Title

Assessing Quality of Function-Based Behavior Supports: Reliability, Concurrent Validity, and Predictive Validity of the Technical Adequacy Tool for Evaluation

Permalink

https://escholarship.org/uc/item/4w54n8dw

Author

Hinant, Rachel

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Assessing Quality of Function-Based Behavior Supports: Reliability, Concurrent
Validity, and Predictive Validity of the Technical Adequacy Tool for Evaluation

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Education

by

Rachel Hinant

March 2022

Dissertation Committee:
      Dr. Austin Johnson, Chairperson
      Dr. Rondy Yu
      Dr. Stephanie Moore

The Dissertation of Rachel Hinant is approved:

_____

_____

_____
                                    Committee Chairperson

University of California, Riverside

Acknowledgement

I am thankful to my advisors Dr. Austin Johnson and Dr. Cathleen Geraghty for each guiding and supporting me through the program and navigating all the challenges that arose along the way. I would not have completed this dissertation without their help. I would also like to thank the members of my committee Dr. Rondy Yu and Dr. Stephanie Moore for their generosity with their time and support.

ABSTRACT OF THE DISSERTATION

Assessing Quality of Function-Based Behavior Supports: Reliability, Concurrent Validity, and Predictive Validity of the Technical Adequacy Tool for Evaluation

by

Rachel Hinant

Doctor of Philosophy, Graduate Program in Education
University of California, Riverside, March 2022
Dr. Austin Johnson, Chairperson

Student problem behavior impacts instruction, learning, and teacher stress; furthermore, teachers are often underprepared to manage student behavior effectively (Clunies-Ross et al., 2008; Flower et al., 2017). However, behavior interventions in schools are often not based on the function of behavior (Blood & Neel, 2007; Cook et al., 2012). The Technical Adequacy Tool for Evaluation (TATE; Iovannone & Romer, 2017) is one rating tool designed to help assess the quality of Behavior Intervention Plans (BIPs), which may help school practitioners measure and improve the quality of behavior interventions for students. The present study evaluated the relationship between BIP quality measured by the TATE and individual behavior outcomes as well as parent ratings of adaptive and problem behaviors. A strong positive correlation coefficient was observed between BIP scores on the TATE and BIP-QE II (Browning-Wright et al., 2013) rating tool. A Poisson regression was used to evaluate whether TATE items 12 through 15 were predictive of frequency of problem behavior after intervention; the TATE item rating presence of a strategy to provide functionally equivalent reinforcement to a replacement behavior was predictive of change in frequency of problem behavior.

None of the TATE items 12 through 15 were a significant predictor of change in adaptive behavior (i.e., ABAS3 scores) or behavior problems (i.e., PDDBI scores). Implications for future research as well as importance of functionally-equivalent reinforcers for appropriate behaviors are discussed.

Table of Contents

List of Tables

List of Figures

Assessing Quality of Function-Based Behavior Supports: Reliability, Concurrent Validity, and Predictive Validity of the Technical Adequacy Tool for Evaluation

Student behavior problems are a common challenge for teachers across all grade levels in both general and special education settings. Teachers frequently report needing to address problem behaviors; almost half (47.4%) of 97 teachers surveyed reported they address student problem behaviors five or more times in a typical day (Clunies-Ross et al., 2008). Talking out of turn was reported as the most frequent and troublesome behavior exhibited by both male and female students (Clunies-Ross et al., 2008). Difficulty managing student problem behavior is related to teacher stress, burnout, and the decision to leave the profession altogether (Evers et al., 2004; Gold, 1985).

Limited behavior support training is available to teachers through teacher education programs, professional development, or consultation after they begin teaching. Training programs, if they provide training in behavior management, more often address universal and topographically-focused strategies rather than teaching strategies to increase appropriate behavior and decrease inappropriate behavior (Flower et al., 2017). Similarly, one survey of teachers found that they report limited confidence and little implementation of individualized strategies, and that they are more confident with universal strategies (Moore et al., 2017). Even with some preparation and training in behavior management, a majority (84%) of teachers wanted more preparation during teacher training, and many (64%) wanted more ongoing professional development in behavior management. Teachers voice a clear need for support and training in how to manage frequent student problem behavior.

When teachers do attempt to manage student behavior, they often select interventions based on the topography of behavior rather than on the hypothesized function of behavior. Given scenarios with information that made the function of behavior apparent, few teachers considered the function of behavior in picking an intervention (Myers & Holland, 2000). Teachers untrained in behavioral assessment inaccurately identify function when simply asked about the possible function of student behavior, compared to a more formal analysis (Calloway & Simpson, 1998). Lack of training on individualized or function-based interventions in teacher education programs and teacher reliance on other teachers for problem-solving and intervention strategies contribute to the phenomenon of teachers unable to adequately address behavior challenges (Myers & Holland, 2000; Wilson et al., 1998). Practicing teachers are in need of consultative support and training to better manage student problem behavior and promote behaviors which contribute to learning and positive outcomes for students.

Applied Behavior Analysis

Many evidence-based strategies for managing student behavior have been developed based on applied behavior analysis (ABA). This theoretical framework assumes that behavior is lawful and can be reliably predicted and understood through its relation to environmental stimuli that precede and follow behavior (Cooper et al., 2007). The lawfulness of behavior allows for the development of strategies that create behavior change by changing the environmental contingencies around the targeted behaviors. Using ABA strategies, teachers can decrease student problem behavior and increase

appropriate behavior by changing what happens before and after the student engages in targeted behaviors (Goh & Bambara, 2012).

Board Certified Behavior Analysts (BCBAs) are individuals who are trained in the use of ABA. While some BCBAs may work in education settings providing settings to students, the majority report that they primarily provide services to people with autism spectrum disorders (Association of Professional Behavior Analysts, 2015). These services may be provided through clinical settings or in-home services. Additionally, BCBAs report working across a wide variety of settings such as with persons with intellectual and developmental disabilities, in higher education, behavioral pediatrics, child welfare, and corrections and delinquency (Association of Professional Behavior Analysts, 2015).

Positive behavior supports (PBS) provide a framework for applying ABA strategies to socially significant behavior changes (Sugai et al., 2000) in school settings. Traditional school approaches to behavior problems often are zero-tolerance, exclusionary practices (e.g., office referral, suspension, expulsion) that are many times ineffective, applied disproportionately to certain groups of students (e.g., African-American boys, students with disabilities), and put students at higher risk of dropping out of school (Government Accountability Office, 2018; Losen et al., 2015; Chu & Ready, 2018). PBS uses behavioral principles to explicitly teach and reinforce socially acceptable behaviors while systematically removing reinforcement for problem behaviors, in contrast to only reacting punitively to problem behaviors after they occur. To do this effectively, PBS requires the collection and analysis of behavioral data prior to

developing behavior supports, during implementation, and following implementation to evaluate effectiveness (Sugai et al., 2000).

Multi-Tiered Systems of Support for Behavior

Within the school system, PBS can be part of a multi-tiered system of support (MTSS) for behavior, allowing for careful resource allocation across an intensifying continuum of supports (Scott et al., 2010). One example of MTSS for behavior is school-wide positive behavior supports (SWPBS), a tiered approach which improves overall student behavioral outcomes by providing universal, targeted, and intensive intervention strategies to students depending on their need and allows for student movement across tiers as behavior improves or worsens (Horner et al., 2009). Essential to all levels of MTSS for behavior is data on problem behaviors, interventions that are highly probable to be efficient and effective, consistency of implementation and practices across all faculty and staff, and ongoing progress monitoring used for data-based decision making (Lewis & Sugai, 1999; Scott et al., 2010).

Universal support, or Tier 1, involves teaching and reinforcing appropriate behaviors for all students in a school (Scott et al., 2010). At the universal level, behavioral expectations are defined and taught for classrooms and entire schools. Reinforcement, often in the form of social praise and generalized reinforcement, is provided to students who engage in behaviors aligned with school expectations. Data collected at the universal level can include office disciplinary referrals (ODRs) or direct observations (Farkas et al., 2012; Scott et al., 2010). Those data can be used to evaluate

effectiveness of universal supports and potential areas for improvement, as well as to identify students in need of more targeted or intensive supports.

Effective Tier 1 plans include effective classroom behavior management strategies, such as providing behavior specific praise, establishing classroom routines, arranging the physical classroom environment to support appropriate behavior, and providing precorrection and reminders to students when appropriate (Scott et al., 2007). Improving classroom management and schoolwide behavior management supports all students in engaging in appropriate behaviors and minimizing inappropriate behaviors (Bradshaw et al., 2010). By improving the school environment for all students and promoting appropriate behaviors, MTSS works in part as a prevention model, providing preventive support for students before behavior problems become more severe with an established and lengthy learning history. Behaviors are learned, and repeated exposure to a contingency can strengthen that learned behavior over time, making it more difficult to remediate later on (Cooper et al., 2007). Effective Tier 1 supports can help prevent problem behaviors from developing or worsening by removing a student's need to engage in inappropriate behaviors.

Targeted supports, commonly referred to as Tier 2, provide intervention for students who need additional strategies to improve behavior. Data (e.g., office disciplinary referrals, teacher nomination) are used to identify students for Tier 2 interventions. Typical Tier 2 interventions should be high-probability interventions (i.e., likely to succeed based on research, demonstrate match to the student and problem behavior) based on data collected about targeted problem behaviors (Scott et al., 2010).

Interventions are based in the principles of ABA, and can involve improving student-teacher relationships, designing effective instruction, and classroom management (Scott et al., 2010). Commonly used Tier 2 interventions with demonstrated effectiveness include check-in/check-out (CICO; Hawken & Horner, 2003), positive peer reporting or tootling (Skinner et al., 2000), and daily behavior report cards (Vannest et al., 2010).

For students who continue to exhibit problem behaviors after Tier 2 interventions have been implemented, more intensive and individualized interventions can be implemented as part of Tier 3. Tier 3 supports involve data collection on observable, measurable behaviors and development of an individualized intervention with strategies to teach and reinforce socially acceptable replacement behaviors, while decreasing target problem behaviors. Often this is accomplished by conducting a functional behavior assessment (FBA) and developing a behavior intervention plan (BIP). An FBA is a flexible assessment procedure to gather information about environmental contingencies around a target problem behavior or behaviors resulting in a testable hypothesis about the function of the behavior (Cooper et al., 2007). The function of a behavior refers to the events that temporally surround the behavior (i.e., antecedents, consequences). Possible functions of behavior include: to obtain attention, to obtain access to a tangible item or activity, to escape or avoid attention, to escape or avoid an item or activity, or self-stimulation (Cooper et al., 2007). The BIP is a document describing specific strategies to be used to make the problem behavior less efficient and less effective, while also teaching a functionally equivalent replacement behavior (FERB) that allows the individual to access the same function as the problem behavior in a socially acceptable

way (O'Neill et al., 2015). When the BIP is implemented, data collection on the problem behavior and the replacement behavior are collected to guide decision making around potential modifications necessary to improve the BIP, plans to gradually withdraw the BIP, or the need for a new FBA and BIP.

Function-Based Intervention

Interventions that are based on the hypothesized function of behavior are more effective than interventions not related to function (Goh & Bambara, 2012; Hurl et al., 2016). This is true both of interventions that are completely individualized and developed for a specific student, as well as for manualized interventions, which may need slight modification to address function (Campbell & Anderson, 2008; Carter & Horner, 2009). Whether an intervention is completely individualized (e.g., a BIP) or is a modified manualized intervention, data about the behavior and the environment must be gathered and analyzed in order for the intervention to be matched to the hypothesized function of problem behavior.

Function-Based Tier 1. Typically, Tier 1 interventions utilize generalized conditioned reinforcers that are likely to be effective for most students in most circumstances. Generalized reinforcers have been paired with many different reinforcers, conditioned and/or unconditioned, over time and as a result of this learning history are effective as a reinforcer for most people at most times (Cooper et al., 2007). Unconditioned reinforcers (e.g., food, water, warmth) are reinforcers for all people without any learning history, but function as reinforcers only when the individual in question is not satiated (e.g., food is not a reinforcer for someone who has just eaten a

large meal). Generalized reinforcers, like social praise, are reinforcing for most humans and aren't impacted by satiation or deprivation in the same manner as unconditioned reinforcers (Cooper et al., 2007). Many Tier 1 interventions utilize token systems where students earn some form of token or point and can exchange that for their choice of reward from a menu of different options (Farkas et al., 2012). These token systems allow for immediate reinforcement of appropriate behavior with the token which is paired with later access to a different reinforcer, such as a toy or treat (Cooper et al., 2007).

