	
AM/PM		**1 2 3 4 5**	
AM/PM		**1 2 3 4 5**	

Appendix B

Behavioral Indicators Data Sheet (for Participant)
Jared

Circle the image/number the best represents how you are feeling when your counselor says it is time to "check-in."

1	I feel happy or excited.
2	I feel calm, relaxed, and engaged in what I'm doing.
3	I feel confused and/or anxious as to why I cannot have what I want
4	I feel like I might lose control and say something mean to someone else
5	I feel out of control of my emotions, like I want to hit someone or yell/curse at somebody

Date: _____

Time	Circle One				
_____AM/PM	**1**	**2**	**3**	**4**	**5**
_____AM/PM	**1**	**2**	**3**	**4**	**5**
_____AM/PM	**1**	**2**	**3**	**4**	**5**
_____AM/PM	**1**	**2**	**3**	**4**	**5**
_____AM/PM	**1**	**2**	**3**	**4**	**5**

Appendix C

Behavioral Indicators Data Sheet (for Observers)
Charlie

Use the descriptive scale below to rate the behavioral affect of the student/client at the specified check-in time. Circle only one number per check-in. Write the time of the check-in and circle either "AM" or "PM" corresponding to the time of day. In the "Activity/Context" section, briefly write what was going on in the student's/client's environment or what s/he was engaged in at the time of the check-in. Additional information can be provided in the 'Notes' section if desirable.

Target Behavior(s): Frustration (screaming/yelling; catastrophizing a situation by verbal expressions/exaggerated statements that the activity is unwanted or impossible).
The topographical definitions of each of these behaviors were included on the researchers' data sheet within the rating scale

1	the child shows signs of being happy or excited as evidenced by smiling and laughing or engaging in physical motor activities or positive social interactions with others (e.g., talking about preferred topics, playing a game of chase, etc.).
2	the child's body is relaxed, calm, and he is engaged in the activity at hand without refusals.
3	The child is showing precursors to frustration by quieting his voice or not responding to other people; may withdraw briefly from the space
4	The child is escalating by whining or repeating questions/statements about the contextual event (e.g., change in routine, making mistake)
5	The child is experiencing heightened levels of frustration as indicated by screaming or yelling, property destruction (ripping materials, throwing them away, etc.), or using exaggerated statements/catastrophic descriptions of a situation where that response does not match the actual level of difficulty or contextual events (e.g., "This is the worst thing ever!"). Crying may also occur.

Date: _____

Time	Activity/Context	Circle Rating	Notes
AM/PM		**1 2 3 4 5**	
AM/PM		**1 2 3 4 5**	
AM/PM		**1 2 3 4 5**	
AM/PM		**1 2 3 4 5**	
AM/PM		**1 2 3 4 5**	

Appendix D

Behavioral Indicators Data Sheet (for Participant)
Charlie

Circle the image/number the best represents how you are feeling when your parent/teacher says it is time to "check-in."

1	I feel happy and/or excited.
2	My body feels calm and relaxed and I feel engaged in what I'm doing.
3	I feel a little nervous or upset about what is happening.
4	I feel like I might lose control of my emotions and I don't want to talk to anyone or I might explode ("closed off")
5	I feel very frustrated and I like everything is going wrong; I feel like I want to shout at the top of my lungs or I feel like crying.

Time	Circle One				
_____AM/PM	**1**	**2**	**3**	**4**	**5**
_____AM/PM	**1**	**2**	**3**	**4**	**5**
_____AM/PM	**1**	**2**	**3**	**4**	**5**
_____AM/PM	**1**	**2**	**3**	**4**	**5**
_____AM/PM	**1**	**2**	**3**	**4**	**5**

Appendix E

Treatment Integrity Checklist for Researchers

Date: _____ Setting: _____ Researcher implementing the treatment: _____

Additional scoring columns are placed on the original data sheet to fit six fidelity checks per session

Step	Circle Rating
1. The researcher checked the participant's Fitbit wristband to indicate that is turned on, charged, and functioning properly (display settings are showing)	+ -
The researcher checked the position of the Fitbit on the participant's wristband to assure the following:	
2. The Fitbit is on the participant's wrist at the time of the check-in/data collection period	+ -
3. The wristband fits snuggly on the participant's wrist (allowing one of the participant's fingers to fit between the wristband and their wrist, and the wristband is secured through the loops and notch *(the wristband does not move loosely when participant shakes his/her wrist)*	+ -
4. The main screen of the wristband rests at one-finger's distance from the participant's wrist bone and is displayed on the outside of the wrist	+ -
5. The researcher has a copy of the *Behavioral Indicators Data Sheet for Observers* and a copy of the *Behavioral Indicators Data Sheet for Participants* for the participant	+ -
6. The researcher instructed the participant to complete their self-check in at the designated time (e.g., "It's time to do your check-in")	+ -
7. The researcher took data at the same time of the participant's check-in for each check-in period	+ -
8. The researcher's data sheet cannot be seen by the participant (the participant should not be able to see what number the researcher circled)	+ -
9. The researcher filled in the time of each check-in	+ -
10. The researcher circled one number on the rating scale of the data sheet per check-in	+ -
11. The researcher wrote the activity/context in which the data collection took place at the time of each check-in	+ -
12. The researcher wrote supplemental information in the "Notes" section of the data sheet as indicated on the sheet	+ -
13. The researcher checked that the participant completed all sections of their data sheet at the check-in times (writing the time and circling a number)	+ -

14. **The researcher did not provide any indication to the participant of what number s/he should circle for their self-rating	+ -

**If the participant indicates that s/he needs help with their check-in rating or is unsure of what to put (e.g., "I don't know what number to circle," "I'm not sure," or something similar), the researcher should answer with a neutral statement, saying "Put your best guess."

