School-Based Interventions Targeting Challenging Behaviors Exhibited by Young Children with Autism Spectrum Disorder: A Systematic Literature Review

Jose R. Martinez, Brittany L. Werch, and Maureen A. Conroy
University of Florida

Abstract: The purpose of this review was to critically examine and summarize the impact of school-based interventions designed to decrease challenging behaviors in young children with Autism Spectrum Disorder (ASD). Reviewed studies employed a single-case experimental design, targeted challenging behaviors, included children 3–8 years old with ASD, and took place in school settings. Interventions were categorized as antecedent-based, function-based, reinforcement, instructional, or multicomponent. In addition to analyzing articles descriptively, effect size estimates were calculated using the Non-overlap of All Pairs and Tau-U methods. A total of 26 studies including 44 children were reviewed. Individual effect size estimates ranged from weak to strong, but on average, the reviewed interventions were effective in decreasing the target behaviors exhibited by the participants. Findings suggest that interventions implemented in school settings can effectively decrease challenging behaviors exhibited by young children with ASD. Directions for future research and practice are also discussed.

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that includes persistent deficits in social communication and social interactions as well as restricted and repetitive patterns of behaviors and interests (American Psychiatric Association, 2013). In addition, individuals with ASD often exhibit challenging behaviors such as physical and verbal aggression, property destruction, non-compliance, and self-injury with more frequency than individuals without disabilities, or with other developmental disabilities (Conroy, Dunlap, Clarke, & Alter, 2005a; Reichow & Barton, 2014). Young children with ASD (i.e., birth to eight years) are at increased risk for developing challenging behaviors because they often lack functional communicative behaviors and thus, use challenging behaviors to communicate their needs and wants (Conroy et al., 2005a). Challenging behaviors interfere with the ability of young children with ASD to attend to and engage in academic activities and/or engage in prosocial interactions with peers and adults (Boyd, McDonough, & Bodfish, 2012; Powell, Dunlap, & Fox, 2006). Consequently, challenging behaviors are considered one of the most significant barriers to these children’s participation in inclusive general education settings, social relationships, and community activities (Dunlap, Iovannone, Wilson, Kincaid, & Strain, 2010). Once challenging behaviors become an established part of children’s behavioral repertoires, these behaviors are not likely to decrease in the absence of timely, effective, and targeted interventions (Conroy et al., 2005a; Reichow & Barton, 2014). As the number of young children with ASD who exhibit challenging behaviors continues to increase, the need to identify research-based interventions (i.e., evidence-based practices [EBPs]) that are effective in decreasing challenging behaviors exhibited by these children becomes evident.

Legal mandates (i.e., Individuals with Disabilities Education Act of 2004 [Public Law 108-446]; No Child Left Behind Act of 2001 [Public Law 107-110]) require teachers to implement EBPs in their classrooms as a means...
to improve the academic, behavioral, and social outcomes of the children with ASD that they serve. Additionally, teachers and related service personnel are required by professional standards and ethics (e.g., Council for Exceptional Children [CEC]) to implement EBPs in their classrooms. Schools are the primary or only source of intervention for some young children with ASD exhibiting challenging behaviors (Machalicek, O’Reilly, Beretvas, Sigafos, & Lancioni, 2007). Therefore, there is a critical need to identify school-based interventions for young children with ASD exhibiting challenging behaviors that are supported by scientific evidence (i.e., EBPs). Previous reviews have identified interventions that are effective in decreasing challenging behaviors in children with ASD, but these reviews have included studies conducted in clinical and school settings (e.g., Conroy et al., 2005a; Horner, Carr, Strain, Todd, & Reed, 2002; Odom et al., 2003). Classrooms are more complex settings (e.g., greater number of children and different levels of teacher training in ASD) to implement interventions than clinical settings, so it is difficult to determine from these literature reviews the effectiveness of the reviewed interventions in school settings.

More recently, Machalicek and colleagues (2007) conducted a literature review of single-case research design studies investigating the impact of school-based interventions on the challenging behaviors exhibited by individuals with ASD between the ages of three and twenty-one. The findings suggested that the interventions implemented in the reviewed studies were successful in decreasing or eliminating the participants’ challenging behaviors (Machalicek et al., 2007). However, the authors did not report the outcomes of the studies based on the various age groups of the participants and thus, it is difficult to determine whether or not there were differential effects on the challenging behaviors exhibited by the participants based on their ages. Consequently, to our knowledge little is known about the implementation of interventions in school settings as a means to decrease challenging behaviors in young children with ASD as no review of empirical literature has been conducted to date focusing solely on this population. Given the emphasis in the recent literature of implementing interventions focused on proactive, positive strategies to decrease challenging behaviors (i.e., Positive Behavior Support [PBS]; Conroy et al., 2005a; Dunlap et al., 2010; Reichow & Barton, 2014), this review will be focused on interventions that follow that approach. These are interventions that focus on proactive, positive strategies for preventing, and decreasing challenging behaviors such as environmental arrangements of antecedent stimuli, positive consequences, and use of direct instructional strategies to teach appropriate replacement behaviors (Conroy et al., 2005a; Dunlap et al., 2010; Dunlap et al., 2003).

As the prevalence of young children with ASD exhibiting challenging behaviors in school settings continues to increase, there is greater pressure on schools and practitioners to implement EBPs that are effective in addressing those behaviors. Therefore, there is a need to determine which interventions are effective in addressing challenging behaviors in young children with ASD. The purpose of this review is to examine and summarize the current research in this area to provide direction for future research and intervention. Specifically, the purpose of this article is to: a) conduct a review of the quality of the literature base in this area using the evidence-based standards developed by the CEC (2014); b) calculate effect size estimates of the reviewed interventions using the non-overlap of all pairs (NAP) and Tau-U methods; c) determine which interventions can be considered an EBP; and d) provide direction for future research in this area.

