

A quantitative synthesis of developmental disability research: The impact of functional assessment methodology on treatment effectiveness

Caitlin H. Delfs & Jonathan M. Campbell

Abstract

Methods and outcomes from functional behavioral assessment have been researched widely over the past twenty-five years. However, several important research questions have yet to be examined sufficiently. This quantitative review of developmental disability research aims to make comparisons of different functional behavioral assessment methodologies, both across and within diagnostic categories. Quantitative synthesis data were used to answer questions regarding behavioral function, assessment type, differences based upon diagnostic category, and treatment effectiveness. Results indicate that assessment methodology does not impact treatment effectiveness, but both identified functions and treatment effectiveness are impacted by diagnosis. Implications for clinicians as well as future research directions are also discussed.

Keywords: developmental disability, autism, intellectual disability, functional assessment, meta-analysis

Although not essential for a diagnosis, Autism Spectrum Disorders (ASD) and Intellectual Disability (ID) are commonly associated with a broad range of maladaptive behaviors including self-injurious behavior (SIB), property destruction, aggression towards others, severe disruptions, and stereotypic behaviors (e.g., body rocking). Maladaptive behaviors can lead to poor social relationships; poor academic success, destruction of property, and serious medical problems, such as tissue damage. For these reasons, the assessment and treatment of such behaviors in individuals with ASD and ID is an important component of any comprehensive approach to rehabilitation.

A behavioral approach to intervening with maladaptive behaviors has been consistently documented as the most efficacious approach for treating aberrant behaviors (Gresham et al., 2004; Campbell, Herzinger, & James, 2007). The key to effective treatment is the identification of the function, or purpose, of the behavior. The most current taxonomy of behavioral function focuses on three types of reinforcement as the major mechanisms maintaining behavior: (a) positive reinforcement, (b) negative reinforcement, and (c) automatic reinforcement. In the last 25 years, there has been a trend toward developing treatments for maladaptive behaviors following determination of the hypothesized functions of the behaviors through Functional Behavior Assessments (FBA). Based on the ascribed function of the target behavior, an appropriate treatment package can be selected. Researchers assessing maladaptive behaviors agree that identifying the function of the target behavior is integral in the treatment selection process; thus FBAs are a core feature in the development of interventions designed to ameliorate aberrant behaviors (Yarborough & Carr, 2000) and required by federal education law (e.g., Individuals with Disabilities Education Act [IDEA], P.L. 105-117, 1997).

Although required by law in some cases, the term FBA is still somewhat vague. Generally, FBA refers to any methodology used to identify the purpose of behavior and encompasses *indirect assessments*, (e.g., interviews, rating scales), *descriptive assessments* (e.g., A-B-C sheets, direct observation with no variable or environment manipulation); and *functional analyses* (FA; e.g., analogue conditions in which antecedent or consequent variables are systematically manipulated within an experimental design). For the purposes of this paper, we are using the term FA to describe all

experimental analyses. The term Behavioral Assessment (BA) refers to those assessments which are non-experimental in nature and includes both indirect and descriptive assessments.

Several researchers have made comparisons across FBA methodologies and, in general, the findings support the FA as the “gold standard” for ascribing function and consequently developing function-based treatments. Paclawskyj et al. (2001) and Durand and Crimmins (1988) both reported positive correlations when comparing FA outcomes to the functions hypothesized by the Questions About Behavioral Function (QABF; Matson & Vollmer, 1995) and the Motivation Assessment Scale (MAS; Durand & Crimmins, 1992), respectively. In contrast, Hall (2005) found that descriptive and experimental methods of FBA agreed only 25% of the time. In almost all published accounts of comparison data, the FA represented the gold standard for validity tests of other types of assessment.

Others have looked beyond comparisons of ascribed function across FBA types and instead assessed intervention outcomes across methodologies. Knowing which FBA methodology is associated with more successful treatment outcomes is imperative. Didden, Korzilius, van Oorsouw, and Sturmey (2006) made comparisons across descriptive and experimental FBAs and found that treatments based on experimental methods resulted in significantly higher treatment effectiveness scores. Herzinger and Campbell (2007) conducted a meta-analysis of autism literature on the assessment and treatment of maladaptive behaviors. The authors found that when using a particular effect size calculation to determine treatment efficacy, treatments that were based on FA results were more effective than those based on BA results.

Any of the three aforementioned reinforcement types (i.e., positive, negative, and automatic) may be the maintaining variable for maladaptive behaviors. However, it is possible that participant characteristics, such as diagnostic category, may also influence the function of maladaptive behaviors. In other words, disability type may align with behavioral function. It may be hypothesized that individuals diagnosed with ID, but without any impairment in socialization or communication, are likely to engage in behaviors that result in access to social positive reinforcement more frequently than those diagnosed with ASD, a disability characterized by marked impairments in socialization and communication. Similarly, it may be hypothesized that individuals with ASD are more likely to engage in behaviors ascribed to automatic functions due to the sensory sensitivities commonly reported in the literature or escape functions to avoid situations in which interactions with others is necessary. In their review of the developmental disability literature, Dawson, Matson, and Cherry (1998) focused on individuals who functioned in the severe to profound range of ID based on the notion that level of cognitive functioning may influence the reasons for the maladaptive behavior (i.e., function). Although Dawson et al. (1998) found no significant differences in ascribed function across diagnostic categories, they did find a pattern of mean differences to support their hypothesis that diagnosis mediates function of maladaptive behaviors. Currently, there is little published research comparing the differences in ascribed functions of maladaptive behaviors across diagnostic categories.