Function-Based Tier 2. Manualized interventions such as CICO, a typical Tier 2 intervention, have research to demonstrate that they are often effective but can be improved when modified to address behavioral function. Campbell and Anderson (2008) modified a CICO intervention that had already been implemented for two 10-year-old boys. CICO interventions typically provide access to adult attention contingent upon appropriate behaviors but based on data gathered by the authors for an FBA, the hypothesized function of both boys' problem behaviors was peer attention. The authors modified the CICO intervention to provide access to peer attention (e.g., choosing what peer to sit next to at lunch) contingent upon positive behavior during independent reading activities. The modified CICO intervention resulted in fewer instances of problem behavior for both students compared to both baseline and the original CICO intervention (Campbell & Anderson, 2008).

McIntosh et al. (2009) implemented an unmodified Tier 2 CICO intervention for 34 elementary aged students. The authors identified the hypothesized function of problem behavior through teacher interviews, and found that for students whose problem behavior

8

was hypothesized to be maintained by attention, CICO was effective in reducing teacher reporting of problem behavior and frequency of office disciplinary referrals (ODRs). For students with escape-maintained behavior, there were no statistically significant effects of the unmodified CICO intervention on student problem behavior, prosocial behavior, or frequency of ODRs (McIntosh et al., 2009). Manualized interventions at the Tier 2 level may be ineffective if they do not address the function of the student's problem behavior.

Function-Based Tier 3. Tier 3 interventions, whether manualized or individually developed, can also be improved when they are modified or designed to address behavioral function. Carter and Horner (2009) put in place modifications to a First Step to Success program, which typically provides adult attention contingent upon appropriate behaviors, to address problem behavior function for three kindergarten students. Using direct observation data, the authors hypothesized that the function of problem behavior for two students was peer attention and for the third was peer attention and escape from difficult academic tasks. The First Step to Success intervention was modified for each of the students to address hypothesized function (e.g., increase peer attention for appropriate behavior, escape task for appropriate behavior, decrease attention for problem behavior). All three students engaged in problem behavior less frequently when the modified intervention was implemented as compared to the unmodified intervention (Carter & Horner, 2009).

Individualized interventions developed using ABA techniques are also more effective when designed to address the function of problem behavior. Briere and Simonsen (2011) conducted FBAs for two middle school students with persistent

problem behaviors that hadn't been ameliorated by prior intervention attempts (e.g., CICO). The FBAs suggested that one student's problem behavior was maintained by attention, while the other student's problem behavior was maintained by escape from academic tasks. The authors designed self-monitoring interventions that provided attention or escape and implemented each for both students. For each student, problem behavior was less frequent when the functionally-relevant intervention was in place compared to baseline and the functionally non-relevant intervention (Briere & Simonsen, 2009).

Filter and Horner (2009) also conducted FBAs to assess the function of two fourth grade students' problem behaviors and developed a function-based intervention for each student. There were greater reductions in problem behavior during periods when the function-based interventions were implemented than when there was no intervention and when there was a non-function-based intervention in place (i.e., reasonable interventions successful for other students but that did not change environmental contingencies around problem behavior). Ellingson et al. (2000) followed similar procedures to collect data on problem behavior for three students with intellectual disabilities (ID) and subsequently develop intervention strategies. For each student, an intervention which matched the hypothesized function of behavior was implemented, alternating with an intervention that addressed an alternative hypothesis (i.e., did not match function). Problem behavior decreased for each student when the function-based intervention was implemented with fidelity (Ellingson et al., 2000). Interventions that address the function of a student's

problem behavior are more effective at reducing problem behavior than interventions that do not address function.

Functional Behavior Assessment

In order to develop an intervention that addresses behavioral function, data pertaining to the potential function of a problem behavior are necessary. An FBA can include data gathered through indirect methods, such as interviews, rating scales, and questionnaires, and through direct methods of observation of behavior in a natural setting such as antecedent, behavior, consequence (ABC) data (Cooper et al., 2007). A complete FBA contains information in six key areas: description of the problem behavior, antecedents that predict the problem behavior, setting events that predict the problem behavior, consequences that maintain the problem behavior, a summary statement or hypothesis about the behavior, and data that support that hypothesis (O'Neill et al., 2015). These six areas can also be grouped into three main outcomes: description of the problem behavior, prediction (e.g., antecedents, setting events), and function (i.e., maintaining consequences; Adams & Dunsmuir, 2009). Information on what the target behavior is, and what in the environment predicts and maintains the behavior, can be used when developing an intervention that addresses the function of problem behavior to effectively decrease problem behavior and increase socially acceptable and desired behaviors.

Some differences exist in the literature around how the process of FBA is conceptualized. Some researchers describe FBA as a process including both indirect and direct methods of data collection (Cooper et al., 2007), while others conceptualize

indirect FBA and direct FBA as distinct processes (Fee et al., 2016). Additionally, researchers differ in how functional analysis (FA) relates to FBA. FA is an experimental method for determining the function of a behavior and involves manipulating the environment while measuring the target behavior (Iwata et al., 1982/1994). In its most widely-known iteration, an FA typically consists of measuring behavior while changing what consequence contingently follows behavior across four conditions (i.e., access to tangibles, task demands, social attention, alone) in order to identify the hypothesized function of behavior (Iwata et al., 1982/1994). Some researchers treat FBA and FA as two distinct assessment frameworks (Camp et al., 2009), while others treat FA as an optional additional piece of an FBA that can be included to corroborate the FBA hypothesis or clarify cases where the FBA suggests multiple functions (Cooper et al., 2007; Martens & Lambert, 2014; O'Neill et al., 2015; Steege & Scheib, 2014).

Effectiveness of FBA. Meta-analyses of studies using FBAs generally indicate that behavioral interventions developed using FBA data are effective, with overall effect sizes for FBA-based interventions implemented across settings, including school and clinical settings, in the moderate to high range. Bruni and colleagues (2017) combined 30 single case design studies of behavioral interventions, some including an FBA and some not, and found an overall Tau-U effect size of .75. The small but insignificant difference between interventions preceded by an FBA and those without an FBA may be due to their inclusion of interventions like cognitive behavioral therapy (CBT) which does not consider function, but which is nonetheless an evidence-based intervention (Ishikawa et al., 2007; Weisz et al., 2006). Hurl et al. (2016) analyzed 13 single case design studies

with a comparison of interventions based on FBA data with interventions not based on function and found an overall effect size of .85. Goh and Bambara (2012) analyzed studies of FBA based interventions in different settings and with different student populations and found FBA based interventions to have a strong effect on student behavior change. Overall, evidence supports the assertion that interventions that use data from an FBA to address the function of student problem behavior are effective in reducing problem behavior across different settings, including clinical and school settings.

The studies included in meta-analyses of FBA effectiveness represent a variety of FBA methodologies which, based on current evidence, do not have a significant impact on effectiveness. Bruni et al. (2017) compared effect sizes for studies employing indirect or descriptive FBAs (i.e., including indirect methods as well as direct observation but no experimental methods) to studies that included an FA and found both studies with descriptive FBAs and those with FAs had equivalent Tau-U effect sizes of .79. Goh and Bambara (2012) compared effect sizes for studies with FBAs with only descriptive methods (e.g., interviews & direct observations), FBAs with only experimental methods (e.g., functional analysis, analogue assessment), and FBAs using both methods. The percent of non-overlapping data (PND) effect size for each type of FBA method was similar: 88% for descriptive only, 87% for experimental only, and 91% for a combination of both. FBAs employing descriptive data collection and those employing experimental methods are both effective ways to develop a hypothesis about behavioral function to be used in intervention, though there may be differences (e.g., severity of problem behavior)

that impact which methods are most effective for different cases. In schools, it may be expected to be effective in most cases to conduct a descriptive FBA consisting of indirect methods like interviewing stakeholders and reviewing records as well as direct methods like directly observing the student, in order to develop a hypothesis statement about the function of the behavior. Interventions based on hypothesized function of behavior derived from an FBA, including interventions developed as part of a BIP, are effective at reducing problem behaviors and increasing appropriate behaviors.

Development and Implementation of Function-Based Supports

A variety of personnel can be tasked with conducting an assessment and developing an intervention or modifying an existing intervention for students with behavior problems. No clear recommendations have emerged from the literature about who should be responsible for function-based supports. In some studies, specialists who are not based in the classroom (e.g., board certified behavior analysts, school psychologists) are responsible for conducting FBAs and developing a corresponding intervention (e.g., Campbell & Anderson, 2008). In other cases, teams which include specialists, administrators, and teachers together collect and analyze FBA data and develop interventions (e.g., Benazzi et al., 2006).

Legal Requirements

Legal guidance regarding the FBA/BIP process is sparse and pertains primarily to students receiving special education services. The Individuals with Disabilities Education Act (IDEA; Individuals with Disabilities Education Act, 2004) requires individualized education plan (IEP) teams to consider the use of strategies that may include positive

behavioral interventions and supports in cases where a student's problem behavior impedes their learning or the learning of other students (Drasgow & Yell, 2001). IDEA also requires the IEP team to meet regarding conducting an FBA and writing a BIP within 10 days of a student being removed from school for more than 10 days, placed in an interim alternative setting due to a drug or weapon offense, or removed in a way that would constitute a change in placement (Drasgow & Yell, 2001). IDEA does not provide guidance or requirements regarding who is responsible for conducting an FBA or developing a BIP, how an FBA is to be conducted or a BIP written, or what is necessary to include in the FBA and BIP (Collins & Zirkel, 2017). State legislation is similarly limited; some states provide additional guidance, but this is typically limited to naming elements to include (e.g., "function" included in FBA, "interventions" included in BIP) rather than clear specification about how FBAs should be conducted and BIPs written (Zirkel, 2011). Best practices in how to complete FBAs and BIPs are not currently reflected in federal and state legislation (Collins & Zirkel, 2017; Zirkel, 2011).

In California, the 1990 Hughes Bill (A.B. 2586), part of California Education Code, was passed to establish requirements for how schools responded to students with behavior problems (Wright & Gurman, 2001). The Hughes Bill required that, for any child with exceptional needs with behavior challenges, a functional analysis assessment (FAA) should be conducted and used to develop a behavior intervention plan. The Hughes Bill required that FAAs include information gathered through direct observation, interviews, and reviews of records. FAAs were required to include systematic observation of the target behavior as well as antecedents and consequences, ecological

analysis of the setting, review of any record of health or medical factors, and review of any past interventions for effectiveness (Wright & Gurman, 2001). FAAs did not require experimental methods such as an FA or analogue assessment. The Hughes Bill also set up requirements around who was required to complete FAAs; FAAs were to be conducted by a behavior intervention case manager (BICM) or someone supervised by a BICM. A BICM was any person designated by the school district as someone trained in positive behavioral interventions, typically someone already employed by the district. The BICM requirement did not set up any new credentialing or degree requirements, like the Board Certified Behavior Analyst (BCBA) credential. The qualifications necessary to be a BICM were to be decided by each special education local planning agency (SELPA; Wright & Gurman, 2001).

The Hughes Bill defined behavior intervention plan (BIP) as a written document developed by the IEP team, including the BICM, that becomes part of the IEP. A BIP needed to include a summary of the FAA, a description of the target behavior that was objective and measurable, goals and objectives specific to that individual's behavior plan, a detailed description of behavior interventions to be used, specific schedules for data collection, criteria for fading the intervention, behavior interventions to be used in noneducational settings, and specific dates for periodic review (Wright & Gurman, 2001). The Hughes Bill also prohibited the use of interventions that could cause physical pain or excessive emotional trauma, and interventions using physical restraints unless performed by trained personnel in an emergency situation. The Hughes Bill stayed part of California law until its repeal in 2013, and has not been replaced by any other requirements related

16

to behavior interventions beyond what is stated in IDEA. It represents one of very few attempts by state legislatures to introduce substantive requirements around behavioral assessment and intervention in the schools.

Case law has typically focused on procedural rather than substantive aspects of FBAs and BIPs, often only requiring that districts have considered an FBA and BIP in order to meet the requirements of a free and appropriate public education (FAPE) for students (Collins & Zirkel, 2017). Courts typically will not rule on whether an FBA or BIP is technically adequate or appropriate, often siding with the district in cases where a parent has argued that an FBA or BIP was not appropriate (Collins & Zirkel, 2017). In the case *Alex R. v. Forestville Valley Community Unit School District* (2004), a parent argued the BIP put in place for their child was inappropriate and did not meet substantive requirements. The court's argument in favor of the district was that there are no substantive requirements for BIPs in the law, preventing the court from judging BIP adequacy.