Method

Search Procedures and Study Selection

To identify appropriate studies for review, the Education Resources Information Center (ERIC), PsychINFO, and Education Full Text electronic databases were searched for peer-reviewed articles containing the following search terms: 1) “autism” or “Asperger”; 2) “preschool” or “early childhood”; 3) “problem behavior” or “challenging behavior”; and 4) “intervention” or “early intervention.” Additionally, the reference list of four key literature reviews were examined to identify addi-
tional articles (i.e., Conroy et al., 2005a; Horner et al., 2002; Machalicek et al., 2007; Odom et al., 2003). A total of 335 articles were identified through searching the electronic databases and reference list of the aforementioned literature reviews. After removing the articles that were duplicated, 284 articles remained. The titles and abstracts of those remaining articles were screened to determine if they should be included in the review. If there was insufficient information within the titles and abstracts of the articles, the full texts were read to determine if they should be included in this review. A total of 63 full-text articles were screened for eligibility, 26 of which met the inclusion criteria set for this review.

Included studies that met criteria had child participants three to eight years of age with a diagnosis of ASD (i.e., autism, Asperger’s syndrome, or pervasive developmental disorder-not otherwise specified [PDD-NOS]), implemented an intervention focused on decreasing the challenging behaviors exhibited by the participants, employed a single-case research design to examine the effects of the intervention on the participants’ target behaviors (e.g., multiple baseline design, or reversal design), occurred between 2000 and June 2015 (following previous seminal reviews including Conroy et al., 2005a; Horner et al., 2002; Machalicek et al., 2007), were published in a peer-reviewed journal, and took place in a preschool or elementary school. Studies were excluded if they used non-experimental designs (e.g., AB designs), did not address outcomes related to challenging behaviors, or utilized group designs. Group design studies were excluded due to the expected small number of studies conducted in this area using this design, and to be able to use uniform effect size estimates (i.e., NAP and Tau-U). However, no studies using a group design were found when the search was conducted. Inclusion criteria only had to be met for one of the participants in a study for the study to be included in this review. For example, if the intervention in a study was implemented in multiple settings (e.g., school and home), the study was included in this review, but only the data of the participants who received the intervention in school settings were used.

Operational definitions and a coding form were developed to record the information extracted from the included studies. The information extracted from each article was coded into four broad categories: characteristics of the participants, characteristics of the studies, type of intervention, and methodological quality of the studies.

Characteristics of the participants. The following characteristics of the participants were coded within this category: a) age, coded as the chronological age of each of the participants; b) diagnosis, coded as autism, Asperger’s syndrome, or PDD-NOS; c) cognitive functioning, coded as above average/average (i.e., an IQ score at or above 70 on a standardized cognitive functioning assessment), below average (i.e., an IQ score below 70 on a standardized cognitive functioning assessment), or cognitive functioning not reported; and d) language/communication skills, coded as above average/average (i.e., a score of 70 or above on a standardized test of communication/language skills, and/or the presence of strong receptive and expressive skills), below average (i.e., a score below 70 on a standardized test of communication/language skills, and/or the inability to use more than two-three word utterances to communicate needs or wants), or language/communication skills not reported.

Characteristics of the studies. The following features of the studies were coded within this category: a) experimental design (e.g., alternating treatments design, reversal/withdrawal design, or multiple baseline design); b) target behaviors, coded as disruptive behaviors (e.g., noncompliance, talking out, out-of-seat behavior), destructive behaviors (e.g., property destruction, aggression, self-injurious behavior), or stereotypic behaviors (e.g., hand flapping); c) intervention setting, coded as natural (i.e., intervention was delivered in an authentic setting/situation), contrived (i.e., intervention was delivered in a situation/setting created by researchers), or combined settings (i.e., the intervention was implemented in a contrived and authentic setting for all of the participants, or implemented in a natural setting for some participants and in a contrived setting for others); d) intervention agent, coded as...
teacher or researcher; and e) functional assessment method (i.e., method used to assess the function of the participants’ challenging behaviors), coded as experimental method (e.g., functional analysis), indirect method (e.g., behavioral interview, questionnaire, or rating scale), descriptive method (e.g., direct observation), a combination of functional assessment methods, or functional assessment method not conducted.

**Type of intervention.** The interventions implemented in the reviewed studies were categorized using the following categories: antecedent-based interventions, function-based interventions, reinforcement interventions, instructional interventions, and multicomponent interventions.

Antecedent-based interventions were defined as interventions in which the participants’ environments were altered before they engaged in challenging behaviors to set the occasion for them to engage in different, non-challenging behaviors (i.e., antecedents associated with desirable behaviors were emphasized in these interventions, and antecedents associated with challenging behaviors were reduced or eliminated). These interventions include providing individuals with visual or verbal cues to forewarn them of changes in activities, or providing individuals with highly preferred items prior to an activity in which they exhibit high rates of challenging behaviors. Function-based interventions were defined as interventions in which the participants were taught socially appropriate communication responses that could be used to obtain the same reinforcers as their challenging behaviors, while simultaneously withdrawing reinforcers of their challenging behaviors (e.g., functional communication training interventions). Reinforcement interventions were defined as interventions in which reinforcements were delivered when the participants engaged in socially appropriate behaviors and were minimized or eliminated when they engaged in challenging behaviors in order to decrease the participants’ likelihood of engaging in challenging behaviors (e.g., differential reinforcement of other behavior [DRO] interventions). Instructional interventions were defined as interventions in which participants were prompted or instructed to engage in socially appropriate behaviors, or were taught to recognize relevant social and physical cues in certain contexts to help them acquire and use socially appropriate behaviors (e.g., social stories). Multicomponent interventions were defined as interventions that implemented more than one of the aforementioned interventions.

**Methodological quality of the studies.** The evidence-based standards for single-case design research developed by the CEC (2014) were used to evaluate the methodological quality of each reviewed study. In order to examine issues related to internal, external, and ecological validity of the included studies, three additional variables were incorporated in the review. These variables included whether the studies collected and reported generalization data, maintenance data, and social validity data.