Although the participant will be of reading age to read the indicators scale on their data sheet, the researcher can read the full scale to the participant if the participant desires as a reminder of what each number means. **The researcher should not encourage or advocate for the participant to circle a particular number, as this will compromise the integrity of the participant's self-report information.**

References

Baer, D. M., Wolf, M. M., & Risley, T. R. Risley. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1*(1), 91-97.

Barrera, F. J., Violo, R. A., & Graver, E. E. (2007). On the form and function of severe self-injurious behavior. *Behavioral Interventions*, *22*(1), 5–33. https://doi.org/10.1002/bin.228

Brigham, T. A. (1983). Self-management: A radical behavioral perspective. In P. Karoly & F. H. Kanfer (Eds.), *Self-management and behavior change: From theory to practice* (pp. 32-59). New York: Pergamon Press.

Catania, A. C. (2013). *Learning*. (5th Edition). Cornwall-on-Hudson, NY: Sloan Publishing.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis*. 3rd ed. Pearson.

Critchfield, T. S., Tucker, J. A., & Vuchinich, R. E. (1998). Self-Report Methods. *Handbook of Research Methods in Human Operant Behavior* (pp. 435-470). Plenum Press.

Feely, M., Seay, K. D., Lanier, P., Auslander, W., Kohl, P. L. (2017). Measuring fidelity in research studies: A field guide to developing a comprehensive fidelity measurement system. *Child and Adolescent Social Work Journal*, *35*(2), 139-152. doi: 10.1007/s10560-017-0512-6

Friman, P. C., Hayes, S. C., & Wilson, K. G. (1998). Behavior analysis of private events is possible, progressive, and non-dualistic: A response to Lamal. *Journal of Applied Behavior Analysis*, *31*(4), 707-708.

Friman, P. C., Hayes, S. C., & Wilson, K. G. (1998). Why behavior analysts should study emotion: The example of anxiety. *Journal of Applied Behavior Analysis, 31*(1), 137-156.

Freeman, R. L., Horner, R. H., & Reichle, J. (1999). Relation between heart rate and Problem behaviors. *American Journal on Mental Retardation*, *104*(4), 330-345. doi: 10.1352/0895-8017(1999)104<0330:RBHRAP>2.0.CO;2

Hoch, J., Moore, T., McComas, J., & Symons, F. J. (2010). Arousal and activity choice in autism: A single-case assessment integrating autonomic and behavioral analysis. *Journal of Applied Biobehavioral Research, 15*(3), 119-133. https://doi.org/10.1111/j.1751-9861.2010.00056.x

Hughes, C. & Lloyd, J. W. (1993). An analysis of self-management. *Journal of Behavioral Education*, *3*(4), 405-425.

Johnston, J. M., Pennybacker, H. S., & Green, G. (2020). *Strategies and tactics of behavioral research and practice*. (*4)*, Routledge.

Kostewicz, D. E., Kubina Jr., R. M., & Cooper, J. O. (2000). Managing aggressive thoughts and feelings with daily counts of non-aggressive thoughts and feelings: A self-experiment. *Journal of Behavior Therapy and Experimental Psychiatry*, *31*, 177-187.

Kubina Jr., R. M., Haertel, M. W., & Cooper, J. O. (1994). Reducing negative inner behavior of senior citizens: The one-minute counting procedure. *Journal of Precision Teaching*, *11*(2), 28-35.

Lamal, P. A. (1998). Advancing backwards. *Journal of Applied Behavior Analysis*, *31*(4), 705-706.

Lydon, S., Healy, O., & Dwyer, M. (2013). An experimentation of heart rate during challenging behavior in autism spectrum disorder. *Journal of Developmental and Physical Disabilities, 25*, 149-170. doi: 10.1007/s10882-012-9324-y

Morsbach, S. K. and Prinz, R. J. (2006). Understanding and improving the validity of self report of parenting. *Clinical Child and Family Psychology Review, 9*(1), 1-21. doi: 10.1007/s10567-006-0001-5

Moskowitz, L. J., Mulder, E., Walsh, C. E., McLaughlin, D. M., Zarcone, J. R., Proudfit, G. H., & Carr, E. G. (2013). A multimethod assessment of anxiety and problem behavior in children with autism spectrum disorders and intellectual disability. *American Journal of Intellectual and Developmental Disabilities, 118*(6), 419-434. doi: 10.1352/1944.7558.118.6.419

Moskowitz, L. J., Walsh, C. M., Mulder, E., McLaughlin, D. M., Hajcak, G., & Carr, E. G. (2017). Intervention for anxiety and problem behavior in children with autism Spectrum disorder and intellectual disability. *Journal of Autism and Developmental Disorders, 47*, 3930-3948. doi: 10.1007/s10803-017-3070-z

Sidman, M. (2008). Reflections on stimulus control. *The Behavior Analyst, 31*, 127-135.

Simoës-Perlant, A., Lemercier, C., Pêcher, C., & Benintendi-Medjaoued, S. (2018). Mood self-assessment in children from the age of 7. *Europe's Journal of Psychology*, *14*(3), 599-620. doi: 10.5964/ejop.v14i3.1408

Skinner, B. F. (1964). Behaviorism at fifty. In T. W. Wann (Ed.) *Behaviorism and phenomenology*, 79-97. Chicago: University of Chicago.

Skinner, B. F. (1953). *Science and human behavior.* New York, NY: The Macmillan Company.