Best Practice

Responsibility for carrying out an FBA and writing a BIP, or implementing other function-based interventions, could vary across a number of factors. A high-severity behavior could necessitate a practitioner with specialist-level knowledge and experience, while strained resources may leave responsibility for behavior management to school staff with more limited training. In a tiered system, the number and training of personnel designated to support with student behavior management may vary across tiers to maximize efficiency of available resources. Clear guidance around who is necessary and

who is unable to conduct behavioral assessment and develop function-based supports is not defined in the literature.

A study by Benazzi and colleagues (2006) suggests that both a person with specialist level knowledge of ABA and a person with knowledge of the classroom setting where the intervention will be implemented are required for the best outcomes. Benazzi and colleagues compared BIPs written by teams that included only members with technical knowledge, only members with classroom knowledge, and both members with technical knowledge and members with classroom knowledge. BIPs were rated by experts to have high technical adequacy and coded as to whether they were written by teams with only people with technical knowledge or by mixed teams (Benazzi et al., 2006). Contextual fit, or how well a BIP matches the values, skills, and resources of the person(s) who will implement the plan, was higher for teams with members with only classroom knowledge or with both types of knowledge (i.e., technical, classroom). Technical adequacy is important to how likely the plan is to be successful if implemented correctly, while contextual fit represents how well the plan fits in the classroom and how likely the person responsible for implementation (e.g., classroom teacher) is to implement the plan as written. Both technical adequacy and contextual fit are important for behavior supports, and are likely to be higher if the team or dyad responsible for plan development includes someone with specialist knowledge of ABA and someone with classroom knowledge specific to where the plan will be implemented.

The social validity of an intervention in ABA refers to the social significance of the intervention goals, the social appropriateness of the procedure employed, and the

social importance of the effects (Wolf, 1978). Gathering subjective information on the appropriateness of an intervention and feedback from intervention consumers and the community on the importance of the intervention is important in making research and ABA interventions meaningful (Wolf, 1978). Including relevant stakeholders (e.g., teacher, parent, classroom aide, student) in the planning process and taking their views and experiences into account can help to improve the social validity of an intervention, in addition to improving contextual fit. Additionally, variables related to the behavior should be taken into consideration when determining appropriate interventions and significance of the goals, for example selecting interventions that limit or do not include exclusionary or restrictive procedures that would isolate a child from their peers and the learning environment in school, or selecting as a goal behaviors that will allow the child more access to natural reinforcements in most environments.

Social validity is also important due to its relationship with treatment integrity. Treatment integrity is the degree to which an intervention or procedure is implemented as it was designed, and is important both for drawing valid conclusions from research and maximizing the likelihood an intervention will be effective (Lane et al., 2004). Evidence-based interventions, such as behavior plans based on the function of target behavior, are unlikely to be effective if they are not implemented with integrity (Wood et al., 2007). Social validity, or acceptability, has an impact on the contextual fit of a plan and the likelihood it will be implemented with fidelity, although it is not sufficient alone as teachers may rate an intervention as acceptable but still struggle to implement consistently and completely (Benazzi et al., 2006; Wood et al., 2007).

In addition to social validity, quality of the BIP is also an important variable. While a BIP, like an FBA, is a flexible and individualized document, there are some general components that research has indicated should be included. BIPs should indicate how the behavior of those responsible for implementation (e.g., teachers, aides, parents) will change and not focus solely on how the individual's behavior will change (O'Neill et al., 2015). Based on the principles of ABA, an individual will alter their behavior when the environmental contingencies surrounding that behavior change. Changes to environmental contingencies (e.g., a target behavior no longer results in access to a reinforcer) are managed by those responsible for implementing a behavior intervention. These changes must be described clearly enough in the BIP to be understood by those tasked with making the changes.

A quality BIP should also be based on the hypothesized function of behavior identified in the FBA (O'Neill et al., 2015). This link between FBA and BIP should be made explicitly. This could be accomplished by including or paraphrasing the summary statement from the FBA which states the target behavior and hypothesized function(s). It could also be done by including a statement detailing how the plan will ensure that appropriate behaviors "compete" successfully with problem behaviors based on hypothesized function of behavior (O'Neill et al., 2015).

As documents anchored in behavioral theory, BIPs should be based on principles and research from the field of ABA (O'Neill et al., 2015). Examples of common principles include reinforcement, punishment, extinction, generalization, and stimulus control. Additionally, any intervention strategy included in a BIP should be based on

quality research. In general, a technically-sound BIP will make problem behaviors irrelevant, inefficient, and ineffective (O'Neill et al., 2015). Problem behaviors are made irrelevant by removing the child's need to engage in the problem behavior; for example, by adjusting the difficulty of reading passages for a child who typically screams to avoid highly difficult reading assignments. Strategies that make problem behaviors irrelevant are antecedent strategies, which adjust the environment prior to the occurrence of problem behavior to make the problem behavior less likely to occur. Problem behaviors are made inefficient by providing the individual with an alternative behavior which is a more efficient way to access reinforcement. The alternative behavior could be less effortful than the problem behavior or provide access to reinforcement more quickly or more often. Problem behaviors are made ineffective by preventing access to the reinforcer that had previously maintained the behavior. The reinforcer could be systematically withheld when the individual engages in problem behavior or could be temporarily withheld while the individual is prompted to engage in the appropriate behavior instead. By making the problem behavior irrelevant, inefficient, and ineffective, BIPs can result in reductions in problem behavior along with increases in appropriate behaviors. For this reason, it is theoretically important for BIPs to have a combination of antecedent strategies, a functionally equivalent replacement behavior which results in access to the same reinforcement as the problem behavior, and a strategy to remove reinforcement for the problem behavior.

Finally, quality BIPs are designed to fit within the setting where they will be implemented (O'Neill et al., 2015). This includes contextual fit, or how well the plan fits

21

with the values of the person(s) tasked with implementation. It also involves how well the plan fits within the natural rhythms and routines of the environment. BIPs should also be an efficient use of time, money, and resources. For instance, a plan which requires the purchase of expensive equipment or designation of a staff to provide constant, individual attention throughout the day are unlikely to fit well in a school with a small budget and limited staff. BIPs should fit with the skills of those who will be responsible for implementation; a person who is untrained and unsure how to carry out a complicated intervention is unlikely to do so with fidelity. Finally, short term results of the BIP should be reinforcing rather than punishing. If a plan results in an immediate increase in problem behavior, as often occurs with extinction procedures, staff are unlikely to continue to implement the plan as their early efforts have been met with a punishing increase in the behavior they do not want to see. BIPs should attempt to provide for early reductions in problem behavior and increases in appropriate behavior, to reinforce those who are implementing the plan.

FBA and BIP Quality in Practice

FBAs and BIPs developed in schools are often of low quality (Blood & Neel, 2007; Cook et al., 2007; Van Acker et al., 2005). In many cases, BIPs are developed without any form of functional assessment of behavior being completed first (Blood & Neel, 2007). Without an assessment to provide information on hypothesized function of behavior, it is unlikely that BIPs will successfully make problem behavior irrelevant, inefficient, and ineffective. Even when an FBA has been conducted, the majority of BIPs continue to be poor quality and generic (Blood & Neel, 2007). BIPs developed in schools

22

have been found to typically be lists of common behavioral strategies like using a token economy or praising desired behavior. While these strategies are based in ABA research and have evidence to support their use, they should be selected only when they fit for the individual in question and are based on function of behavior. For a student who engages in verbal aggression in order to escape difficult math tasks, praising the child for working on math tasks does not meet the same function (i.e., escape) as the problem behavior and is therefore unlikely to make that problem behavior irrelevant, inefficient, or ineffective. Very few behavior plans developed in schools have been found to include individualized strategies, including antecedent and consequent strategies, based on hypothesized function of problem behavior (Blood & Neel, 2007). Additionally, when the teams who are tasked with developing BIPs do not have training in positive behavior supports, BIPs are of inadequate quality and unlikely to result in positive student outcomes (Cook et al., 2012).

Less information is available regarding the quality of FBAs and BIPs developed in clinical settings. A survey by Roscoe and colleagues assessed functional assessment practices by practitioners working in agencies that serve individuals with developmental disabilities (Roscoe et al., 2015). They found that practitioners most frequently reported conducting descriptive assessments to guide selection of behavioral treatment. A majority of respondents indicated that functional analysis is the most informative assessment tool, though only a third of respondents reported typically using functional analysis to inform behavioral intervention. A survey of BCBAs conducted by Oliver and colleagues found similar results. A majority of respondents reported regularly conducting FBAs, primarily

descriptive methods. Most respondents reported that functional analyses were the most useful method of assessing function of behavior, but less than half of respondents reported using functional analysis in practice (Oliver et al., 2015).

Legal Guidance

While lack of training is one factor impacting quality of FBAs and BIPs developed in school settings, limited legal guidance is also a likely factor. The lack of clear requirements or guidelines in both federal and state legislation as to who is qualified to conduct an FBA and develop a BIP, what methodology should be followed when developing an FBA or BIP, and what information is necessary to include, could be contributing to the inconsistency and poor quality of FBAs and BIPs written in school settings (Collins & Zirkel, 2017; Zirkel, 2011). Requiring IEP teams to consider use of FBAs, BIPs, and positive behavior supports in general, without providing guidance or clarity on that process, may have contributed to schools being poorly prepared to implement best practice assessments and interventions and the subsequent low quality of FBAs and BIPs developed there.

Similarly, the narrow interpretation in case law of the FBA to BIP process may contribute to the low quality of FBAs and BIPs in school settings. Courts typically rule in favor of school districts in cases regarding FBAs and BIPs and interpret the law regarding FBA and BIP requirements narrowly (Zirkel, 2011). The lack of any substantive requirements for FBA and BIP components or quality in federal and state law has also been incorporated into case law decisions, leading courts to rule in favor of school

districts when parents argue FBAs or BIPs were inappropriate or poor quality (Collins & Zirkel, 2017).

Research Guidance

Functional behavior assessment is a multi-component, flexible process which may vary from case to case in exactly how it is carried out. The research literature provides a wealth of information about possible ways to carry out an FBA in order to accomplish the purpose of an FBA: developing a hypothesis about the function(s) of target behavior(s). Determining a course of action for any given FBA often requires a throughout knowledge of the research literature on functional assessment of behavior as well as relevant ethics of behavior analysis. There are a multitude of choices to make in completing an FBA: deciding how many stake-holders to interview, how many observations to conduct, what dimension(s) of behavior to measure, whether to include a functional analysis, and so on. For school professionals without extensive training in ABA this process may become difficult to navigate. There are some structured and semi-structured methods of FBA data collection, some of which, like the Functional Assessment Checklist: Teachers and Staff (FACTS; McIntosh et al., 2008) and the Motivation Assessment Scale (MAS; Shogren & Rojahn, 2003), have been evaluated through research studies. Many more standard tools of FBA do not have research to provide information on psychometric properties and rigor of the data gathered through those tools.

Research on the FBA methods used by school professionals, including school psychologists, is also limited (Johnson et al., 2019). Some research on school psychologists' and other school professionals' behaviors and attitudes towards FBA are

available through larger surveys relating to general school psychologists practices (Lewis et al., 2008) and school wide behavior support (Sullivan et al., 2011), in addition to research on the attitudes and perspectives of researchers and trainers (Scott et al., 2000).

Johnson et al. (2019) conducted a survey of school psychologists' practices when conducting FBAs. Of the 199 respondents, 90% reported they were responsible for conducting FBA. Average number of FBAs respondents reported conducting in a typical school year was 8.1, averaged across respondents. Thirty-nine percent of respondents reported they do not conduct FBAs as part of a team. In regards to what components they include in a typical FBA, 98% of respondents reported including a definition of behavior, 94% included a review of records, 94% conducted interviews, 97% conducted observations, and 60% administered rating scales. When asked whether they used a standardized form to conduct interviews, 60% answered yes. 56% of respondents reported using a standardized form to conduct observations. Eighty-three percent of respondents reported including an alternative behavior in their FBA report, and 87% reported including a desired behavior in the report (Johnson et al., 2019).

Scott et al. (2000) conducted a survey of 30 individuals responsible for teaching or conducting research related to functional assessment of behavior. This survey consisted of questions related to individuals' perceptions on behaviors that warrant conducting an FBA, whether measurement methods such as ABC data collection, interviews, and record reviews are necessary or sufficient parts of an FBA. While respondents indicated that no particular data collection method was sufficient in and of itself, they did report the following methods as "sometimes" or "always" necessary parts

of an FBA: interviews, record reviews, rating scales, ABC data collection, target behavior definitions, and multiple observations (preferably three or more).