The current study aims to answer the following specific research questions: 1) Is treatment more effective when following an experimental functional analysis (FA) or a non-experimental behavioral assessment (BA) for individuals with developmental disabilities?, 2) Is there a predominant observed function based on the type of assessment (i.e., FA or BA) for either (a) individuals with ASD, (b) individuals with ID, or (c) individuals with ASD and ID?, 3) Does ascribed function differ depending on diagnostic category?, 4) Does the observed function of the behavior, regardless of FA method used, have an impact on the effectiveness of treatment?, and 5) Is treatment effectiveness impacted by diagnosis? For example, do individuals with ASD who function in the range of ID show poorer response to behavioral treatment than those without a diagnosis of ASD?

METHOD

Study Identification and Selection

For the years 2000 through 2005, published functional assessments of problem behavior for individuals with developmental disabilities were identified through searches of PsycLit, ERIC, and MedLine databases using appropriate search terms, such as subject descriptions (e.g., autism, mental retardation, intellectual disability), target behaviors (e.g., self-injurious behaviors, aggression, problem behaviors), and assessment type (e.g., applied behavior analysis, functional assessment, functional analysis). Published studies were identified by issue-by-issue hand searches of the following journals: *American Journal on Intellectual and Developmental Disabilities*, previously known as *American Journal of Mental Retardation*, *Behavioral Interventions*, *Behavior Modification*, *Education and Training in Developmental Disabilities*, previously known as *Education and Training in Mental Retardation and Developmental Disabilities*, *Journal of Applied Behavior Analysis*, *Journal of Association of People with Severe Handicaps*, *Journal of Autism and Developmental Disorders*, *Journal of Intellectual Disability Research*, *Intellectual and Developmental Disabilities* previously known as *Mental Retardation*, and *Research in Developmental Disabilities*. Also, timely references (i.e., citations between 2000 and 2005) from each article found through the literature search were reviewed for possible inclusion.

Studies were selected for inclusion if the following criteria were satisfied. First, studies were selected if they were published in peer reviewed journals between January 2000 and December 2005, following the introduction of FBA as a requirement of the Individuals with Disabilities Act (IDEA, 1997). Second, only single-subject designs were included and then only if a participant was diagnosed with ASD or ID. If the participants were described as being “autistic-like,” or developmentally delayed, they were also included. Third, a FBA had to be conducted and results reported, with maladaptive behaviors as the target behaviors of treatment. If an article did not report treatment data in the following format, it was not included in the treatment effectiveness analyses: (a) data points, not just mean scores, were reported; (b) baseline data and intervention data were reported; and (c) if the intervention procedures targeted reduction of stereotyped, self-stimulatory, self-injurious, destructive, disruptive, or aggressive behaviors. If an article included multiple participants or studies only partially met inclusionary criteria, only those components that met criteria were included in the review. There were no exclusion criteria for age or gender of participants or assessment/intervention setting. Articles that met some criteria (e.g., ASD diagnosis, targeted reduction of problem behavior) but did not meet others (e.g., no functional assessment reported) were not included (e.g., Pace and Toyer, 2000; Scattone et al., 2002).

Estimating Effects of Behavioral Interventions

Effect size calculations

There are several methods for assessing effectiveness data using both regression and non-regression approaches. Frequently reported summary methods have involved the calculation of Mean Baseline Reduction (MBLR), Percentage of Non-overlapping Data (PND), and Percentage of Zero Data (PZD; Campbell, 2003). Other methods, such as the Percentage of data points Exceeding the Median (PEM; Ma, 2006) and Improvement Rate Difference (IRD; Parker & Hagan-Burke, 2007) could have been selected for use and comparison. Olive and Smith (2005) found merit in both the MBLR and PND for calculating non-regression effect sizes for single subject designs. Based on their common usage in related literature reviews, the following three effect sizes based on nonregression approaches were calculated per intervention in the current study: MBLR, PND, and PZD.

The MBLR is calculated by subtracting the mean of treatment observations from the mean of baseline observations then dividing by the mean of baseline observations and multiplying by 100 (Campbell, 2003; Lundervold & Bourland, 1988; O'Brien & Repp, 1990). The PND statistic is calculated

as the percentage of treatment data that did not overlap with baseline data points (Scruggs, Mastropieri, & Casto, 1987). If a baseline phase reported one or more data points of zero, then the same number of data points was excluded in the treatment phase prior to calculation of the PND (Didden, Duker & Korzilius, 1997). The PND can range from 0 to 100%. According to Scruggs, Mastropieri, Cook, and Escobar (1986) a PND greater than 90% reflects a highly effective treatment, a PND of 70-90% is considered a fair treatment outcome, and a PND of less than 50% indicates unreliable/ineffective intervention. The PZD statistic is calculated by locating the first intervention data point that reached zero and computing the percentage of data points that reached zero including the first zero point (Scotti et al., 1991). The PZD score is considered a more stringent efficacy indicator as it requires target behaviors to reach and stay at zero levels throughout treatment to be considered effective. Campbell (2004) noted that the PZD score represents a "degree of behavior suppression versus degree of behavior reduction" (p. 235). PND and PZD scores have been found to be independent indicators of treatment outcome (Scotti et al., 1991; Campbell, 2003) and have been used in several studies to measure the effectiveness of treatments (Didden, Korzilius, van Oorsouw, & Sturmey, 2006; Herzinger & Campbell, 2007).