The purpose of the CEC’s evidence-based standards is to provide researchers, practitioners, and stakeholders with a guide that can help them identify methodologically sound intervention studies in the field of special education and categorize the evidence base of those interventions (CEC, 2014). There is a set of eight quality indicators within these evidence-based standards that must be met by studies in order for them to be classified as methodologically sound (i.e., without meaningful threats to the validity of the findings): 1) context and setting (i.e., studies provide sufficient information regarding the critical features of the contexts and settings in which they were conducted); 2) participants (i.e., studies provide sufficient information to identify the population of participants to which results may generalize to); 3) intervention agents (i.e., studies provide sufficient information regarding the critical features of the intervention agent[s]); 4) description of the intervention (i.e., studies provide sufficient information regarding the critical features of the implemented interventions so they can be easily understood and replicated); 5) implementation fidelity (i.e., studies assess and report implementation fidelity related to adherence and dosage using direct, reliable measures); 6) internal validity (i.e., studies provide sufficient evidence that the independent variable[s] cause changes in the dependent variable[s]); 7) dependent variables (i.e., outcome measures are applied appropriately.
to gauge the effects of the implemented interventions); and 8) data analyses (i.e., data analyses are appropriately conducted). According to the CEC (2014), the conservative approach of requiring studies to meet all of the quality indicators to be classified as methodologically sound increases the likelihood that only the highest quality and most trustworthy studies are considered when categorizing the evidence base of interventions in the field of special education.

Using a rubric following the criteria specified by the CEC (2014), studies were coded as whether or not they were methodologically sound (i.e., adequately addressed the methodological issues specified by each quality indicator). Methodologically sound studies then were coded as having positive, neutral or mixed, or negative effects. According to the CEC (2014), a study is considered to have a positive effect when a functional relationship is established between the implemented intervention(s) and the target behavior(s) for at least 75% of the participants, it has a minimum of three total participants, and the data for none of the participants show evidence of a functional relationship in a non-therapeutic direction. A study is considered to have negative effects when a functional relationship in a non-therapeutic direction is established between the implemented intervention(s) and the target behavior(s) for at least 75% of the participants, and the study has a minimum of three total participants. Lastly, a study is considered to have neutral or mixed effects when the criteria for neither positive nor negative effects are met (CEC, 2014).

After classifying each reviewed study as methodologically sound or not and determining the effects of the methodologically sound studies, the evidence base of each intervention was categorized (CEC, 2014). Interventions were classified as evidence-based practices (i.e., interventions with at least five methodologically sound studies with positive effects, and at least 20 total participants across studies), potentially evidence-based practices (i.e., interventions with at least five methodologically sound studies with positive effects, and at least 20 total participants across studies), potentially evidence-based practices (i.e., interventions with at least five methodologically sound studies with positive effects, and at least 20 total participants across studies), or practices with insufficient evidence (i.e., interventions that do not meet the criteria for any of the other categories; CEC, 2014).

The CEC’s evidence-based standards were selected to evaluate the methodological quality of the reviewed studies due to the focus of these standards on special education research (CEC, 2014). In addition, these standards allow researchers, practitioners, and stakeholders to incorporate the findings of numerous methodologically sound studies into reports that summarize findings across studies, categorizing the evidence base of the interventions implemented in those studies. Other evidence-based standards (e.g., What Works Clearinghouse [WWC] evidence-based practice standards; WWC, 2014) require researchers, practitioners, and stakeholders to mostly focus on judging the evidence of individual studies making it more difficult to summarize findings across studies and categorize the evidence base of the interventions.

Effect Size Estimates

Two nonparametric effect size estimates were calculated to provide an estimate of the magnitude of intervention effect. Effect size estimates were calculated using the NAP (Parker & Vannest, 2009) and Tau-U (Parker, Vannest, Davis, & Sauber, 2011a) methods. According to Brossart, Vannest, Davis, and Patience (2014), effect size estimates have advantages and disadvantages and thus, it is better to calculate more than one effect size estimate when synthesizing the literature. Moreover, it is important to pair effect size estimates with visual analysis of the data because current effect size estimates do not fully address the complexity of single case research design (e.g., autocorrelation, variability, number of available data points; Whalon, Conroy, Martinez, & Werch, 2015). NAP and Tau-U
were chosen because they closely align with visual analysis and do not require test assumptions that are not met in single case research (i.e., random sampling from a population that is normally distributed, samples that are independent from one another, and two or more groups with equal variances; Parker, Vannest, & Davis, 2011b). Both NAP and Tau-U are not as sensitive to outliers and small data points since they include all data points in their calculations (Parker et al., 2011b), and both have been found in the literature to yield stronger statistical power than other nonparametric effect size estimates (e.g., percentage of non-overlapping data; Whalon et al., 2015).

NAP compares each baseline phase data point to each intervention phase data point to obtain a percentage of non-overlapping data that shows improvement (i.e., NAP = [Pos + .5 X Ties] / Total; Parker et al., 2011b). Effect size estimates calculated with NAP can be interpreted as ineffective (i.e., below .50), questionable (i.e., between .51 and .70), effective (i.e., between .71 and .90), and very effective (i.e., above .91; Parker et al., 2011b). Tau-U is the percent of data that demonstrate improvement over time by comparing all data points in each phase ([Pos–Neg]/Pairs; Parker et al., 2011a). Tau-U can be interpreted using the guidelines set out by Rakap (2015) as ineffective (i.e., below .50), questionable (i.e., between .51 and .65), effective (i.e., between .66 and .92), and very effective (i.e., above .93). Tau-U and NAP are similar, but NAP is the percentage of non-overlap whereas Tau-U is the percentage of non-overlap minus overlap with the ability to control for trend (Parker et al., 2011a), making Tau-U the more conservative effect size estimate (Whalon et al., 2015). Effect size estimates were calculated using a web-based calculator (Vannest, Parker, & Gonen, 2011). The .pdf files of the studies, along with Biosoft’s UnGraph for Windows (version 5.0), were used to extract the data from the graphs of all of the reviewed studies in order to calculate the effect size estimates. Although the data from UnGraph is often reliable and valid, in some instances the symbols on the graphs overlap and UnGraph can miss or misrepresent data points (Whalon et al., 2015). Therefore, the data extracted from UnGraph was compared with visual analysis to ensure reliability and validity. Errors in the extracted data were changed to better represent visual analysis. UnGraph data representing the number of sessions, frequency counts, and percentages were rounded to the nearest whole number (Shadish et al., 2009).