In terms of school psychologists' practices, Lewis et al. (2008) surveyed practicing school psychologists in regard to general practices of members and non-members of the National Association of School Psychologists (NASP) and included questions on frequency of conducting FBAs. The majority of school psychologists (58% of NASP members, 74% of non-members) reported conducting fewer than 10 FBAs in a typical year. NASP members reported a mean of 17 FBAs conducted annually ($SD = 22$). Non-members reported a mean of 14 FBAs in a typical year ($SD = 31$).

Sullivan et al. (2011) surveyed NASP member school psychologists on their experiences with positive behavior supports and interventions. Nearly all respondents (82%) reported that their site used FBAs. The majority of respondents reported having had graduate coursework, supervised fieldwork, and/or had attended workshops or conferences on functional behavior assessment, while only two percent of respondents reported no training in FBAs. There was a significant relationship between reported training experiences and school setting (i.e., rural, suburban, or urban) such that school psychologists working in urban schools were more likely to receive supervised experience related to FBA than those in rural schools. Additionally, there was a significant relationship between training experiences and education level; school psychologists with nondoctoral degrees were more likely to report having received in-service training and supervised experience in FBA.

Limited research provides a general picture that FBAs are being conducted by a majority of school psychologists, though the number of FBAs in a typical year varies widely. Most school psychologists report some training or coursework in FBA. More detailed information about type of training, duration, and, most importantly, effectiveness of that training in improving quality of behavior assessment and supports offered by school psychologists is limited. This lack of information on practical application of FBA in school settings further complicates efforts to understand what is necessary and sufficient for a good quality FBA which will guide an effective BIP to result in meaningful changes in student behavior.

Treatment Integrity

Quality of FBA and BIP are also important to assess and improve in order to improve treatment integrity of interventions. Treatment integrity refers to the fidelity with which an intervention is implemented as it was written and designed. Omitting steps or components of an intervention would indicate low treatment integrity. Behavior interventions, whether for individuals, groups of students, or entire schools, are significant undertakings and not always completed with integrity. Among research studies of MTSS (including academics and behavior) and school-wide primary prevention programs, fewer than half (46%) included a measure of treatment integrity (Bruhn et al., 2015). Treatment integrity outcomes also vary depending on the type of measure used and by different raters (Lane et al., 2008). Treatment integrity is important because it impacts the conclusions that can be drawn about a piece of research, including those related to effectiveness and generalizability. If school staff do not implement

interventions with fidelity, this can have an impact on the effectiveness of the

intervention as well as conclusions that can be drawn. If fidelity or integrity is not

measured in any way, either by the researcher or by the school staff tasked with

developing or implementing the intervention, valuable information is lost, and school

staff may fail to implement interventions with integrity. If interventions for behavior are

not implemented with integrity, they are less likely to be effective, and school personnel

are less likely to buy in to individualized and function-based behavior interventions as

well as potentially ABA in general in the future.

Measuring Quality

In order to ensure that FBAs and BIPs developed in school settings are of high

quality and likely to be effective, there must be a way to measure quality. This presents

some challenges, as the FBA to BIP process must, by nature, be flexible and

individualized to the student and specific behaviors to be targeted. As previously

discussed, while there are general components of FBA which are typically included, there

is no specific and standardized way to complete an FBA. The literature generally agrees

direct observation of the individual is important, but does not go so far as to specify, for

example, that three observations each of 20 minute duration across two settings using

ABC data recording is standard practice. The procedure is flexible and adaptable to the

specific context of each FBA depending on a variety of factors including the nature of the

behavior to be targeted.

One option which has been developed to aid school staff tasked with developing

behavior supports is the Technical Adequacy Tool for Evaluation (TATE; Iovannone &

Romer, 2017). The TATE is a rating tool developed by researchers based on the theoretically-essential components of FBAs and BIPs per the research literature. It was subsequently reviewed by a panel of national experts and revised again. The resulting tool includes nine items specific to FBAs and nine items specific to BIPs, each of which is rated on a scale of zero, one, or two as to the presence and completeness of that item. Items rated a zero indicate an absence of that component, while a two would indicate that component is complete and of good quality. Total scores range between zero and 18 points, with a higher score indicating a more adequate plan.

There is some prior research on the TATE, including data suggesting good inter-rater reliability of the TATE (ICC = 0.94, $p < 0.01$). In addition, convergent validity with another rating scale of BIP quality, the Behavior Support Plan Quality Evaluation (BSP-QE), was large ($d = 0.49$, $p < 0.01$; Iovannone & Romer, 2017). No research evidence is available yet on the relationship between TATE scores on an FBA and BIP and actual student outcomes in terms of reductions in problem behavior and increases in appropriate behavior. Research to support and inform the predictive validity of the TATE (i.e., do higher quality scores lead to better behavior outcomes) and classification accuracy of TATE scores (i.e., do scores above a certain cut point indicate a "quality" BIP which will lead to improvements in behavior) is needed to better inform the utility of this tool.

As previously discussed, theoretically important components of BIPs include antecedent strategy, functionally equivalent replacement behavior, a strategy to provide the same reinforcement for the replacement behavior as was maintaining the problem behavior, and a strategy to remove reinforcement for the problem behavior (O'Neill et al.,

2015) TATE item 12 measures whether there is a strategy which addresses and modifies antecedent events linked to the hypothesis of behavior. TATE item 13 measures whether a socially valid replacement behavior is identified and related to the function identified in the behavioral hypothesis. TATE item 14 measures whether there is a strategy to provide the same outcome/function to the replacement behavior as was provided to the problem behavior in order to reinforce the replacement behavior. TATE item 15 measures whether there is a strategy to eliminate the maintaining consequence of the problem behavior. These four TATE items were believed to be the most theoretically relevant to behavioral outcomes following implementation of the BIP. TATE item 16 deals with the need for a crisis plan, which, while an important consideration for safety of the individual, is not identified in the literature as an important element for behavioral change. Item 17 on the TATE targets whether there is a plan for collecting data on problem and replacement behaviors following BIP implementation. This is important in practice to evaluate the effectiveness of the intervention but is not theorized to directly impact behavior change due to BIP implementation. The final TATE item targets the presence of a plan to collect data on fidelity of implementation. Fidelity is also an important variable in terms of plan effectiveness, as a plan that is not implemented is unlikely to have the desired impact on behavior.

Another tool available to rate quality of BIPs (but not FBAs) is the Behavior Intervention Plan Quality Evaluation Tool, 2nd edition (BIP-QE II; Browning-Wright et al., 2013). This rating tool was developed based on theory of important components of BIPs developed in school settings as informed by the research literature. It includes 12

items related to BIP components which are each scored zero, one, or two points. Similar to the TATE, scores of zero indicate that component is lacking while a score of two would indicate that component is included and good quality. Possible scores range from zero to 24 total points, with higher points indicating higher quality plans. The developers provide additional cut points which categorize BIPs into either a weak plan, underdeveloped plan, good plan, or superior plan. Similar to the TATE, no research evidence yet exists which links these BIP-QE II scores to actual student behavior outcomes. No gold standard tool for evaluation FBAs and BIPs exists currently; prior research studies of FBA and BIP quality primarily use researcher-developed checklists based on theoretically important components of FBAs and BIPs.

The TATE and BIPQE-II were both developed specifically to address quality of behavioral interventions in school settings. Young children with behavioral challenges often receive services in a variety of settings beyond schools. Children on the autism spectrum also receive services through clinic- and home-based programs which present with their own advantages and disadvantages (Leaf et al., 2018). No rating tool is developed yet to specifically measure the quality of behavioral assessments and interventions delivered in clinical settings; however, it is equally important that individuals in clinical settings receive high quality interventions that are most likely to result in desired behavior changes. The current study will investigate that TATE using data derived from clinical settings and discuss implications in both clinical and school-based settings.

Current Study

The aim of the current study is to investigate the reliability and predictive validity of the TATE. Some information already exists regarding the reliability of the TATE as well as its agreement with other, similar scales which assess BIP quality. This study will aim to continue to assess this relationship, specifically between the TATE and the BIP-QE II. There is limited information on how predictive TATE scores are of actual behavior outcomes, and whether there is a relationship between TATE scores and behavior outcomes such that higher TATE scores predict decreases in problem behavior and/or increases in appropriate behavior.

Research Questions

1. Are scores derived from the TATE and the BIP-QE II significantly correlated in the predicted direction, such that as TATE scores increase so do BIP-QE II scores?

2. Are BIP quality indicators, measured using the TATE items, predictive of behavior outcomes?

3. After controlling for pre-intervention Adaptive Behavior Assessment System, 3rd Edition (ABAS3) rating scale scores, does the quality of an FBA and/or BIP predict post-intervention scores on the ABAS3?

4. After controlling for pre-intervention Pervasive Developmental Disability Behavior Inventory (PDDBI) rating scale scores, does the quality of an FBA and/or BIP predict post-intervention scores on the PDDBI?

Methods

Participants and Setting

Existing FBAs and BIPs were collected from a non-profit agency in the Midwest which offered Early Intensive Developmental and Behavioral Intervention (EIDBI) services. EIDBI services were provided by this agency at a variety of locations in the upper Midwestern United States. Locations were in both urban settings with populations of approximately 40,000 to 60,000 people to rural settings with populations ranging from 9,000 to 13,000 residents. EIDBI services are ABA-based services provided in order to educate, train, and support parents and families, promote people's independence and participation in their community, and improve long-term outcomes and quality of life for people and their families (Minnesota Department of Human Services, 2021). Services can be provided in the home, community settings, and in clinic settings. Typically, individuals enrolled in EIDBI programming receive between 10 and 40 hours of service per week.

Individuals who qualify for EIDBI services are persons under the age of 21 who have autism spectrum disorder (ASD) or a related condition, have had a comprehensive multi-disciplinary evaluation which establishes medical necessity of EIDBI services, and are enrolled in a qualifying health care program (Minnesota Department of Human Services, 2021). The FBAs and BIPs which were collected were written prior to the onset of this study and were de-identified prior to inclusion. As such, no specific information on the individuals for whom the assessments and plans were conducted is available.

A total of one FBA and 67 BIPs were collected for the study. Due to minimal data availability, the FBA was eliminated from data analysis. Two of the BIPs was eliminated from analysis due to absence of any data on that individual's behavior outcomes after the study. A total of 65 BIPs were retained for analysis in the study. BIPs were written to address a variety of problem behaviors, including off-task behavior, self-injurious behavior, aggressive behavior, and elopement. BIPs were developed by Board Certified Behavior Analysts (BCBAs) or individuals training or practicing under the direction of a BCBA. Other information about the individuals who developed the BIPs such as training or years of experience is also not available due to process of de-identifying BIPs prior to inclusion.

Data Collectors

De-identified BIPs were provided to the researcher for the study. The primary researcher was a doctoral student in school psychology. BIPs were also scored by a second rater who was a post-doctoral resident with a doctoral degree in clinical psychology. Both raters had previous exposure to ABA through graduate level coursework as well as practical experience through practicum and internship placements. Neither rater was a board certified behavior analyst (BCBA) at the time of the study.

Measures and Materials

All BIPs were scored using both the Technical Adequacy Tool for Evaluation (TATE) and the Behavior Intervention Plan Quality Evaluation Tool, 2nd edition (BIPQE-II) as a measure of BIP quality. The TATE is a rating tool developed based on theoretically essential components of FBAs and BIPs (Iovannone & Romer, 2017). It

includes nine items specific to FBAs and nine items specific to BIPs which are each given a score of zero, one, or two based on completeness and adequacy of the item. Total scores on the TATE range from 0 to 18, with high scores indicating more adequate plans. As no FBAs were available as part of this research study, the nine items related to FBA quality were eliminated from the current study. In addition, the first two items specific to the BIP reference the presence, timeliness, and relationship of the FBA for the specific BIP in question. These two items, items 10 and 11 on the TATE, were also therefore eliminated from the study. Seven items from the TATE were retained. Total scores of BIPs using the remaining TATE items ranged from zero to 14, with higher scores indicating more adequate plans. The BIPQE-II is a rating tool developed based on research and theory of important components of BIPs (Browning-Wright et al., 2013). It includes twelve items which rate presence and quality of various components of BIPs, with each item scored either zero, one, or two. Total scores range from zero to 24 points, with higher points related to higher quality plans. As none of the items on the BIPQE-II relate specifically to an FBA, all items were retained for the current study.