Handling multiple outcomes, participants, assessment types, and experimental phases

Several rules have been established for the coding of assessment type. Using Herzinger and Campbell's (2007) coding system, functional assessment type was coded as either: (a) FA (strictly adhering to guidelines set forth in Iwata et al., 1982), (b) modified FA, (c) brief FA, (d) partial experimental, (e) A-B-C sheet, (f) rating scales (e.g., MAS, QABF), (g) informal assessment, or (h) other. Under the modified FA code, FAs that included sessions not described by Iwata et al. (e.g., tangible) or sessions that differed in length of time were coded. Brief FAs, such as those described by Northup and colleagues (1991) and summarized by Derby and colleagues (1992) were coded. The partial experimental category was used for FAs in which antecedent variables were manipulated, but there were no programmed consequences for target behavior, such as structured descriptive assessments described by Anderson and Long (2002). Later, these groups (i.e., FA, modified FA, brief FA, and partial experimental) were consolidated in order to unify all the experimental analyses. The studies coded as A-B-C sheets, rating scales, and informal assessments were consolidated to form the BA, or non-experimental category. Thus, articles were coded as FA if any environmental variables (antecedents, consequences, or both) were altered as opposed to those designated as BA which did not include variable manipulation.

Consistent with the methodology of Herzinger and Campbell (2007), if two different types of FBAs (e.g., FA and BA) were used with a participant, the methods were coded separately with the possibility of two functions and different treatments identified. If a participant's problem behavior was assessed using multiple BA methods (e.g., MAS, parent interview, and observation) as is often done in both clinical and educational settings, the assessments were coded as a combination. In such a case, the coding resulted in one effect size unless the BAs yielded different functions or different treatments for each method.

Studies that reported on multiple outcomes or multiple participants required separate effect size calculations for each outcome for each participant. When more than one problem behavior was targeted for a participant and separate data points were reported, individual effect sizes were calculated per problem behavior per participant rather than arbitrary selection of one behavior. This approach was used in order to capture all available data regarding each participant and each problem behavior.

Single case designs vary (e.g., A-B-A-B) and effect sizes can be calculated from varied contrasts (Allison & Gorman, 1993). In the present study, the effect sizes were calculated between the first non-treatment phase and the last treatment phase, per Faith et al.'s (1996) recommendations and implemented in Campbell (2003) and Herzinger and Campbell (2007). In designs that compared multiple treatments (e.g., A-B-A-C), the initial baseline and final treatment phase were coded. Although it is not ideal to make

comparisons between baseline and subsequent intervention phases that are separated by both time and experience, this was necessary given the limitations of the meta-analysis format used in the current study. In studies using a multi-element or alternating treatments design, both treatments were coded unless a final “best treatment alone” condition was conducted. In this case, the initial baseline phase and final treatment condition were coded.

Data extraction and variables coded

For the necessary analyses in the present study, the graphs provided by the articles were transformed into raw data via a ruler. The distance between each point and the abscissa was calculated in millimeters and rounded to the nearest 0.5. The data conversion procedure has been used by Allison, Faith, and Franklin (1995), Campbell (2003), and Herzinger and Campbell (2007) with a high degree of inter-rater reliability.

The following participant information was coded when available: participant’s age, gender, race, level of intellectual functioning, secondary diagnoses, years since diagnosis prior to study, and years of prior treatment. The following assessment/pre-intervention data were coded: target behavior, type of FBA, ascribed function(s) of behavior, type of intervention used, length of session, treatment setting, and type of therapist. Targeted behaviors were coded as: aggression, property destruction, disruptive behaviors (e.g., spitting), vocalizations, SIB, and stereotyped behaviors. If relevant, specific types of SIB were also recorded.

The following intervention data were coded: type of intervention, type of experimental design, inter-rater reliability, number of baseline data points, number of final phase treatment points, and attempt to generalize. The types of intervention coded included non-contingent reinforcement, differential reinforcement, punishment, timeout, extinction, sensory extinction, FCT, combined treatments, and other interventions. Based on Herzinger and Campbell (2007), these categories were later consolidated into six categories: (a) reinforcement only, (b) punishment only, (c) extinction only, (d) reinforcement and punishment, (e) extinction plus reinforcement or punishment, and (f) other.

Reliability of data extraction and coding decisions

Eighteen articles were randomly selected for independent coding by advanced graduate students in Psychology, who had experience working with individuals with autism and ID, and inter-rater agreement was established. The 18 articles (21.69% of all articles) included 30 separate assessments (15.07% of all assessments) and 26 different participants (18.05% of all participants). Inter-rater agreement, with a mean of 99.71% and a range of 95.98% to 100% across all coded variables, was determined by the percent agreement method ($\# \text{ of agreements} / \# \text{ of agreements} + \# \text{ of disagreements} \times 100$).

Inferential statistical procedures

Different statistical procedures were used to answer each research question. Three one-way ANOVAs were used to examine research question 1 (a comparison of treatment effectiveness for FA and BA for individuals in one of three diagnostic categories). A non-parametric Chi-square test of non-independence was used to assess research question 2 (possible bias in assessment outcomes based on FBA methodology). A non-parametric Chi-square test of non-independence was used to examine research question 3 (assessment of impact of diagnostic category on FBA outcome). Three one-way ANOVAs were used to address research question 4 (treatment effectiveness as impacted by function). Treatment effectiveness means for each effect size statistic were compared for each of seven functional categories. Three one-way ANOVAs were used to examine research question 5 (treatment effectiveness as impacted by diagnostic category). Treatment effectiveness means for each effect size statistic were compared across three diagnostic categories.

RESULTS

Table 1 provides information on the characteristics of participants and studies. This review included 83 articles reporting on 144 participants with a total of 199 separate studies (i.e., assessments and/or treatments). The 199 studies were collected from a total of eight journals, with the *Journal of Applied Behavior Analysis* contributing the highest percentage of articles (48.7%). Based on the previously determined mutually exclusive categories, the majority of participants fell into the ID only category (54.9%), followed by autism and ID (24.3%), followed by autism only (12.5%), and finally unspecified developmental disability (8.3%). Those reported as having unspecified developmental delays were later combined with the ID only category based on the lack of autism characteristics mentioned, following further analysis of the participant descriptions reported in primary articles. The ratio of males to females in this study was 1.5 : 1. For individuals described as having autism or autistic characteristics, the gender ratio was 3.5 : 1 in favor of males, which is similar to prior reviews documenting the higher prevalence of autism in males. Also consistent with prior reviews documenting the prevalence of ID in individuals with autism, the majority of the participants diagnosed with autism functioned in the range of mental retardation (79.9%) or were considered “untestable” via standardized, formal intelligence testing.