Effect size estimates were calculated for all of the reviewed studies, not only the ones found to be methodologically sound through the CEC’s evidence-based standards, to facilitate the quantitative comparison of the effectiveness of the implemented interventions across studies. Specifically, given the small number of studies that met inclusion criteria for this review, an even smaller number of methodologically sound studies were expected. Therefore, calculating effect size estimates only for methodologically sound studies would have made it impractical to compare the effectiveness of the interventions across studies. For example, calculations of the effect size estimates for each type of intervention, based on participants’ specific characteristics, would have been unfeasible using just methodologically sound studies (e.g., effect size estimates of instructional interventions on the participants with an average/above average cognitive ability).

Reliability

Once an initial determination was made by the first author as to whether or not an article met the inclusion criteria, the second author independently applied the same inclusion/exclusion criteria to a randomly selected 20% (n = 57) of the articles. In the case of a disagreement, the study in question was reread and discussed by the authors until an agreement was reached. After the final list of studies to be included in the review was obtained, the first author extracted and coded all of the relevant data from each study into the four aforementioned categories. The second author then independently extracted and coded all of the target data from a random sample of 20% (n = 5) of the included studies. There were 40 items across the four coding categories, per study, on which there could be agreements or disagreements. Percentage agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements, and multiplying by 100. Interobserver agreement across all
categories was 89% (range = 78–100%). However, when a disagreement occurred between the authors, the study in question was discussed until a consensus was reached.

Results

Characteristics of the Participants

A total of 26 studies published between 2002 and 2013 investigating the impact of school-based positive behavioral interventions on the challenging behaviors exhibited by 44 children with ASD were reviewed. The age of the participants ranged between 3 and 8 years with a mean age of six years. Table 1 provides descriptive information about the participants included in this review. Thirty-five (80%) participants were school-aged (i.e., over 5 years old), and nine (20%) participants were preschool-aged (i.e., 3 to 5). Descriptions of cognitive ability were provided for 37 (84%) of the participants. Out of those 37 participants, twenty-eight (76%) were described as performing below average while nine (24%) were described as performing in the above average range. With regards to language/communication skills, most participants (n = 34, 77%) were reported as having below average receptive and expressive language skills. Ten (23%) participants were reported as having above average/average receptive and expressive language skills. Table 2 provides information on effect size estimates by characteristics of the participants and type of intervention.

Characteristics of the Studies

The majority of studies employed a multiple baseline or multiple probe design (n = 11, 42%) to assess the effects of the implemented intervention on the participants’ target behaviors, while nine (34%) studies applied a reversal design, and six (24%) studies applied an alternating treatments design. Destructive behaviors were targeted in 12 (46%) of the studies and disruptive behaviors and stereotypic behaviors were targeted each in seven (27%) of the studies. Twenty-two (85%) of the studies were conducted in natural settings, while the rest of the studies (n = 4, 15%) were conducted in contrived settings. Interventions were implemented by teachers in 12 (46%) of the studies, and implemented by researchers in 14 (54%) of the studies. Most of the studies conducted assessments in order to determine the function of the participants’ target behaviors and thus, guide the development and implementation of the interventions. More specifically, 13 (50%) of the studies determined the function of the participants’ target behaviors utilizing experimental methods, three studies (12%) used direct observation methods, five (19%) studies used indirect methods, and two (8%) studies used combined methods. Three (11%) studies did not conduct assessment methods to determine the function of the participants’ target behaviors. Maintenance and generalization were each measured and reported in six (23%) studies. Finally, eight studies (31%) measured and reported information related to the social validity of the intervention.

Type of Intervention and Methodological Quality of the Studies

Eleven (43%) studies implemented instructional interventions, six (23%) studies implemented function-based interventions, four (15%) studies implemented antecedent-based interventions, three (12%) studies implemented reinforcement interventions, and two (7%) studies implemented multicomponent interventions. Table 1 provides information on the type of intervention implemented in each study. Table 3 provides a summary of the methodological quality ratings of the studies.