Interrater Reliability

Interrater reliability was calculated using the ratings of two separate raters. Each rater scored 100% of BIPs included in the study using both the TATE and BIPQE-II. Prior to rating, the primary researcher trained the second rater on the TATE and BIPQE-II using the rating scales and sample BIPs not included in the study. Three sample BIPs were rated using both the TATE and BIPQE-II by the two raters. On the third BIP both raters demonstrated 100% agreement on scores on each item on the TATE and BIPQE-II.

After training on sample BIPs to 100% agreement, both raters independently rated the BIPs included in the study. Scores were counted as agreement if both raters assigned the same value (i.e., 0, 1, or 2) on an item, and counted as disagreement if rater assigned different values. Total agreement was calculated using the number of agreements divided by the total number of items rated per rating tool. Interrater agreement on the TATE was 98%, with 7 disagreements out of 462 items rated. Interrater agreement on the BIPQE-II was 98.5%, with 12 disagreements out of 792 items rated. Items on which the two raters disagreed were discussed by both raters until an agreement was reached.

Behavior Outcomes

In order to assess for effectiveness of the BIP at improving client behavior outcomes, de-identified data collected by the non-profit EIDBI provider on behavior outcomes was also obtained. Data on behavior outcomes included measures of frequency, duration, and percent of intervals that target behavior occurred. The data provided was a measure of behavior targeted for decrease (e.g., aggression, self-injurious behavior, off-task behavior); no data on behaviors targeted for increase was provided. Three months after the BIP was implemented was selected as the timepoint to sample behavior outcome data as this is the time-frame used by the non-profit to review individual client goals and treatment plans in general and when a goal was included in a BIP it typically targeted this three-month time frame. Typically, behavior outcomes for an individual following implementation of a BIP would be reviewed using visual analysis to study immediacy of change in level, trend, and variability of behavior. However, analyses planned for the

current study require data that can quantified and compared across individuals. Visual analysis was not used to review behavior outcomes for the current study.

In addition, standardized measures of participants' adaptive skills as well as autism symptoms and problem behaviors were included. The non-profit EIDBI provider administers both the Adaptive Behavior Assessment System, 3rd edition (ABAS3; Harrison & Oakland, 2015) and Pervasive Developmental Disorder Behavior Inventory (PDDBI; Cohen & Sudhalter, 2005) rating scales to parents and caregivers at six month intervals after each client enrolls in EIDBI services in order to track progress. The ABAS3 is a standardized measure of an individual's adaptive functioning as rated by a caregiver who knows them well and indicates how dependent that individual is on caregivers in their environment to meet daily needs as compared to other same-aged peers. It was standardized using a nationally representative sample of 4,500 individuals between the ages of 0 to 89 years (Harrison & Oakland, 2015). The ABAS3 results in a number of scores; the General Adaptive Composite was used in this study. Test-retest reliability for this score was calculated with a subset of 265 individuals from the standardization sample with good reliability of the GAC for parent forms for children 0 to 5-years-old ($r = 0.83$) and for children 5 to 21-years-old ($r = 0.83$; Harrison & Oakland, 2015). Internal consistency of the ABAS parent forms for the GAC was high ($r = 0.99$; Harrison & Oakland, 2015). Interrater reliability of the ABAS parent forms was also highest for the GAC score, with estimates between 0.83 and 0.85 (Harrison & Oakland, 2015). Strong correlations were reported between the ABAS3 and ABASII GAC scores ($r = 0.88$), and between the ABAS3 GAC score and the Vineland measure of adaptive

behavior, specifically the adaptive behavior composite score, with correlation coefficients of 0.77 and 0.80 for the 0- to 5-year-old form and the 5- to 21-year-old form respectively (Harrison & Oakland, 2015).

The PDDBI is a standardized measure of autism symptoms as rated by a caregiver and compared to other same-aged peers who have a diagnosis of autism spectrum disorder (ASD). The standardization sample of the PDDBI included 369 parents of children between the ages of 1 year 6 months and 12 years 5 months with approximately 86% of the cases rated by parents diagnosed with autism and 12% of cases diagnosed with Pervasive Developmental Disorder-NOS (Cohen & Sudhalter, 2005). The Autism composite score was used from the PDDBI as well as the Aggression score, as this is the score most directly related to inappropriate behaviors targeted by participants' behavior plans, and therefore was expected to be most likely to be sensitive to changes in participant behavior. Internal consistency of the Autism composite score was reported as an alpha of 0.96, and internal consistency of the aggression scale was reported as an alpha of 0.89 (Cohen & Sudhalter, 2005). Test-retest reliability over a period of 12 months was reported to have a corrected coefficient of 0.65 for the Autism composite and 0.57 for the aggression scale, while test-retest reliability over a 2-week interval was reported to be 0.81 for the Autism composite and 0.74 for the aggression scale (Cohen & Sudhalter, 2005). Correlations between the autism composite and the aggression scale scores with the total score on the Childhood Autism Rating Scale (CARS) were reported to be 0.48 and 0.20 respectively. Interrater reliability coefficients for the autism

composite and aggression scale scores were reported to be 0.32 and 0.27 respectively (Cohen & Sudhalter, 2005).

Analyses

Prior to running analyses for specific research questions, internal consistency of the TATE was measured using coefficient alpha. Analyses were run using the statistical software R version 4.0.2 (R Core Team, 2020). The alpha coefficient for TATE BIP scores to assess internal consistency was calculated using the psych package (Revelle, 2021).

Research Question One

To evaluation the first research question whether scores on the TATE and BIP-QE II were correlated in the expected direction, a Pearson correlation was calculated. Only total scores on the TATE and BIP-QE II were included in this analysis. As no FBAs were available for the analyses, total TATE scores were based on only items rating the BIPs. Correlations between individual items were not calculated due to each rating tool having a different number of items and no clear item to item correspondence. Total scores on the TATE and BIP-QE II were expected to be positively correlated; a BIP of good quality would be expected to earn a high score on both quality rating tools while a BIP or poor quality would be expected to earn a low score on both tools. A scatterplot of TATE and BIPQE-II total scores was created using the R package ggplot2 (Wickham, 2016). The correlation between total scores was calculated using the psych package in R (Revelle, 2021). Correlation between TATE BIP total score and BIPQE-II total score was analyzed and interpreted using Cohen's 1988 guidelines.

Assumptions of the Pearson correlation are that variables are continuous, related pairs, without outliers, and data are linear (Field & Miles, 2010). Scores on both the TATE and the BIP-QE II are continuous numbers. Scores were in related pairs as each BIP was rated using both rating tools, resulting in two quality scores for each BIP. Upon completing a scatterplot of TATE and BIP-QE II total scores, no outlying data points were observed and there was observed to be a linear relationship (Figure 1).

Research Question Two

The second research question asks whether there are BIP components, as measured by the TATE items, which are predictive of behavioral outcomes. Individual BIP items were used in this analysis rather than the total BIP score on the TATE because of the low observed internal consistency of the TATE scores. Additionally, it was theorized that certain items on the TATE could be more important to behavior outcomes, such as items related to the replacement behavior or antecedent or consequence strategies, than other items such as whether a data collection plan was included for progress monitoring. TATE items 12, 13, 14, and 15 were selected for this analysis as they were theorized to be most likely to impact behavioral outcomes following plan implementation.

A Poisson regression was used to analyze TATE item scores on behavior outcomes. For this analysis, only the 50 BIPs which included progress monitoring data based on frequency of behavior were used. These outcomes were all measured as a count of how frequently the behavior had occurred, which fits the assumption for Poisson regression that the response variable is a count per unit of time or space and all counts

41

were positive integers. Further, when graphed in a histogram, the behavior outcome data follow a positively-skewed distribution which also fits the assumption of a Poisson distribution (see Figure 2). Mean and variance of progress monitoring data on frequency of behavior were measured to determine if the assumption that means and variances are equal was met. Overall, variance was slightly higher than the mean of frequency of behavior following BIP implementation ($m = 2.85$, $var = 7.91$). When the data were grouped by frequency of behavior at baseline, the mean and variance of progress monitoring data were roughly equal for observations where baseline frequency was less than or equal to five ($m = 1.50$, $var = 1.85$) and somewhat unequal for observations where baseline frequency was greater than five ($m = 5.47$, $var = 10.27$). Only modest violations of the means equal to variances assumption were observed, and the Poisson regression was used (Roback & Legler, 2021). Poisson regression analyses were conducted in R using the stats package (R Core Team, 2020) as well as the sandwich package (Zeileis, 2006; Zeileis et al., 2020) to calculate robust standard errors.

Research Question Three

To evaluate research question three regarding the prediction of post-intervention ABAS3 scores using scores on TATE items, a multiple regression analysis was planned. Multiple regression assumes a linear relationship between the outcome variable and independent variables. This assumption was assessed using scatterplots of post-intervention ABAS3 scores with TATE scores and was met. The assumption of no multicollinearity in the data was checked using variance inflation factor (VIF) through the car package in R (Fox & Weisburg, 2019). None of the VIF values were high enough

to indicate a problem with multicollinearity. Normal distribution of the residuals of the regression were checked using a QQ plot with the ggpubr program in R (Kassambara, 2020) which met the assumption. After assumptions of multiple regression were checked (Field & Miles, 2010), the multiple regression was run using the stats package in R (R Core Team, 2020) while controlling for pre-intervention ABAS3 scores.

Research Question Four

To evaluate research question four regarding the prediction of post-intervention PDDBI scores using scores on TATE items, a multiple regression analysis was planned. Multiple regression assumes a linear relationship between the outcome variable and independent variables. This assumption was assessed using scatterplots of post-intervention PDDBI scores with TATE scores and was met. The assumption of no multicollinearity in the data was checked using variance inflation factor (VIF) through the car package in R (Fox & Weisburg, 2019). None of the VIF values were high enough to indicate a problem with multicollinearity. Normal distribution of the residuals of the regression were checked using a QQ plot with the ggpubr program in R (Kassambara, 2020) which met the assumption. After assumptions of multiple regression were checked (Field & Miles, 2010), the multiple regression was run using the stats package in R (R Core Team, 2020) while controlling for pre-intervention PDDBI scores.

Results

A total of 65 BIPs were included for analysis. To be included the BIP had to be accompanied by data on behavior which was related to the BIP (e.g., BIP lists aggression as a target behavior and data on frequency of aggression is available). Of these, 44

included parent/caregiver completed ABAS3 forms both before and after the BIP was written. 56 included parent/caregiver completed PDDBI forms with a score for the Autism composite, while 40 included a score for the Aggression scale. The reason for this difference is likely that some caregivers were administered the standard PDDBI which does not include scores for the Aggression scale, while other caregivers were likely administered the extended parent form of the PDDBI which does include the aggression scale.

Of the 65 BIPs with data on behavior outcomes, 1.5% measured behavior by percent of opportunities ($n = 1$), 1.5% measured behavior by percentage of intervals when behavior occurred ($n = 1$), 20% measured duration of target behavior in minutes ($n = 13$), and 77% measured behavior by frequency ($n = 50$). All behavior outcomes measured inappropriate behaviors which were targeted for decrease, such as aggression, off-task behavior, self-injurious behavior, eloping, and non-contextual vocalizations. Baseline rates of behavior were considered as the first five data points included in the data provided. These five data points were averaged to provide one number indicative of baseline behavior. For most BIPs, the five baseline data points were collected within a one to two week period, though there was up to a three week span during baseline data collection. The five data points closest to three months from the date on the BIP were used to calculate progress, again through averaging the five data points to provide one number that indicated progress. Similarly, post-intervention data points ranged from a one to three week span.

While not related to a specific research question, internal consistency of BIP related items on the TATE was also calculated using coefficient alpha, resulting in a standardized alpha of 0.44. Item 18 on the TATE was eliminated due to demonstrating zero variance. Internal consistency of the BIPQE-II was also calculated and resulted in a standardized alpha of 0.59.

Notably relative to these low alpha coefficients, both the items on the TATE and the items on the BIPQE-II, while all theoretically related to overall BIP quality, are each measuring a separate component of the BIP being evaluated. A BIP which has any one component would not necessarily be expected to have any other particular components. For example, a BIP with an antecedent strategy to reduce problem behavior might or might not also have a plan to teach a replacement behavior or a plan to collect data on treatment integrity of the plan implementation. It is not surprising, then, that internal consistency of both instruments was low as the items on the TATE and on the BIPQE-II are measuring distinct components of BIPs.