Table 1. Participant and Study Characteristics

Characteristic	n	%
Gender		
Male	87	60.4
Female	57	39.6
Main Diagnostic Category		
ID	79	54.9
Autism/ID	35	24.3
Autism	18	12.5
Dev. Disability	12	8.3
Level of Intellectual Disability (IQ range)		
Severe (≤ 39)	79	54.9
Not reported	28	19.4
Moderate (54-40)	26	18.1
Mild (70 - 55)	9	6.3
Untestable/Other	2	1.4
Language Ability		
Nonverbal/Mute	52	36.1
Not reported	48	33.3
Some functional language	39	27.1
Average language	4	2.8
Echolalic	1	.7
Journal		
Journal of Applied Behavior Analysis	97	48.7
Behavioral Interventions	39	19.6
Research in Developmental Disabilities	27	13.6
American Journal on Mental Retardation	18	9.0
Journal of Autism and Developmental Disorders	9	4.5
Behavior Modification	7	3.5
Other	2	1.0
Total N	199	
Number of participants per article		
1	48	57.8
2	15	18.1
3	14	16.9
4	4	4.8
5	2	2.4
Total N	83	

Table 2 provides detailed information about FBA types, behavioral interventions, and experimental quality. Studies employed both experimental (77.4%) and non-experimental (22.6%) methods of FBA. Under the FA umbrella, the majority of assessments were modified FAs (53.9%), which

are based on the analogue conditions of Iwata et al., but tailored in terms of the specific conditions used in the analysis. The type of BA most often reported was described as informal assessment (73.3%). The A-B-A-B experimental design (i.e., reversal and withdrawal) was the most commonly used, reported in 31.2% of the studies, followed by the multiple baseline design (29.6%). Studies in the meta-analysis omitted follow-up data collection (82.4%) more often than they included follow-up data collection (14.1%), $\chi^2(1, N=199) 219.03, p < .001$. Generalization data were omitted from the studies (71.4%) more often than reported (28.6%), $\chi^2(1, N=199) 89.89, p < .001$. Inter-rater reliability for FBA sessions was reported in 88.9% of articles with a median of 98.0 (range, 80.0% to 100.0%). For treatment sessions, inter-rater reliability data were reported in 100% of articles with a median of 98.0 (range, 70.2% to 100.0%).

Table 2. Assessment, Intervention, and Experimental Characteristics

Characteristic	n	%	
Type of functional behavioral assessment			
Experimental	154	77.3	
Modified session EFA	83		
EFA (Iwata et al.)	49		
Brief EFA	16		
Partial Experimental	4		
Other	2		
Non-experimental	45	22.6	
Informal Assessment	33		
Combination of BA types	4		
Descriptive Assessment	4		
ABC sheet	3		
Not reported	1		
Type of intervention			
Reinforcement only	106	53.3	
Extinction and Reinforcement or Punishment	53	26.6	
Other/Not reported	20	10.1	
Reinforcement and Punishment	10	5.0	
Extinction only	6	3.0	
Punishment only	4	2.0	
Experimental design			
Reversal/Withdrawal	62	31.2	
Multiple Baseline	59	29.6	
Multiple Treatment Comparison	27	13.6	
Alternating Treatments	24	12.1	
Combination	15	7.5	
Simple A-B	11	5.5	
Other	1	0.5	
Follow-up data collected			
No	164	82.4	
Yes	28	14.1	
Not Reported	7	3.5	
Attempt to generalize behavior			
No/Not reported	142	71.4	
Yes	57	28.6	
Characteristic	r	M	SD
Reliability of observations			
Inter-rater reliability			
FBA	80.0-100.0	75.6	36.8

Table 3 depicts that relationship between FA methodology type and treatment effectiveness. Three independent samples *t*-tests indicated that when comparing FA and BA across diagnostic categories, there were no significant differences in treatment effectiveness as measured by the three effect

sizes calculated. Results from the independent samples t-tests and reported means, standard deviations, and ranges of effect size calculations are presented in Table 3.

Table 3. *Descriptive Statistics for Three Effect Sizes by FBA Methodology*

	<u>M</u>	<u>SD</u>	<u>Min</u>	<u>Max</u>
FA				
MBLR	81.29	25.93	-49.03	100.00
PND	81.06	30.53	0.00	100.00
PZD	58.77	35.36	0.00	100.00
<u>BA</u>				
MBLR	75.86	29.65	-14.14	100.00
PND	77.09	32.67	0.00	100.00
PZD	51.12	35.66	0.00	100.00

Results from One-way ANOVAs

MBLR: $F(1, 197) = 1.43, n.s$

PND: $F(1, 197) = .57, n.s$

PZD: $F(1, 197) = 1.62, n.s$

Note. FA = functional analysis; BA = behavioral assessment; MBLR = mean baseline reduction; PND = percentage of non-overlapping data; PZD = percentage of zero data; M = mean; SD = standard deviation; Min = minimum value; Max = maximum value. Descriptive statistics are presented for 199 treatment outcomes.