Instructional interventions. Nineteen (42%) children, with an average age of six (range 3–8), participated in instructional interventions. Instructional interventions were developed to target disruptive behaviors (e.g., screaming, yelling, crying, or loud humming; Agosta, Graetz, Mastroianni, & Scruggs, 2004; Crozier & Tincani, 2005; Iskander & Rosales, 2013; Ozdemir, 2008; Scattone, Wilczynski, Edwards, & Rabian, 2002), stereotypic behaviors (e.g., vocal stereotypy; Ahearn, Clark, MacDonald, & Chung, 2007; Brownell, 2002; Miguel, Clark, Tereshko, & Ahearn, 2009), and destructive behaviors (e.g., pushing, biting or kicking; Frea, Arnold, & Vittimberga, 2001; Machalicek et al., 2009; Mancil, Haydon, & Whitby, 2009).
<table>
<thead>
<tr>
<th>Reference</th>
<th>N</th>
<th>Age</th>
<th>Cognitive Functioning</th>
<th>Language Skills</th>
<th>Design</th>
<th>Target Behavior</th>
<th>Intervention</th>
<th>NAP</th>
<th>Tau-U</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agosta et al. (2004)</td>
<td>1</td>
<td>6</td>
<td>BA</td>
<td>BA</td>
<td>Reversal</td>
<td>Disruptive</td>
<td>Instructional</td>
<td>.76</td>
<td>.23</td>
</tr>
<tr>
<td>Ahearn et al. (2007)</td>
<td>3</td>
<td>3-7</td>
<td>BA</td>
<td>NR</td>
<td>Reversal</td>
<td>Stereotypic</td>
<td>Instructional</td>
<td>1.0</td>
<td>.98</td>
</tr>
<tr>
<td>Blair et al. (2007)</td>
<td>1</td>
<td>6</td>
<td>BA</td>
<td>BA</td>
<td>MBL</td>
<td>Destructive</td>
<td>Function-based</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Braithwaite et al. (2000)</td>
<td>1</td>
<td>7</td>
<td>BA</td>
<td>BA</td>
<td>MBL</td>
<td>Destructive</td>
<td>Function-based</td>
<td>.99</td>
<td>.98</td>
</tr>
<tr>
<td>Brownell (2002)</td>
<td>2</td>
<td>6</td>
<td>AA</td>
<td>AA</td>
<td>Reversal</td>
<td>Stereotypic</td>
<td>Instructional</td>
<td>.95</td>
<td>.90</td>
</tr>
<tr>
<td>Buckley and Newchok (2005)</td>
<td>1</td>
<td>7</td>
<td>BA</td>
<td>NR</td>
<td>Reversal</td>
<td>Destructive</td>
<td>Function-based</td>
<td>.95</td>
<td>.90</td>
</tr>
<tr>
<td>Carnett et al. (2014)</td>
<td>1</td>
<td>7</td>
<td>AA</td>
<td>AA</td>
<td>Alt. treatment</td>
<td>Disruptive</td>
<td>Reinforcement</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Conroy et al. (2005)</td>
<td>1</td>
<td>6</td>
<td>AA</td>
<td>AA</td>
<td>Alt. treatment</td>
<td>Stereotypic</td>
<td>Antecedent-based</td>
<td>.90</td>
<td>.64</td>
</tr>
<tr>
<td>Crozier and Tincani (2005)</td>
<td>1</td>
<td>8</td>
<td>AA</td>
<td>AA</td>
<td>Reversal</td>
<td>Disruptive</td>
<td>Instructional</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Falconata et al. (2013)</td>
<td>1</td>
<td>7</td>
<td>BA</td>
<td>NR</td>
<td>Reversal</td>
<td>Destructive</td>
<td>Function-based</td>
<td>1.0</td>
<td>.97</td>
</tr>
<tr>
<td>Frea et al. (2001)</td>
<td>1</td>
<td>4</td>
<td>BA</td>
<td>BA</td>
<td>MBL</td>
<td>Destructive</td>
<td>Instructional</td>
<td>.93</td>
<td>.87</td>
</tr>
<tr>
<td>Haley et al. (2010)</td>
<td>1</td>
<td>8</td>
<td>AA</td>
<td>AA</td>
<td>Alt. treatment</td>
<td>Stereotypic</td>
<td>Antecedent-based</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Iskander and Rosales (2013)</td>
<td>1</td>
<td>8</td>
<td>AA</td>
<td>NR</td>
<td>MBL</td>
<td>Disruptive</td>
<td>Instructional</td>
<td>.79</td>
<td>.59</td>
</tr>
<tr>
<td>Lang et al. (2010)</td>
<td>4</td>
<td>4-7</td>
<td>BA</td>
<td>BA</td>
<td>Alt. treatment</td>
<td>Stereotypic</td>
<td>Antecedent-based</td>
<td>.85</td>
<td>.68</td>
</tr>
<tr>
<td>Machalicek et al. (2007)</td>
<td>2</td>
<td>6-7</td>
<td>BA</td>
<td>BA</td>
<td>MBL</td>
<td>Destructive</td>
<td>Instructional</td>
<td>.77</td>
<td>.54</td>
</tr>
<tr>
<td>Mancil et al. (2009)</td>
<td>3</td>
<td>6-8</td>
<td>BA</td>
<td>BA</td>
<td>Reversal</td>
<td>Destructive</td>
<td>Instructional</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Miguel et al. (2009)</td>
<td>1</td>
<td>4</td>
<td>BA</td>
<td>NR</td>
<td>Reversal</td>
<td>Stereotypic</td>
<td>Instructional</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>O’Neil et al. (2001)</td>
<td>1</td>
<td>6</td>
<td>BA</td>
<td>BA</td>
<td>MBL</td>
<td>Disruptive</td>
<td>Function-based</td>
<td>.91</td>
<td>.49</td>
</tr>
<tr>
<td>Ozdemir (2008)</td>
<td>2</td>
<td>7-8</td>
<td>AA</td>
<td>AA</td>
<td>MBL</td>
<td>Disruptive</td>
<td>Instructional</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Reichle et al. (2010)</td>
<td>2</td>
<td>4</td>
<td>BA</td>
<td>BA</td>
<td>Alt. treatment</td>
<td>Destructive Reinforcement</td>
<td></td>
<td>.92</td>
<td>.85</td>
</tr>
<tr>
<td>Rispoli et al. (2011)</td>
<td>3</td>
<td>5-6</td>
<td>BA</td>
<td>BA</td>
<td>Alt. treatment</td>
<td>Destructive</td>
<td>Antecedent-based</td>
<td>1.0</td>
<td>.89</td>
</tr>
<tr>
<td>Scattone et al. (2002)</td>
<td>2</td>
<td>7</td>
<td>BA</td>
<td>BA</td>
<td>MBL</td>
<td>Disruptive</td>
<td>Instructional</td>
<td>.76</td>
<td>.30</td>
</tr>
<tr>
<td>Schindler and Horner (2005)</td>
<td>3</td>
<td>4-5</td>
<td>BA</td>
<td>BA</td>
<td>MBL</td>
<td>Destructive</td>
<td>Function-based</td>
<td>.88</td>
<td>.77</td>
</tr>
<tr>
<td>Strain et al. (2011)</td>
<td>2</td>
<td>5-8</td>
<td>BA</td>
<td>BA</td>
<td>MBL</td>
<td>Destructive</td>
<td>Multicomponent</td>
<td>.96</td>
<td>.93</td>
</tr>
<tr>
<td>Taylor et al. (2005)</td>
<td>1</td>
<td>4</td>
<td>BA</td>
<td>BA</td>
<td>Reversal</td>
<td>Stereotypic Reinforcement</td>
<td></td>
<td>.79</td>
<td>.59</td>
</tr>
<tr>
<td>Waters et al. (2009)</td>
<td>2</td>
<td>6</td>
<td>BA</td>
<td>BA</td>
<td>MBL</td>
<td>Destructive</td>
<td>Multicomponent</td>
<td>.92</td>
<td>.83</td>
</tr>
</tbody>
</table>