Research Question One: Correlation of TATE and BIPQE-II

All 65 BIPs included in the study were scored using both the TATE and the BIPQE-II. TATE items related to FBAs (i.e., items 1 through 11) were not scored as no FBAs were available for inclusion. The seven remaining items on the TATE were scored, leaving total possible scores ranging between 0 and 14 (as possible scores for each item ranged from 0 to 2). All items on the BIPQE-II were scored and total possible scores ranged from 0 to 24. Descriptive statistics on the TATE and BIPQE-II are presented in Table 1.

TATE and BIPQE-II scores were expected to be positively correlated. This was based on the assumption that both rating forms are a measure of BIP quality and would therefore be related. Correlation was tested by calculating the Pearson product-moment correlation coefficient using R. There was a positive correlation between scores on the TATE and on the BIPQE-II, $r = 0.56$, $p < 0.001$, indicating that as BIP scores on the TATE increased, so do scores on the BIPQE-II. By Cohen's (1988) guidelines, this indicates a strong correlation because r is greater than .50. There was a strong correlation in the expected direction between BIP scores on the TATE and BIPQE-II. A scatterplot of total scores on the TATE and BIPQE-II is presented in Figure 1.

Research Question Two: TATE Item Scores Predictive of Behavior Outcomes

To evaluate whether scores on the TATE were predictive of behavior outcomes, individual TATE items were evaluated as predictors of frequency of problem behavior (i.e., behavior targeted for decrease in the BIP) post-BIP implementation. TATE items 12, 13, 14, and 15 were each selected as predictor variables in the model as these items were theorized to be most related to behavior outcomes. Item 12 is related to presence of an antecedent strategy in the BIP. Item 13 rates whether a functionally equivalent replacement behavior is determined. Item 14 rates whether there is a strategy to reinforce the replacement behavior to meet the same function as was hypothesized to maintain the problem behavior. Item 15 rates presence of a strategy to eliminate the maintaining consequence for the problem behavior. These four components were hypothesized to be the most impactful for behavior outcomes.

The outcome variable for the model was frequency of problem behavior. BIPs which included behavior outcome data measured in another way, such as duration of behavior or percentage of opportunities, were not included in this analysis. Only BIPs with behavior outcome data measured as a frequency or count per unit of time were included as these data met the criteria for a Poisson regression. The frequency of problem behavior was taken from progress monitoring graphs which indicated how many times a discrete behavior had occurred per day or session of treatment. Baseline data on behavior was included as a control variable in this model. Baseline data was also a measure of frequency of behavior, how many times per day or session the discrete target behavior had occurred. The model analyzed with Poisson regression included the TATE items 12 through 15 as predictor variables, baseline frequency of behavior as a control, and frequency of behavior post-intervention as the outcome or dependent variable.

Results of this model were that TATE items 12, 13, and 15 were not significant predictors of frequency at the $p = .01$ level of behavior following BIP implementation, with p-values of 0.03, 0.23, and 0.57 respectively. TATE item 14 was a significant predictor at the .01 level with a coefficient of 0.62 ($p = 0.001$). The expected log count for a one-unit increase in score on TATE item 14 is .62. Additionally, and as expected, baseline data on frequency of behavior was also a significant predictor of frequency of behavior following BIP implementation, with a coefficient of 0.03 ($p < .001$).

The model was run again with only the significant predictor TATE item 14 and the control variable of baseline behavior frequency included. Robust standard errors were also calculated to control for mild violation of the distribution assumption of means equal

to variances, along with corresponding parameter estimates and p-values. This information is available in Table 2. The fit of this model was assessed using a goodness-of-fit chi-squared test with the residual deviance. The chi-squared test was significant ($p = .001$), suggesting that the data do not fit the model well. As the assumptions were met, there may be a predictor variable missing from the current study.

Research Question Three: TATE Items Predict ABAS3 Scores

A multiple regression analysis using the ordinary least squares (OLS) method was used to evaluate whether scores on TATE items were predictive of adaptive behavior as measured by ABAS3 scores following BIP implementation. TATE items 12, 13, 14, and 15 were again selected as predictor variables due to theoretical importance to behavior change. Standard scores on the ABAS3 general adaptive composite (GAC) completed by parents after the BIP was implemented were used as the continuous outcome variable. Scores on the ABAS3 GAC completed by parents prior to implementation of the BIP were included as a control variable. There were a total of 44 BIPs included in this analysis; 15 BIPs were not included as there was either no pre-intervention ABAS3, no post-intervention ABAS3, or no ABAS3 data whatsoever.

Results of the multiple regression indicated none of the TATE items were significant predictors of ABAS3 GAC scores after BIP implementation when ABAS3 GAC scores prior to BIP implementation were accounted for. Pre-intervention ABAS3 GAC scores were a significant predictor of post-intervention ABAS3 GAC scores ($p < .001$). The p values for TATE items 12, 13, 14, and 15 were as follows: $p = 0.27$, $p = 0.54$, $p = 0.88$, and $p = 0.19$ respectively. None of the TATE items included for analysis

were significant predictors of adaptive behavior as measured by the ABAS3 following implementation of the BIP when controlling for ABAS3 scores prior to BIP implementation.

Research Question Four: TATE Items Predict PDDBI Scores

A multiple regression analysis using OLS was used to evaluate whether scores on TATE items were predictive of autism symptoms as measured by PDDBI scores following BIP implementation. TATE items 12, 13, 14, and 15 were again selected as predictor variables due to theoretical importance to behavior change. Standard scores on the PDDBI autism composite completed by parents after the BIP was implemented were used as the continuous outcome variable. Scores on the PDDBI autism composite completed by parents prior to implementation of the BIP were included as a control variable. As with the ABAS3, there were a total of 44 BIPs included in this analysis; 15 BIPs were not included as there was either no pre-intervention PDDBI, no post-intervention PDDBI, or neither. These were not the same 44 BIPs as were included in the third research question; several BIPs included PDDBI scores but did not have ABAS3 scores and vice versa.

Results of the multiple regression indicated none of the TATE items were significant predictors of PDDBI autism composite scores after BIP implementation when PDDBI autism composite scores prior to BIP implementation were accounted for. Pre-intervention PDDBI autism composite scores were a significant predictor of post-intervention PDDBI autism composite scores ($p < .001$). The p values for TATE items 12, 13, 14, and 15 were as follows: $p = 0.69$, $p = 0.15$, $p = 0.38$, and $p = 0.65$

respectively. None of the TATE items included for analysis were significant predictors of autism symptoms as measured by the PDDBI following implementation of the BIP when controlling for PDDBI scores prior to BIP implementation.

Another multiple regression analysis was conducted to evaluate whether any TATE item scores were predictive of the aggression scale score on the PDDBI. The aggression scale score on the PDDBI was also used due to its theoretical relation to development and implementation of a BIP as this scale is based on parent perception of their child's disruptive and aggressive behaviors which are typical targets for BIPs. TATE items 12, 13, 14, and 15 were again selected as predictor variables due to theoretical importance to behavior change. Standard scores on the PDDBI aggression scale completed by parents after the BIP was implemented were used as the continuous outcome variable. Scores on the PDDBI aggression scale completed by parents prior to implementation of the BIP were included as a control variable. A total of 30 BIPs were included in this analysis; 29 were not included due to absence of pre-intervention, post-intervention, or both PDDBI aggression scale scores.

Results of the multiple regression indicated none of the TATE items were significant predictors of PDDBI aggression scale scores after BIP implementation when PDDBI aggression scale scores prior to BIP implementation were accounted for. Pre-intervention PDDBI aggression scale scores were a significant predictor of post-intervention PDDBI aggression scale scores ($p < .001$). The p values for TATE items 12, 13, 14, and 15 were as follows: $p = 0.32$, $p = 0.23$, $p = 0.18$, and $p = 0.29$ respectively. None of the TATE items included for analysis were significant predictors of aggressive

and defiant behaviors as measured by the PDDBI following implementation of the BIP when controlling for PDDBI scores prior to BIP implementation.

Discussion

The purpose of this study was to evaluate the validity of the TATE tool in assessing the quality of BIPs. This was assessed by gathering pre-existing BIPs from a clinical EIDBI provider where ABA services were provided to children with autism or related conditions. In addition to BIPs, pre-existing data collected by the provider on client behavior outcomes as well as parent ratings on the ABAS3 and PDDBI were obtained. BIPs were not developed and implemented as part of this study, only evaluated after-the-fact.

Internal consistency of the TATE was evaluated and resulted in a standardized alpha of 0.44, while internal consistency of the BIP-QE II resulted in a standardized alpha of 0.59. The low internal consistency of both tools may be related to the limited sample size of the current study. It may also be due to the different components of a BIP measured by each item of each tool. A BIP with one component would not necessarily be expected to also include other components; for example, a BIP that includes a plan to withhold the consequence that previously maintained a problem behavior would not necessarily be expected to also have a plan to collect progress monitoring data. While all of the items measured by the scales are theoretically important to a high quality BIP, there isn't necessarily a theoretical relationship between the items that would lead to the expectation of high internal consistency. Additionally, certain items or components of the

BIP may be more or less impactful than others in terms of behavior change following BIP implementation.

The TATE and the BIPQE-II are two theoretically related tools, both used to evaluate the quality or adequacy of BIPs. It was predicted that BIP scores on the TATE and on the BIPQE-II would positively correlate because a high-quality BIP would in theory score high on both instruments, while a low-quality BIP would have a low score on both instruments. Scores on BIPs obtained for this study did correlate in the expected direction, with a strong correlation of 0.56. This is in line with a previous estimate of the correlation between TATE and Behavior Support Plan Quality Evaluation (BSP-QE), an earlier version of the BIPQE-II, which yielded a large correlation ($d = 0.49$, $p < 0.01$; Iovannone & Romer, 2017). The strong correlation between scores on the TATE and scores on other instruments designed to assess BIP quality is evidence supporting the convergent validity of the TATE tool in assessing quality of behavior intervention plans.

Information to inform predictive validity of the TATE was also assessed by analyzing whether TATE item scores were predictive of changes to behaviors targeted by the BIP as well as more distal outcomes including adaptive skills and behavior impairment as rated by parents. TATE items 12, 13, 14, and 15 were selected for this analysis. TATE item 12 rates whether there is an antecedent strategy which directly address and modifies antecedents identified in the FBA and described with enough detail for implementation. TATE item 13 rates whether there is a minimum of one socially valid replacement behavior that will be taught to the student and described in enough detail for implementation. TATE item 14 rates if there is a minimum of one strategy that

will reinforce the replacement behavior and provide the same outcome or function as the problem behavior did, described with enough detail to implement. TATE item 15 rates if there is a minimum of one strategy which eliminates the maintaining consequence of the problem behavior which is described with sufficient detail to implement (Iovannone & Romer, 2017).

TATE items 12, 13, and 15 were not significant predictors of frequency of problem behavior following implementation of the BIP, while controlling for baseline frequency of problem behavior. TATE item 14 was a significant predictor such that high scores on item 14 were significantly related to reductions in frequency of problem behavior. Ensuring that a behavior plan includes an explicit and clearly described strategy to provide functionally equivalent reinforcement following occurrence of an appropriate replacement behavior may be particularly important to ensuring improvements in behavior outcomes following plan implementation. This makes theoretical sense as a core principle of behavioral intervention is providing an alternate, appropriate way for an individual to meet whatever need or function they achieved previously with an inappropriate behavior.

However, the Poisson regression model still was not a good fit for the data, and important predictor variables may be missing from the model. One crucial element of understanding plan effectiveness is data on whether the plan was implemented to fidelity. No information on fidelity of implementation was available as part of this study; it would be a worthwhile endeavor for future research to evaluate the relationship between BIP quality and eventual behavior outcomes.

TATE scores were not predictive of parental ratings of adaptive skills or behavior impairment as measured by the ABAS3 and PDDBI. The ABAS3 and PDDBI are both measures of more global behavior. They may therefore be less sensitive to the changes in specific behavior outcomes directly aligned with individual BIPs. There are several other possible reasons for the results observed in the current study, as well as limitations to the current dataset and subsequent analyses. The current study utilized only pre-existing BIPs and behavior data; there was no measure of treatment integrity or indication of whether the BIP was implemented at all. While quality of BIP is important to treatment outcomes, there are certainly other variables which would impact outcomes including whether the plan was implemented as written, and to what degree. Plan implementation could have varied due to many factors, including familiarity and training of staff, buy-in of staff and other service providers, and extent to which the client participated in services. The pre-existing data that was collected spanned a period of time prior to and during the COVID-19 pandemic, which may have impacted or interrupted participation in EIDBI services and therefore interrupted BIP implementation during services.