Table 4 shows the results from the comparison of ascribed function related to functional assessment methodology. The data showed that there is a relationship between the type of methodology (e.g., FA, BA) used in the assessment and the result of that assessment, $\chi^2(5, N = 199) 19.81, p < .01$. This finding indicates that there are significant differences in the function ascribed to target behaviors depending on the type of functional assessment used. The results indicated that FA procedures more likely result in a social positive reinforcement function (i.e., behavior maintained by access to tangible items or social attention) and BA are more likely to result in automatic functions. In addition, BA most often indicated a single function maintaining target behaviors as opposed to target behaviors that are multiply maintained.

Table 4. *Frequencies for Assessment Type and Function*

Assessment Type	<u>Function</u>					
	<u>POS RF</u>	<u>NEG RF</u>	<u>AUT</u>	<u>COM</u>	<u>UND</u>	<u>OTH</u>
FA	49	36	31	27	9	2
BA	12	12	16	1	0	4
Total	61	48	47	28	9	6

Note. FA = functional analysis; BA = behavioral assessment; POS RF = social positive reinforcement; NEG RF = social negative reinforcement; AUT = automatic; COM = combination; UND = undifferentiated; OTH = other

Table 5 shows the comparison of ascribed function as impacted by diagnostic category. A non-parametric chi-square test of non-independence showed that there are significant differences in the ascribed functions of target behaviors across diagnostic categories, $\chi^2(10, N = 199) 29.22, p < .01$. The data indicated that individuals diagnosed with ASD alone and ASD and ID were most often identified as having target behaviors maintained by social negative functions (e.g., escape from tasks). Individuals diagnosed as functioning within the range of ID were more often identified as having maladaptive behaviors maintained by social positive contingencies.

Table 5. *Frequencies for Diagnostic Category and Function*

Diagnostic Category	<u>Function</u>					
	<u>POS RF</u>	<u>NEG RF</u>	<u>AUT</u>	<u>COM</u>	<u>UND</u>	<u>OTH</u>
ASD	8	14	4	1	3	0
ID	36	15	26	16	1	6
ASD/ID	17	19	17	11	5	0
Total	61	48	47	28	9	6

Note. ASD = autism diagnosis; ID = intellectual disability diagnosis; ASD/ID = autism and ID diagnoses; POS RF = social positive reinforcement; NEG RF = social negative reinforcement; AUT = automatic; COM = combination; UND = undifferentiated; OTH = other

Table 6 depicts data from treatment effectiveness as impacted by ascribed function. The results from three one-way ANOVAs indicated no differences for ascribed function for the three calculated effect sizes. Treatment effectiveness, as assessed by the (a) MBLR statistic, $F(5, 193) = .46, n.s.$, (b) PND statistic, $F(5, 193) = 1.10, n.s.$, and (c) PZD statistic, $F(5, 193) = 1.72, n.s.$ was not significantly affected by the function of the problem behavior. Table 6 includes data regarding the means, and standard deviations for the three calculated effect sizes.

Table 6. Means and Standard Deviations for Three Effect Sizes by Ascribed Function

Function	<u>MBLR</u>	<u>PND</u>	<u>PZD</u>
POS RF	82.29 (24.29)	83.81 (24.91)	61.57 (29.20)
<u>NEG RF</u>	77.74 (27.54)	75.46 (32.30)	54.16 (38.85)
<u>AUT</u>	79.22 (30.35)	82.60 (34.37)	50.10 (39.24)
<u>COM</u>	77.87 (28.24)	71.43 (36.72)	58.61 (34.57)
<u>UND</u>	81.62 (26.11)	82.82 (29.10)	61.37 (36.13)
<u>OTH</u>	95.02 (5.79)	98.07 (3.24)	89.55 (15.18)

Note. POS RF = social positive reinforcement; NEG RF = social negative reinforcement; AUT = automatic; COM = combination; UND = undifferentiated; OTH = other; MBLR = mean baseline reduction; PND = percentage of non-overlapping data; PZD = percentage of zero data

Table 7 shows the results of treatment effectiveness across diagnostic category. The results from three one-way ANOVAs indicated no differences for ascribed function for two of the three calculated effect sizes. Treatment effectiveness, as assessed by the (a) MBLR statistic, $F(2, 196) = .45, n.s.$, and (b) PND statistic, $F(2, 196) = .51, n.s.$ was not significantly affected by the diagnosis of the individual. However, treatment effectiveness as assessed by the PZD statistic did indicate significant differences of treatment effectiveness across diagnostic categories, $F(2, 196) = 4.36, p < .01$. When assessing treatment effectiveness with the PZD statistic, the most stringent efficacy indicator, treatment is significantly more effective for individuals with ID than individuals diagnosed with ASD. Table 7 includes data regarding the means and standard deviations for the three calculated effect sizes.

Table 7. Means and Standard Deviations for Three Effect Sizes by Diagnostic Category

Diagnostic Category	<u>MBLR</u>	<u>PND</u>	<u>PZD</u>
ASD	80.84 (13.47)	75.81 (29.39)	45.84 (32.04)
<u>ID</u>	81.27 (28.62)	81.89 (30.30)	63.23 (34.52)
<u>ASD/ID</u>	77.13 (28.57)	78.95 (33.44)	50.32 (37.07)

Note. ASD = autism diagnosis; ID = intellectual disability diagnosis; ASD/ID = autism and ID diagnoses; MBLR = mean baseline reduction; PND = percentage of non-overlapping data; PZD = percentage of zero data