**Note.** BA = Below average; AA = Above average; NR = Not reported; MBL = Multiple baseline.
Social stories were the instructional intervention most frequently implemented in the reviewed studies (i.e., implemented by 63% of the studies categorized as instructional interventions). Other instructional interventions consisted of teaching participants socially appropriate behaviors that were incompatible with their challenging behaviors (Ahearn et al., 2007; Miguel et al., 2009), or using visual aids to teach participants socially appropriate behaviors (Frea et al., 2001; Machalicek et al., 2009). The effect size estimates of instructional interventions were highly variable (NAP ranging from .76 to 1.0; Tau-U from .23 to 1.0). However, the mean NAP was .92 (range .77–1.0) and Tau-U was .84 (range .72–1.0), indicating that overall instructional interventions are effective in decreasing the challenging behaviors exhibited by young children with ASD.

**Methodological quality.** Out of the eleven studies that implemented instructional interventions, five (45%) did not collect treatment fidelity data (Agosta et al., 2004; Ahearn et al., 2007; Brownell, 2002; Frea et al., 2001; Miguel et al., 2009), and one (10%) did not report all of the necessary data to meet the quality indicators related to internal validity (Iskander & Rosales, 2013). Therefore, only five studies (45%) were considered to be methodologically sound and were used to determine whether instructional interventions can be considered EBP. Out of the five methodologically sound studies, three had positive effects (Crozier & Tincani, 2005; Mancil et al., 2009; Ozdemir, 2008). The other two studies had mixed/neutral effects (Machalicek et al., 2009; Scattone et al., 2002). These studies were classified as having mixed/neutral effects because they did not have a minimum of three participants (CEC, 2014). Therefore, instructional interventions are considered to potentially be an EBP in terms of their effectiveness in decreasing challenging behaviors exhibited by young children with ASD in school settings (i.e., more than four methodologically sound studies implemented instructional interventions, but the ratio of studies with positive effects to studies with neutral/mixed effects was less than 2:1).

**Function-based interventions.** A total of eight (17%) children, with an average age of six (range 4–7), participated in the six studies that implemented a function-based intervention. All of the studies targeted destructive behaviors (e.g., hitting, kicking, or pushing; Blair, Umbreit, Dunlap, & Jung, 2007; Braithwaite & Richdale, 2000; Buckley & Newchok, 2005; Falcomata, Muething, Gainey, Hoffman, & Fragale, 2013; O’Neil & Sweetland-Baker,
implemented functional communication training interventions. These interventions consist of teaching individuals socially appropriate communicative behaviors that serve the same function as their challenging behaviors in order to replace their challenging behaviors. Overall, function-based interventions were effective in decreasing the challenging behaviors exhibited by the participants (i.e., mean NAP of .96 [range .88–1.0] and Tau-U of .85 [range .49–1.0]).

**Methodological quality.** Out of the studies that implemented function-based interventions, four (66%) did not collect implementation fidelity data (Braithwaite & Richdale, 2000; Buckley & Newchok, 2005; Falcomata et al., 2013; O’Neil & Sweetland-Baker, 2001). Only two studies (34%) met all of the CEC’s quality indicators and thus, were considered to be methodologically sound studies (Blair et al., 2007; Schindler & Horner, 2005). Both of these studies had positive effects. Given that there were less than five methodologically sound studies with positive effects that implemented function-based interventions (i.e., the minimum of studies needed for an intervention to be considered EBP), function-based interventions are considered to potentially be an EBP (CEC, 2014).

**Antecedent-based interventions.** Nine (20%) children, with an average age of five (range 4–8), participated in antecedent-based interventions. These interventions were implemented to target stereotypic behaviors (e.g., vocal stereotypy; Conroy, Asmus, Sellers, & Ladwig, 2005b; Haley, Heick, & Luiselli, 2010; Lang et al., 2010) and destructive behaviors (e.g., biting or attempts to bite; Rispoli et al., 2011). Antecedent-based interventions consisted of providing participants with unrestricted access to items that maintained their challenging behaviors prior to activities in which they exhibited high rates of challenging behaviors (Rispoli et al., 2011), embedding preferred materials into activities in which participants exhibited high rates of challenging behaviors (Lang et al., 2010), or providing participants with visual cues to indicate to them times in which it was acceptable and times in which it was not acceptable for them to engage in challenging behaviors (Conroy et al., 2005b; Haley et al., 2010). Overall, an-