Setting of service delivery is also an important consideration. The TATE tool was developed to support school staff with development of quality, function-based behavior interventions. In school settings, staff may have had limited exposure to ABA in general, including the concept of behavior as predictable and purposeful, as well as function-based behavior intervention specifically. While surveys of school psychologists show that a majority of these professionals have some training or exposure to ABA (Sullivan et al., 2011), there is less information on the ABA training experiences that teachers,

paraprofessionals, administrators, and other school staff who may be tasked with implementation of BIPs have received.

In the EIDBI setting where the BIPs for this study were developed, all staff have some degree of training and supervision in ABA. Staff at the providing agency were either Board Certified Behavior Analysts (BCBAs), Board Certified Assistant Behavior Analysts (BCaBAs), or Registered Behavior Technicians (RBTs) who are under the supervision of a BCBA, or individuals who were in training to achieve one of these certifications. Staff with training in ABA could be reasonably assumed to have some knowledge of the functions of behavior (i.e., attention, tangible, escape/avoidance, automatic) and of common methods of behavior change (e.g., reinforcement, extinction, functional communication training). It is possible that in the cases observed in this study where a poorly rated BIP and decreases in inappropriate behaviors were observed, staff were implementing interventions to address hypothesized function of inappropriate behaviors without these interventions having been written down in a plan specific to a target behavior. In this case as well, information on treatment integrity and what strategies are implemented would be important to understanding the relationship between BIP quality and behavior outcomes.

Potentially, due to differences between service delivery settings, it may be beneficial to develop a tool for evaluating behavioral assessment and intervention plan quality specific to clinical settings. The same components of assessment and intervention theorized to be integral to behavior change as evidenced by the research literature would be important considerations of such a tool. The standards required to achieve a higher

score might be different for FBAs and BIPs developed in a clinical setting than those in a school due to the difference in training and resources available in a clinical setting. For clinical settings where a BCBA develops the FBA and BIP and oversees progress, and where children receive a high intensity of ABA-based services, more rigorous standards for quality might be appropriate to ensure maximum benefit for children receiving such high-intensity services.

Research evaluating the link between TATE scores on FBAs and BIPs developed in the school setting is important to be able to judge the utility of this tool to guide school-based programming. In settings where staff are not tasked primarily with implementing ABA interventions and do not have significant training and/or supervision in ABA, there may be more of an apparent relationship between quality of BIPs as measured by the TATE and actual student behavior outcomes. Should this be the case, the TATE would be a useful tool for measuring and improving quality of function-based behavior supports for students. School staff could integrate the TATE into existing school-wide behavior supports or MTSS as a tool to aid in improving the quality of intervention offered to students in need of intensive, individualized supports. It can also be integrated into consultation as a tool for consultant and consultee to evaluate the quality and likelihood of success of individualized interventions developed as part of the consultative relationship.

Another significant limitation of this study was the absence of any functional behavior assessments included in the data analysis. Since only one de-identified FBA was provided to the researcher along with the de-identified BIPs, it could not be assumed that

BIPs were developed by the agency staff after completing some form of assessment (e.g., an FBA). An informal or brief assessment of behavioral function may have been routine, or even more complete FBAs which were simply not inputted into the agency's medical records system where it could be retrieved and de-identified in order to be included in the current study. Quality and completeness of these FBAs, if they were completed, would have likely had an impact on the robustness of the behavioral hypothesis and the subsequent function-based interventions as well as effected the likelihood that behaviors would change in the desired direction. It is crucial for function-based behavior interventions to have accurately identified the function of the target behavior and provided a replacement behavior which meets the same function (O'Neill et al., 2015). The quality of the FBA and resulting behavioral hypothesis are crucial elements of the function-based behavior support process, and the data missing from the current study in terms of whether there was assessment into behavioral function and the quality of that assessment is a significant limitation to the conclusions which can be drawn about the utility of the TATE. More than half of the TATE items were not included in this study due to the absence of FBAs in the de-identified data provided to the researcher. Further research is needed to assess whether TATE scores are a significant predictor of individual behavior outcomes following implementation of intervention when the TATE is used in its entirety to rate the combination of FBA and BIP.

Another application of the TATE which was not evaluated in the present study due to the limitations of de-identified, pre-existing data, is its use as a tool to guide development of function-based behavior supports. The TATE can be used to evaluate

FBAs and BIPs that have already been written as one tool in identifying why a particular intervention may not have been successful in decreasing inappropriate behaviors and/or increasing appropriate behaviors. It can also be used to evaluate a plan that has been written prior to implementation, to guide decisions about whether a plan needs further adjustment before it is put into practice. The TATE might also be utilized as a guiding document to assist the developer of a function-based behavior intervention as they go through the process of assessing for behavioral functioning and writing a plan. While not a checklist of to-do items for a good FBA and BIP, the TATE could serve as a reminder of theoretically important components of the FBA to BIP process and guide the provider in planning how to conduct an assessment and what to include in an intervention.

In this way, the TATE could be a useful tool during the consultative relationship. Prior research has shown that behavior interventions are most likely to have a good contextual fit and have high likelihood of implementation with fidelity if they are developed by a team of professionals with expert knowledge of behavior as well as those with knowledge of the setting where the plan will be implemented (Benazzi et al., 2006). This lends itself well to a behavior consultation model where a consultant with expert knowledge of behavior works with the consultee who has knowledge of the classroom setting where the plan will be implemented and who will likely be responsible for plan implementation. In a consultation relationship, the consultant could use the TATE to help guide their information gathering and development of the behavior plan while integrating information from the consultee to ensure good contextual fit and acceptability of the plan. This might allow school psychologists with some training in ABA but who are not

BCBAs to ensure the technical soundness of the plans they develop while in consultation with teachers or other school professionals. This may help with improving the quality of FBAs and BIPs developed in school settings, which are notoriously of poor quality (Blood & Neel, 2007; Cook et al., 2007; Van Acker et al., 2005). Development of technically sound, acceptable plans which are then implemented with fidelity is crucial to improving the educational outcomes of students who engage in significant behaviors that disrupt their own learning and the learning of others. Future research into the utility of the TATE in evaluating FBAs and BIPs as well as in guiding the process of development of function-based behavior supports continues to be a necessary endeavor to improving the quality of services provided in school settings.

Conclusion

This study aimed to evaluate the validity of the TATE as a tool for assessing the quality of BIPs. Pre-existing, de-identified BIPs were scored using the TATE and resulting TATE scores were analyzed for their correlation with BIPQE-II scores as well as predictive power for specific behavior outcomes as well as more global outcomes on the ABAS3 and PDDBI rating scales. TATE scores were strongly, positively correlated with scores on the BIPQE-II ($r = 0.56$, $p < 0.001$). Higher scores on TATE item 14, related to providing reinforcement for a replacement behavior which met the same function of the problem behavior, were predictive of a reduction in frequency of problem behavior following BIP implementation. No TATE item scores were predictive of more global outcomes measured by the ABAS3 and PDDBI parent rating scales.

References

A.B. 2586, 1990 California Legislature (Cal. 1990).

Adams, K., & Dunsmuir, S. (2009). Functional behaviour assessment: Addressing severe, complex and enduring needs. *Educational and Child Psychology, 26*(4), 82-101.

Association of Professional Behavior Analysts (2015). 2014 U.S. professional employment survey: A preliminary report. http://www.csun.edu/~bcba/2014-APBA-Employment-Survey-Prelim-Rept.pdf. Accessed Feb 2022.

Benazzi, L., Horner, R. H., & Good, R. H. (2006). Effects of behavior support team composition on the technical adequacy and contextual fit of behavior support plans. *The Journal of Special Education, 40*(3), 160-170.

Blood, E. & Neel, R. S. (2007). From FBA to implementation: A look at what is actually being delivered. *Education and Treatment of Children, 30*(4), 67-80.

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes. *Journal of Positive Behavior Interventions, 12*(3), 133-148.

Briere, D. E., & Simonsen, B. (2011). Self-monitoring interventions for at-risk middle school students: The importance of considering function. *Behavioral Disorders, 36*(2), 129-140.

Browning-Wright, D., Mayer, G. R., & Saren, D. (2013). Behavior intervention plan quality evaluation scoring guide II. *Positive Environments, Network of Trainers (PENT)*. Retrieved from: www.pent.ca.gov.

Bruhn, A. L., Hirsch, S. E., & Lloyd, J. W. (2015). Treatment integrity in school-wide programs: A review of the literature (1993-2012). *Journal of Primary Prevention, 36,* 335-349.

Bruni, T. P., Drevon, D., Hixson, M., Wyse, R., Corcoran, S., & Fursa, S. (2017). The effect of functional behavior assessment on school-based interventions: A meta-analysis of single-case research. *Psychology in the Schools, 54*(4), 351-369.

Calloway, C. J., & Simpson, R. L. (1998). Decisions regarding functions of behavior: Scientific versus informal analyses. *Focus on Autism and Other Developmental Disorders, 13*(3), 167-175.

Camp, E. M., Iwata, B. A., Hammond, J. L., & Bloom, S. E. (2009). Antecedent versus consequent events as predictors of problem behavior. *Journal of Applied Behavior Analysis, 42*(2), 469-483.

Campbell, A., & Anderson, C. M. (2008). Enhancing effects of check-in/check-out with function-based support. *Behavioral Disorders, 33*(4), 233-245.

Carter, D. R., & Horner, R. H. (2009). Adding function-based behavioral support to first step to success: Integrating individualized and manualized practices. *Journal of Positive Behavior Interventions, 11*(1), 22-34.

Chu, E. M., & Ready, D. D. (2018). Exclusion and urban public high schools: Short- and long-term consequences of school suspensions. *American Journal of Education, 124*(4), 479-509.

Clunies-Ross, P., Little, E., & Kienhuis, M. (2008). Self-reported and actual use of proactive and reactive classroom management strategies and their relationship with teacher stress and student behaviour. *Educational Psychology, 28*(6), 693-710.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates, Publishers.

Cohen, I. L., & Sudhalter, V. (2005). *PDD behavior inventory: Professional manual*. Psychological Assessment Resources.

Collins, L. W., & Zirkel, P. A. (2017). Functional behavior assessments and behavior intervention plans: Legal requirements and professional recommendations. *Journal of Positive Behavior Interventions, 19*(3), 180-190.

Cook, C. R., Crews, S. D., Browning-Wright, D., Mayer, G. R., Gale, B., Kraemer, B., & Gresham, F. M. (2007). Establishing and evaluating the substantive adequacy of positive behavioral support plans. *Journal of Behavioral Education, 16,* 191-206.

Cook, C. R., Mayer, G. R., Browning-Wright, D., Kraemer, B., Wallace, M. D., Dart, E., Collins, T., & Restori, A. (2012). Exploring the link among behavior intervention plans, treatment integrity, and student outcomes under natural educational conditions. *The Journal of Special Education, 46*(1), 3-16.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Pearson.

Drasgow, E., & Yell, M. L. (2001). Functional behavioral assessments: Legal requirements and challenges. *School Psychology Review, 30*(2), 239-251.

Ellingson, S. A., Miltenberger, R. G., Stricker, J., Galensky, T. L., & Garlinghouse, M. (2000). Functional assessment and intervention for challenging behaviors in the classroom by general classroom teachers. *Journal of Positive Behavior Interventions, 2*(2), 85-97.

Evers, W. J. G., Tomic, W., & Brouwers, A. (2004). Burnout among teachers: Students' and teachers' perceptions compared. *School Psychology International, 25*(2), 131-148.

Farkas, M. S., Simonsen, B., Migdole, S., Donovan, M. E., Clemens, K., & Cicchese, V. (2012). Schoolwide positive behavior support in an alternative school setting: An evaluation of fidelity, outcomes, and social validity of tier 1 implementation. *Journal of Emotional and Behavioral Disorders, 20*(4), 275-288.

Fee, A., Schieber, E., Noble, N., & Valdovinos, M. G. (2016). Agreement between questions about behavior function, the motivation assessment scale, functional assessment interview, and brief functional analysis of children's challenging behaviors. *Behavior Analysis: Research and Practice, 16*(2), 94-102.

Field, A., & Miles, J. (2010). *Discovering statistics using SAS.* SAGE Publications LTD.

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

Flower, A., McKenna, J. W., & Haring, C. D. (2017). Behavior and classroom management: Are teacher preparation programs really preparing our teachers? *Preventing School Failure: Alternative Education for Children and Youth, 61*(2), 163-169.

Fox, J., & Weisburg, S. (2019). *An {R} companion to applied regression, Third Edition*. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

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

Gold, Y. (1985). Burnout: Causes and solutions. *The Clearing House, 58*(5), 210-212.