DISCUSSION

Summary of Findings

The current study focused on a series of questions regarding the impact of functional assessment methodology, functional assessment outcome, and diagnostic category on treatment effectiveness. When comparing treatment outcomes for interventions based on both experimental and non-experimental FBA for all three diagnostic groups, no significant differences were found. This finding adds to the mixed results currently found in the literature regarding comparisons of FBA methodologies. Findings also indicate that FBA methodology itself moderates the outcomes of the assessment. For example, FA were more likely to result in maladaptive behaviors being identified as maintained by social positive reinforcement contingencies as opposed to BA which most often identified automatic functions. There were also differences in the ability of FBA methodologies to detect multiply-maintained behaviors. One potential hypothesis to explain these findings is the possibility of rater bias inherent in many forms of BA. The BA is often dependent on the ability of the rater to report accurate information and, like other rating scales, is subject to rater biases. In addition, the goal of many BA is to identify the most significant function of the maladaptive behavior. Potentially useful information may be lost when the assessment methodology assumes maladaptive behaviors are maintained by only one function and are not multiply-maintained. Significant differences in FBA functional outcomes were found depending on the type of assessment conducted. These findings differ from previous research that indicates that FA and BA methodologies themselves do not impact the results of the assessment (e.g., Herzinger & Campbell, 2007).

Diagnostic category was identified as significantly affecting ascribed function, regardless of FBA type used. For individuals with ASD, maladaptive behavior was more likely to be identified as being maintained by social negative reinforcement (i.e., escape). These findings support the hypothesis that marked impairment in socialization and communication, as evinced by a diagnosis of ASD, may influence the function of maladaptive behavior. It is possible that individuals with ASD may avoid situations due to communication and/or socialization deficits. The results indicate that the functions of maladaptive behavior, in and of themselves, have no significant impact on treatment effectiveness. That is, the ascribed function of behavior does not mediate the effectiveness of interventions. Although previous research (e.g., Vollmer, 1994; Piazza, Hanley, & Fisher, 1996), has indicated that behaviors maintained by some functions are more difficult to treat, the results of the current study and others (Herzinger & Campbell, 2007) do not support that notion. A significant relationship was observed when assessing the impact of diagnostic category on treatment effectiveness. Interventions were more successful for individuals with ID, but not diagnosed with ASD, than for any other category when assessed using the PZD statistic. Follow-up comparisons were made across diagnostic categories to help explain this finding. Due to the overrepresentation of males in the two ASD categories (i.e., ASD only and ASD/ID), three one-way ANOVAs were conducted. Treatment effectiveness was not significantly affected by the gender of the individual as measured by all three effect sizes. This indicated that the higher PZD scores for individuals with ID (and without reported characteristics of ASD) was not due to unequal gender ratios across diagnostic categories. Comparisons of communication ability were also made across three diagnostic categories to determine if language skills were the moderating variable impacting treatment effectiveness. The results of a one-way ANOVA indicated no significant differences in average level of communication across the diagnostic categories. These follow-up comparisons across diagnostic categories could not explain the differences in treatment outcome as measured by the PZD statistic.

Implications for clinicians

The results of the quantitative review are relevant to practicing clinicians regarding the assessment and treatment of severe maladaptive behaviors exhibited by individuals with developmental disabilities. The most salient finding, with immediate implications for practitioners, is the effect of diagnostic category on identified function of maladaptive behaviors and treatment outcome effectiveness.

Although these preliminary results should be interpreted with caution and a thorough analysis that includes rigorous experimental control and replication is warranted, there may be some immediate implications. If an individual's diagnosis can be indicative of the maintaining contingencies of exhibited maladaptive behaviors, knowledge of that diagnosis could inform treatment development from the beginning. Similarly, the impact of diagnostic category on treatment effectiveness, when assessed by the PZD statistic, could inform decisions regarding treatment outcome expectations at the onset of evaluation and treatment development. Currently, some behavioral clinicians and researchers do not use diagnostic categories to describe the participants in their studies. This may be a nod to the behavioral perspective that focuses on observable behavior and making treatment decisions based on that rather than assumed characteristics implied by diagnostic associations. However, the results of the current study suggest that diagnostic category may be linked to the functional relationships for maladaptive behaviors. Procedural concerns regarding the use of FBA methodologies that are prone to particular results (i.e., ascribing function to behavior based on inherent methodological flaws) are also important to consider. If the results of FBAs can be hypothesized by simply knowing the type of FBA methodology used, the results cannot be considered valid. This finding has immediate implications for clinicians who interpret the results of any FBA without further assessment and continuous evaluation. The implication that FBA methodologies may impact the results (i.e., ascribed function of target behaviors) reinforces the notion that interventions should be continuously evaluated for effectiveness, rather than assumed effective because they are based on the ascribed function of the maladaptive behavior.

Limitations of the literature

Conducting a meta-analysis allows researchers to synthesize the findings of several primary articles that utilize single subject research to determine "general findings". However, these findings are inherently impacted by the quality of the primary articles included. The literature reviewed for the current meta-analysis contained several limitations. One limitation is the possibility that articles that are selected for publication are biased or skewed in some ways. For example, studies that report poor treatment effectiveness may go unpublished and thus the average effect sizes reported within this review represent overestimates. Also, FBAs that have undifferentiated results and are not further assessed may not be published and therefore not included in the current dataset.

In addition, many articles did not include potentially useful information about the characteristics of the participants. Basic demographic data such as race, age, and even diagnosis were often not reported in the primary articles. Location of assessment and treatment sessions was also excluded from most articles. Sharing this type of information is imperative for appropriate replication and extension in future studies. In addition, its exclusion may impact the conclusions that can be drawn in the context of a meta-analysis such as this. In addition, the use of data across multiple publications without appropriate reporting may be another issue. The lack of information provided and the possible effects of this exclusion have been reported by others (Fisher, Piazza, & Hanley, 1998). However, many researchers are still excluding important information about participants and methodological design from their studies.