### Table 3

<table>
<thead>
<tr>
<th>CEC’s EBP Criteria</th>
<th>Quality Indicators</th>
<th>Studies that adequately addressed each indicator</th>
<th>Studies that did not adequately address each indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>Participants</td>
<td>n = 26 (100%)</td>
<td>n = 0</td>
</tr>
<tr>
<td>Intervention Agent</td>
<td>Intervention Description</td>
<td>n = 26 (100%)</td>
<td>n = 0</td>
</tr>
<tr>
<td>Intervention Fidelity</td>
<td>Outcome Measures</td>
<td>n = 26 (100%)</td>
<td>n = 0</td>
</tr>
<tr>
<td>Internal Validity</td>
<td>Data Analysis</td>
<td>n = 26 (100%)</td>
<td>n = 0</td>
</tr>
<tr>
<td>Social Validity</td>
<td>General Maint.</td>
<td>n = 26 (100%)</td>
<td>n = 0</td>
</tr>
</tbody>
</table>

2001; Schindler & Horner, 2005) and implemented functional communication training interventions. These interventions consist of teaching individuals socially appropriate communicative behaviors that serve the same function as their challenging behaviors in order to replace their challenging behaviors. Overall, function-based interventions were effective in decreasing the challenging behaviors exhibited by the participants (i.e., mean NAP of .96 [range .88–1.0] and Tau-U of .85 [range .49–1.0]).

**Methodological quality.** Out of the studies that implemented function-based interventions, four (66%) did not collect implementation fidelity data (Braithwaite & Richdale, 2000; Buckley & Newchok, 2005; Falcomata et al., 2013; O’Neil & Sweetland-Baker, 2001). Only two studies (34%) met all of the CEC’s quality indicators and thus, were considered to be methodologically sound studies (Blair et al., 2007; Schindler & Horner, 2005). Both of these studies had positive effects. Given that there were less than five methodologically sound studies with positive effects that implemented function-based interventions (i.e., the minimum of studies needed for an intervention to be considered EBP), function-based interventions are considered to potentially be an EBP (CEC, 2014).

**Antecedent-based interventions.** Nine (20%) children, with an average age of five (range 4–8), participated in antecedent-based interventions. These interventions were implemented to target stereotypic behaviors (e.g., vocal stereotypy; Conroy, Asmus, Sellers, & Ladwig, 2005b; Haley, Heick, & Luiselli, 2010; Lang et al., 2010) and destructive behaviors (e.g., biting or attempts to bite; Rispoli et al., 2011). Antecedent-based interventions consisted of providing participants with unrestricted access to items that maintained their challenging behaviors prior to activities in which they exhibited high rates of challenging behaviors (Rispoli et al., 2011), embedding preferred materials into activities in which participants exhibited high rates of challenging behaviors (Lang et al., 2010), or providing participants with visual cues to indicate to them times in which it was acceptable and times in which it was not acceptable for them to engage in challenging behaviors (Conroy et al., 2005b; Haley et al., 2010). Overall, an-
ecedent-based interventions were effective in decreasing the challenging behaviors exhibited by the participants with a mean NAP of .94 (range .85–1.0) and Tau-U of .80 (range .68–1.0).

Methodological quality. Out of the four studies that implemented antecedent-based interventions, one (25%) did not collect implementation fidelity data (Conroy et al., 2005b). The other three studies (75%) met all of the quality indicators and had positive effects (Haley et al., 2010; Lang et al., 2010; Rispoli et al., 2011). Therefore, antecedent-based interventions are considered to potentially be an EBP (CEC, 2014).

Reinforcement interventions. A total of four (9%) children, with a mean age of five (range 4–7), participated in the studies that implemented reinforcement interventions. Destructive behaviors (e.g., hitting or biting others; Reichle, Johnson, Monn, & Harris, 2010), disruptive behaviors (e.g., screaming; Carnett et al., 2014), and stereotypic behaviors (e.g., vocal stereotypy; Taylor, Hoch, & Weissman, 2005) were each targeted by one study. Reinforcement interventions involved the delivery of preferred tangibles or edibles to the participants for engaging in socially appropriate behaviors as a means to reduce their challenging behaviors. Overall, these interventions were effective in decreasing the challenging behaviors exhibited by the participants with a mean NAP of .91 (range .79–1.0) and Tau-U of .82 (range .59–1.0).

Methodological quality. Out of the three studies categorized as implementing reinforcement interventions, two (66%) met all of the quality indicators and had positive effects (Carnett et al., 2014; Reichle et al., 2010). The other study (34%) did not collect implementation fidelity data (Taylor et al., 2005). Since there was only one methodologically sound study with positive effects, reinforcement interventions are considered to have insufficient evidence in terms of their effectiveness in decreasing challenging behaviors exhibited by young children with ASD in school settings.

Multicomponent interventions. Four (9%) children, with an average age of five (range 5–8), participated in multicomponent interventions. Destructive behaviors (e.g., kicking or biting others) were targeted by the two studies that implemented multicomponent interventions (Strain, Wilson, & Dunlap, 2011; Waters, Lerman, & Hovanetz, 2009). One study used visual schedules to teach routines to the participants and reinforce their appropriate behaviors (while ignoring their challenging behaviors) in order to decrease transition related challenging behaviors (Waters et al., 2009). The other study created index cards that listed behavioral expectations for the activities in which participants exhibited high rates of challenging behaviors, taught participants to look at the index cards to remind themselves of what they needed to do during these activities (i.e., self-management), and reinforced them for following the behavioral expectations (Strain et al., 2011). These interventions were effective in decreasing the challenging behaviors exhibited by the participants with a mean NAP of .94 (range .92–.96) and Tau-U of .88 (range .83–.93).

Methodological quality. Only one study (50%) met all of the CEC quality indicators and thus, was considered to be methodologically sound (Strain et al., 2011). The other study (50%) did not collect implementation fidelity data (Waters et al., 2009). Since there was only one methodologically sound study with positive effects, multicomponent interventions are considered to have insufficient evidence in terms of their effectiveness in decreasing challenging behaviors exhibited by young children with ASD in school settings.