Government Accountability Office (2018). Report to congressional requesters: Discipline disparities for black students, boys, and students with disabilities. *K-12 Education*, retrieved from: https://www.gao.gov/products/GAO-18-258.

Harrison, P. L., & Oakland, T. (2015). *ABAS-3: Adaptive behavior assessment system* (3rd edition). Western Psychological Services.

Hawken, L. S., & Horner, R. H. (2003). Evaluation of a targeted intervention within a schoolwide system of behavior support. *Journal of Behavioral Education, 12*(3), 225-240.

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions, 11*(3), 133-144.

Hurl, K., Wightman, J., Haynes, S. N., & Virues-Ortega, J. (2016). Does a pre-intervention functional assessment increase intervention effectiveness? A meta-analysis of within-subject interrupted time-series studies. *Clinical Psychology Review, 47*, 71-84.

Individuals with Disabilities Education Act, 20 U.S.C. §1400-1415 (2004).

Iovannone, R. & Romer, N. (2017, February). *The FBA/BIP technical adequacy tool for evaluation (TATE): Improving practice*. Paper session presented at the meeting of the National Association of School Psychologists, Texas.

Ishikawa, S., Okajima, I., Matsuiko, H., Sakano, Y. (2007). Cognitive behavioral therapy for anxiety disorders in children and adolescents: A meta-analysis. *Child and Adolescent Mental Health, 12*(4), 164-172.

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*(2), 197-209. (Reprinted from *Analysis and Intervention in Developmental Disabilities, 2*(1), 3-20, 1982).

Johnson, A. H., Goldberg, T. S., Hinant, R. L., & Couch, L. K. (2019). Trends and practices in functional behavior assessments completed by school psychologists. *Psychology in the Schools, 56*(3), 360-377.

Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr

Lambert, A. M., Tingstrom, D. H., Sterling, H. E., Dufrene, B. A., & Lynne, S. (2015). Effects of tootling on classwide disruptive and appropriate behavior of upper-elementary students. *Behavior Modification, 39,* 413-430.

Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity: An essential – but often forgotten – component of school-based interventions. *Preventing School Failure: Alternative Education for Children and Youth, 48*(3), 36-43.

Lane, K. L., Kalberg, J. R., Bruhn, A. L., Mahoney, M. E., & Driscoll, S. A. (2008). Primary prevention programs at the elementary level: Issues of treatment integrity, systematic screening, and reinforcement. *Education and Treatment of Children, 31*(4), 465-494.

Leaf, J. B., Leaf, R., McEachin, J., Cihon, J. H., & Ferguson, J. L. (2018). Advantages and challenges of a home- and clinic-based model of behavioral intervention for individuals diagnosed with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 48,* 2258-2266.

Lewis, T. J., & Sugai, G. (1999). Effective behavior support: A systems approach to proactive schoolwide management. *Focus on Exceptional Children, 31*(6), 1-23.

Lewis, M. F., Truscott, S. D., & Volker, M. A. (2008). Demographics and professional practices of school psychologists: A comparison of NASP members and non-NASP school psychologists by telephone survey. *Psychology in the Schools, 45*(6), 467-482.

Losen, D., Hodson, C., Keith, M. A. II, Morrison, K., & Belway, S. (2015). Are we closing the school discipline gap? Los Angeles, CA: UCLA.

Martens, B. K., & Lambert, T. L. (2014). Conducting functional behavior assessments for students with emotional/behavioral disorders. In H. M. Walker & F. M. Gresham (Eds.), *Handbook of evidence-based practices for emotional and behavioral disorders: Applications in schools* (pp. 243-257). New York, NY: Guilford Press.

McIntosh, K., Borgmeier, C., Anderson, C. M., Horner, R. H., Rodriguez, B. J., & Tobin, T. J. (2008). Technical adequacy of the functional assessment checklist: Teachers and staff (FACTS) FBA interview measure. *Journal of Positive Behavior Interventions, 10*(1), 33-45.

McIntosh, K., Campbell, A. L., Carter, D. R., & Dickey, C. R. (2009). Differential effects of a tier two behavior intervention based on function of problem behavior. *Journal of Positive Behavior Interventions, 11*(2), 82-93.

Minnesota Department of Human Services (2021, July 22). Early intensive developmental and behavioral intervention (EIDBI) benefit. MN Department of Human Services.

https://www.dhs.state.mn.us/main/idcplg?IdcService=GET_DYNAMIC_CONVE
RSION&RevisionSelectionMethod=LatestReleased&dDocName=dhs16_195658

Moore, T. C., Wehby, J. H., Oliver, R. M., Chow, J. C., Gordon, J. R., & Mahany, L. A.
(2017). Teachers' reported knowledge and implementation of research-based
classroom and behavior management strategies. *Remedial and Special Education,
38*(4), 222-232.

Myers, C. L., & Holland, K. L. (2000). Classroom behavioral interventions: Do teachers
consider the function of the behavior? *Psychology in the Schools, 37*(3), 271-280.

O'Neill, R. E., Albin, R. W., Storey, K., Horner, R. H., & Sprague, J. R. (2015).
*Functional assessment and program development for problem behavior: A
practical handbook* (3rd ed.). Stamford, CT: Cengage Learning.

Oliver, A. C., Pratt, L. A., & Normand, M. P. (2015). A survey of functional behavior
assessment methods used by behavior analysts in practice. *Journal of Applied
Behavior Analysis, 48*(4), 817-829.

R Core Team (2020). R: A language and environment for statistical computing. R
Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-
project.org/.

Revelle, W. (2021) psych: Procedures for Personality and Psychological Research,
Northwestern University, Evanston, Illinois, USA, https://CRAN.R-
project.org/package=psych Version = 2.1.9.

Roback, P., & Legler, J. (2021). *Beyond multiple linear regression: Applied generalized
linear models and multilevel models in R*. CRC Press, Taylor & Francis Group.

Roscoe, E. M., Phillips, K. M., Kelly, M. A., Farber, R., & Dube, W. V. (2015). A
statewide survey assessing practitioners use and perceived utility of functional
assessment. *Journal of Applied Behavior Analysis, 48*(4), 830-844.

Scott, T. M., Alter, P. J., Rosenberg, M., & Borgmeier, C. (2010). Decision-making in
secondary and tertiary interventions of school-wide systems of positive behavior
support. *Education and Treatment of Children, 33*(4), 513-535.

Scott, T. M., Meers, D. T., & Nelson, C. M. (2000). Toward a consensus of functional
behavioral assessment for students with mild disabilities in public school
contexts: A national survey. *Education & Treatment of Children, 23*(3), 265–285.

Scott, T. M., Park, K. L., Swain-Bradway, J., & Landers, E. (2007). Positive behavior
support in the classroom: Facilitating behaviorally inclusive learning

environments. *International Journal of Behavioral Consultation and Therapy, 3*(2), 223-235.

Shogren, K. A., & Rojahn, J. (2003). Convergent reliability and validity of the questions about behavior function and the motivation assessment scale: A replication study. *Journal of Developmental and Physical Disabilities, 15*, 367–375.

Skinner, C. H., Cashwell, T. H., & Skinner, A. L. (2000). Increasing tootling: The effects of a peer-monitored group contingency program on students' reports of peers' prosocial behaviors. *Psychology in the Schools, 37,* 263-270.

Steege, M. W., & Scheib, M. A. (2014). Best practices in conducting functional behavior assessments. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative decision making* (pp. 273-286). Bethesda, MD: National Association of School Psychologists.

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., Scott, T., Liaupsin, C., Sailor, W., Turnbull, A. P., Turnbull, H. R. III, Wickham, D., Wilcox, B., & Ruef, M. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions, 2*(3), 131-143.

Sullivan, A. L., Long, L., & Kucera, M. (2011). A survey of school psychologists' preparation, participation, and perceptions related to positive behavior interventions and supports. *Psychology in the Schools, 48*(10), 971-985.

Van Acker, R., Boreson, L, Gable, R. A., & Potterton, T. (2005). Are we on the right course? Lessons learned about current FBA/BIP practices in schools. *Journal of Behavioral Education, 14*(1), 35-56.

Vannest, K. J., Davis, J. L., Davis, C. R., Mason, B. A., & Burke, M. D. (2010). Effective intervention for behavior with a daily behavior report card: A meta-analysis. *School Psychology Review, 39*(4), 654-672.

Weisz, J. R., McCarty, C. A., & Valeri, S. M. (2006). Effects of psychotherapy for depression in children and adolescents: A meta-analysis. *Psychological Bulletin, 132*(1), 132-149.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wilson, C. P., Gutkin, T. B., Hagen, K. M., & Oats, R. G. (1998). General education teachers' knowledge and self-reported use of classroom interventions for working

with difficult-to-teach students: Implications for consultation, prereferral intervention, and inclusive services. *School Psychology Quarterly, 13*(1), 45-62.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*(2), 203-214.

Wood, B. K., Umbreit, J., Liaupsin, C. J., & Gresham, F. M. (2007). A treatment integrity analysis of function-based intervention. *Education and Treatment of Children, 30*(4), 105-120.

Wright, D. B., & Gurman, H. B. (2001). *Positive interventions for serious behavior problems: Best practices in implementing the Hughes Bill (A.B. 2586) and the positive behavioral intervention regulations* (Rev. ed.). Sacramento, CA: California Department of Education.

Zeileis, A. (2006). "Object-Oriented Computation of Sandwich Estimators." _Journal of Statistical Software_, *16*(9), 1-16. doi: 10.18637/jss.v016.i09 (URL: https://doi.org/10.18637/jss.v016.i09).

Zeileis, A., Köll, S., & Graham, N. (2020). "Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R." _Journal of Statistical Software_, *95*(1), 1-36. doi: 10.18637/jss.v095.i01 (URL: https://doi.org/10.18637/jss.v095.i01).

Zirkel, P. A. (2011). State special education laws for functional behavioral assessment and behavior intervention plans. *Behavior Disorders, 36*(4), 262-278.

Table 1

Descriptive Statistics for TATE and BIPQE-II Total Scores

Variable	N	*M*	*SD*	Minimum	Maximum
TBIPTotal	65	4.77	2.07	0	10
BIPQETotal	65	7.63	2.80	3	14

Note. TBIPTotal represents total score on the TATE items scored for BIPs. BIPQETotal represents total score on the BIPQE-II for BIPs.

Table 2

Parameter Estimates with Robust Standard Errors for Poisson Regression

Variable	Estimate	Robust SE	*p value*	Lower Limit	Upper Limit
Goal1BL	0.03	0.005	< .001	0.02	0.04
TBIP14	0.89	0.196	< .001	0.51	1.27

Note. Goal1BL represents frequency of target behavior prior to BIP implementation. TBIP14 represents score on TATE item 14.
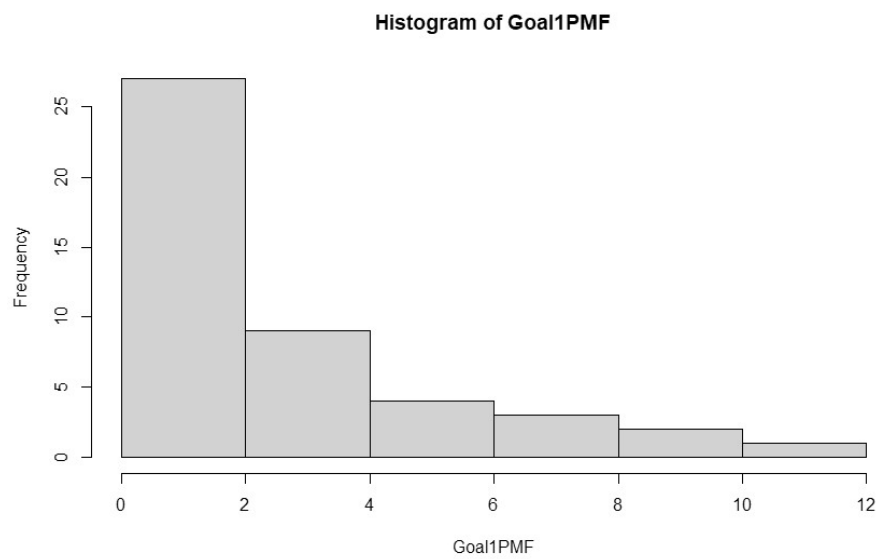
Figure 1

Scatterplot of BIPQE-II Scores by TATE Scores

Figure 2

Histogram of Frequency of Behavior Post-intervention

Note. Goal1PMF represents frequency of target behavior following BIP implementation