Though not initially coded and recorded, follow-up reviews of a sample of the included literature indicated that less than 25% of the studies reported procedural fidelity inter-rater reliability. However, when it was reported it was typically reported only for intervention phases and never for BAs. It is difficult to make comparisons across FBA methodologies if there is no guarantee that the methods were implemented as intended. Also, data common to multiple investigations may have unintentionally been coded more than once in a quantitative synthesis such as this if not noted by the primary article author. That is, some investigators may have presented treatment outcomes on the same participants in separate published articles without acknowledging these circumstances. In some cases, articles did not meet inclusion criteria because a diagnosis of ASD or claim that participant was "autistic-like" was not explicitly stated. The lack of information presented could affect not only the results of the analyses but also attempts to generalize the findings.

Limitations of the current study

Conclusions of the review must be considered within the context of its limitations. One limitation of this research synthesis is the exclusion of unpublished studies, including unpublished theses and dissertations. It is possible that the studies included represent a skewed portion of the population and are not representative of the whole. It is also possible that published articles that met inclusion criteria may have been unintentionally excluded from the review.

Another limitation of the current study is that subgroups of both assessment and treatment types were combined throughout the analyses. For example, FAs reported as “modified” and “brief” analogue sessions were included under the FA category, along with traditional FAs. Categories were combined in order to assess the effectiveness of experimental versus non-experimental assessments rather than specific subtypes of assessment. Existing research indicates possible differences in outcomes for subtypes of experimental analyses (e.g., Hanley, Iwata, & McCord, 2003) as well as subtypes of non-experimental analyses (e.g., Arndorfer et al., 1994; Cunningham & O’Neill, 2000). Combining all types of experimental analyses under the FA category may have influenced the validity of the results for all subtypes. The same is true for the BA category. Similarly, coding intervention groups into six categories, including three groups comprised of multiple components, may not capture the differences between specific types of treatment (e.g., verbal reinforcement, tangible reinforcement).

The procedure used to calculate effect sizes used for comparison could be considered another limitation of the current study. Treatment effectiveness was summarized by examining the first baseline and last treatment phase reported in the primary studies. The choice to use these phases was necessary for legitimate comparison of non-regression-based effect sizes. For example, in some studies, several different treatments were assessed and reported in an A-B-C-D design. In this case, the rate of behavior reported in phase D was compared to the baseline data reported in phase A to determine effectiveness of treatment. However, this choice resulted in a loss of information available in published reports that may have altered effect sizes in unknown ways.

Recommendations for Future Research

The results of the current study indicate that diagnostic category impacts the assessment and treatment of individuals who engage in maladaptive behaviors. Systematic assessment of diagnostic categories and possible influencing characteristics (i.e., level of intellectual disability; level of communication ability) may help further guide treatment development for these individuals. For example, knowing the general outcome of a particular combination of participant characteristics, targeted problem behavior, and ascribed function could influence treatment selection and, in turn, outcome. As indicated as a limitation of the current study, direct comparisons of different types of experimental FAs and non-experimental BAs might be useful. For example, in the current study all methods of experimental assessments were subsumed under the FA category. Brief FAs were categorized with full-length FAs and were not directly compared with other, less time intensive assessment methodologies. Future research may include a direct comparison between specific subtypes of FAs to BAs. Another line of research that would yield comparable results would be to design a single subject study of original data collection in which comparisons are made for individuals who have been administered a multitude of assessments (e.g., an interview, MAS rating scale, brief FA, and extended FA), comparing both assessment outcome and treatment effectiveness.

References

Allison, D.B., Faith, M.S., & Franklin, R.D. (1995). Antecedent exercise in the treatment of disruptive behavior: A meta-analytic review. *Clinical Psychology: Science and Practice*, 2, 279-303.

- Allison, D.B., & Gorman, B.S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behaviour Research and Therapy*, *31*, 621-631.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: Fourth Edition Text Revision*. Washington, DC: American Psychiatric Association.
- Anderson, C.M. & Long, E.S. (2002). Use of a structured descriptive assessment methodology to identify variables affecting problem behavior. *Journal of Applied Behavior Analysis*, *35*, 137-154.
- Arndorfer, R.E., Miltenberger, R.G., Woster, S.H., Rortvedt, A.K., & Gaffaney, T. (1994). Homebased descriptive and experimental analysis of problem behaviors in children. *Topics in Early Childhood Special Education*, *14*, 64-87.
- Campbell, J.M. (2003). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: A quantitative synthesis of single-subject research. *Research in Developmental Disabilities*, *24*, 120-138.
- Campbell, J.M. (2004). Statistical comparison of four effect sizes for single-subject designs. *Behavior Modification*, *28*, 234-246.
- Campbell, J.M., Herzinger, C.H., & James, C.L. (2007). Evidenced based therapies for autistic disorder and pervasive developmental disorders. In R.G. Steele, T.D. Elkin and M.C. Roberts, (Eds.) *Handbook of evidence-based therapies for children and adolescents* (pp. 371-388). New York: Springer.
- Cunningham, E., & O'Neill, R.E. (2000). Comparison of results of functional assessment and analysis methods with young children with autism. *Education and Training in Mental Retardation and Developmental Disabilities*, *35*, 406-414.
- Dawson, J.E., Matson, J.L., & Cherry, K.E. (1998). An analysis of maladaptive behaviors in persons with autism, PDD-NOS, and mental retardation. *Research in Developmental Disabilities*, *19*, 439-448.
- Derby, K.M., Wacker, D.P., Sasso, G., Steege, M., Northup, J., Cigrand, K., & Asmus, J. (1992). Brief functional assessment techniques to evaluate aberrant behavior in an outpatient setting: A summary of 79 cases. *Journal of Applied Behavior Analysis*, *25*, 713-721.
- Didden, R., Duker, P.C., & Korzilius, H. (1997). Meta-analytic study on treatment effectiveness for problem behaviors with individuals that have mental retardation. *American Journal on Mental Retardation*, *101*, 387-399.
- Didden, R., Korzilius, H., van Oorsouw, W., & Sturmey, P. (2006). Behavioral treatment of challenging behaviors in individuals with mild mental retardation: Meta-analysis of single-subject research. *American Journal on Mental Retardation*, *111*, 290-298.
- Durand, V.M., & Crimmins, D.B. (1988). Identifying the variables maintaining self-injurious behavior. *Journal of Autism and Developmental Disorders*, *18*, 99-117.
- Durand, V.M., & Crimmins, D.B. (1992). *The motivation assessment scale (MAS) administration guide*.

- Topeka, KS: Monaco and Associates.
- Faith, M.S., Allison, D.B., & Gorman, B.S. (1996). Meta-analysis of single-case research. In R.D. Franklin, D.B. Allison, & B.S. Gorman, (Eds.), *Design and analysis of single-case research* (pp. 245-277). Hillsdale, NJ: Lawrence Erlbaum.
- Fisher, W.W., Piazza, C.C., & Hanley, G.P. (1998) Informing the reader of the presence of data common to multiple investigations. *Journal of Applied Behavior Analysis, 31*, 703-704.
- Gresham, F.M., McIntyre, L.L., Olson-Tinker, H., Dolstra, L., McLaughlin, V., & Van, M. (2004). Relevance of functional behavioral assessment research for school-based interventions and positive behavioral support. *Research in Developmental Disabilities, 25*, 19-37.
- Hall, S.S. (2005). Comparing descriptive, experimental, and informant-based assessments of problem behaviors. *Research in Developmental Disabilities, 26*, 514-526.
- Hanley, G.P., Iwata, B.A., & McCord, B.E. (2003). Functional analysis of problem behavior: A review. *Journal of Applied Behavior Analysis, 36*, 147-185.
- Herzinger, C.V., & Campbell, J.M. (2007). Comparing functional assessment methodologies: A quantitative synthesis. *Journal of Autism and Developmental Disorders, 37*, 1430-1445.
- Individuals with Disabilities Education Act Amendments (1997). Pub.L. No. 105–117, USC 1400 et seq.
- Iwata, B.A., Dorsey, M.F., Slifer, K.J., Bauman, K.E. & Richman, G.S. (1982). Toward a functional analysis of self-injury. *Analysis and Intervention in Developmental Disabilities, 2*, 3-20.
- Lundervold, D., & Bourland, G. (1988). Quantitative analysis of treatment of aggression, self-injury, and property destruction. *Behavior Modification, 12*, 590-617.
- Ma, H. (2006). An alternative method for quantitative synthesis of single-subject researches. *Behavior Modification, 30*, 598-617.
- Matson, J.L., & Vollmer, T.R. (1995). *User's guide: Questions about behavioral function (QABF)*. Baton Rouge, LA: Scientific Publishers Inc.
- Northup, J., Wacker, D., Sasso, G., Steege, M., Cigrand, K., Cook, J., & DeRaad, A. (1991). A brief functional analysis of aggressive and alternative behavior in an outclinic setting. *Journal of Applied Behavior Analysis, 24*, 509-522.
- O'Brien, S., & Repp, A.C., (1990). Reinforcement-based reductive procedures: A review of 20 years of their use with persons with severe to profound mental retardation. *Journal of the Association for Persons with Severe Handicaps, 15*, 148-159.
- Olive, M.L., & Smith, B.W., (2005) Effect size calculations and single subject designs. *Educational Psychology, 25*, 313-324.
- Pace, G.M., & Toyer, E.A., (2000). The effects of a vitamin supplement on the pica of a child with severe mental retardation. *Journal of Applied Behavior Analysis, 33*, 619-622.
- Parker, R.I., & Hagan-Burke, S., (2007). Median-based overlap analysis for single case data. *Behavior*

Modification, 31, 919-936.

- Paclawskyj, T.R., Matson, J.L., Rush, K.S., Smalls, Y., & Vollmer, T.R. (2001). Assessment of the convergent validity of the Questions About Behavioral Function scale with analogue functional analysis and the Motivation Assessment Scale. *Journal of Intellectual Disability Research, 45*, 484-494.
- Piazza, C.C., Hanley, G.P., & Fisher, W.W. (1996). Functional analysis of cigarette pica. *Journal of Applied Behavioral Analysis, 29*, 437-439.
- Scattone, D., Wilczynski, S.M., Edwards, R.P., & Rabian, B. (2002). Decreasing disruptive behaviors of children with autism using social stories. *Journal of Autism and Developmental Disorders, 32*, 535-543.
- Scotti, J.R., Evans, I.M., Meyer, L.H., & Walker, P. (1991). A meta-analysis of intervention research with problem behavior: Treatment validity and standards of practice. *American Journal on Mental Retardation, 96*, 233-256.
- Scruggs, T.E., Mastropieri, M.A., Cook, S.B., & Escobar, C. (1986). Early intervention for children with conduct disorders: A quantitative synthesis of single-subject research. *Behavioral Disorders, 11*, 260-271.
- Vollmer, T.R. (1994). The concept of automatic reinforcement: implications for behavioral research in developmental disabilities. *Research in Developmental Disabilities, 15*, 187-207.
- Yarborough, S.C. & Carr, E.G. (2000). Some relationships between informant assessment and functional analysis of problem behavior. *American Journal on Mental Retardation, 105*, 130-151.

Author contact information:

Caitlin Herzinger Delfs, PhD, BCBA-D
Marcus Autism Center
1920 Briarcliff Rd
Atlanta, GA 30329
(404) 785-9426