Discussion
As the prevalence of young children with ASD exhibiting challenging behaviors in school settings continues to increase, there is a critical need to identify educational interventions that are supported by scientifically-rigorous evidence (i.e., EBPs). This review focused on synthesizing the literature on interventions implemented in school settings as a means to decrease challenging behaviors exhibited by young children with ASD. A total of 26 articles published within the past 15 years met the inclusion criteria set for this review. An encouraging finding was that overall all of the reviewed interventions were effective in decreasing the challenging behaviors exhibited by the participants. However, none of the interventions could be considered an EBP due to the small literature base in this area (i.e.,
less than five methodologically sound studies with positive effects, with more than 20 total participants across studies, were conducted for each type of intervention; CEC, 2014). Therefore, more research in this area is needed. Further research assessing the effectiveness of school-based interventions aimed at decreasing challenging behaviors exhibited by young children with ASD would aid in the identification of interventions that are supported by scientifically rigorous evidence.

Research suggests that many young children with ASD who exhibit challenging behaviors lack appropriate communication skills or task-related behaviors needed to thrive academically and socially and thus, need to be systematically taught those skills and behaviors (Conroy et al., 2005a). Consequently, an optimistic finding was the emphasis in the literature on targeting socially appropriate behaviors in addition to the challenging behaviors exhibited by the participants. In addition, to be considered effective from a clinical standpoint, interventions should not only reduce engagement in challenging behaviors, but also increase engagement in socially appropriate behaviors that will ultimately facilitate individuals’ inclusion into broad societal contexts (Dunlap et al., 2010). To that end, researchers are encouraged to continue to develop and evaluate interventions in which challenging behaviors and socially appropriate behaviors are targeted.

It is known that challenging behaviors (e.g., stereotypic behaviors) manifest in the early childhood years in individuals with autism (Boyd et al., 2012) and that these behaviors interfere with their everyday activities, including opportunities to socially interact with peers, engage in academic instruction, and perform learned tasks accurately (Boyd et al., 2012). Therefore, researchers are encouraged to continue to conduct studies with young children with ASD to find interventions that are effective in improving their developmental outcomes (Boyd et al., 2012). It is important to note that the majority of the participants included in this review were described as performing below average cognitively and socially. Therefore, more research is needed to determine the impact of the implemented interventions on the challenging behaviors exhibited by young children with ASD who demonstrate a wide range of cognitive abilities and language/communication skills.

Most of the reviewed studies conducted functional behavioral assessments to determine the function of the participants’ challenging behaviors (i.e., 23 [89%] studies conducted functional behavioral assessments) and indicated a direct link between functional assessment outcomes and the implemented interventions. This finding is important because research suggests that the effects of interventions linked to the outcomes of functional behavioral assessments tend to be more durable (Conroy et al., 2005a). Moreover, most of the studies that conducted functional behavioral assessments (n = 13, 57%) conducted experimental methods. Researchers are encouraged to continue using experimental functional assessment methods to identify the function of the challenging behaviors exhibited by young children with ASD as other assessment methods have limitations (Patterson, Smith, & Jelen, 2010). For example, research has shown that descriptive assessment methods (e.g., direct observations) often produce a high proportion of false positives for the attention function (Patterson et al., 2010). Research has found that attention is often correlated with challenging behaviors (e.g., stereotypic behaviors) regardless of their function(s) because practitioners and parents usually provide attention (e.g., a verbal reprimand) when individuals engage in those behaviors. Similarly, the information gathered though indirect assessment methods (e.g., teacher/parents interviews) can be highly subjective, prone to bias, and produce inaccurate information and thus, these assessments may erroneously identify the function(s) of challenging behaviors (Miltenberger, 2008).

Even though all of the reviewed studies were conducted in school settings, the majority of the interventions were implemented by researchers rather than teachers (i.e., authentic change agents). This suggests that teachers were not an integral part of the interventions implemented in the studies, thus decreasing the external and social validity support for the use of these interventions within schools by teachers. Researchers are encouraged to conduct studies in which teachers are the intervention agents. This will further expand our knowledge regarding the practicality, feasibil-
ity, and efficacy of teachers as implementers of these interventions in order to decrease challenging behaviors in young children with ASD. Moreover, the increasing emphasis on implementing EBPs in school settings highlights the need for future research examining the efficacy of behavioral interventions as implemented by teachers within their classrooms.

Lack of measurement generalization, maintenance, and social validity was a limitation found in most of the studies, which decreases the generalizability of the research. Without measurement of generalization, it is unknown whether the implemented interventions would be effective in settings other than the ones in which the interventions were implemented, or if the interventions would be effective when implemented by different intervention agents. Therefore, future research should gather generalization data by measuring the effectiveness of the implemented interventions across settings and intervention agents. Similarly, future research should include data on the maintenance of the implemented interventions. The duration of the effects of an intervention could have important implications for practice, as a short duration of an effect could limit the validity and feasibility of interventions. Lastly, without social validity data it is difficult to validate the appropriateness of the behaviors targeted by the interventions and to determine whether or not the outcomes of the interventions were socially acceptable, relevant, and useful to the participants and/or their families. Therefore, researchers are encouraged to collect data regarding the social validity/acceptability of the interventions implemented in their studies.

Limitations of the Literature Review

One possible threat to the validity of this literature review is publication bias (Rothstein, Sutton, & Borenstein, 2005). Publication bias occurs when researchers do not submit for publication studies with negative findings or journals do not publish studies that fail to produce positive results (Rothstein et al., 2005). The findings of this review should be interpreted with caution as publication bias may have influenced the results (e.g., the impact of the reviewed interventions on the participants’ challenging behaviors). For example, this could be the reason why none of the reviewed studies had negative effects when the CEC (2014) evidence-based standards were applied to assess their methodological quality and effects of the interventions implemented in them. In addition, this review included a small number of studies and thus, there may be insufficient data to draw firm conclusions about the effectiveness of the reviewed interventions on the challenging behaviors exhibited by the participants. Lastly, it is possible that some published articles that could have met the inclusion criteria set for this review may not have been captured in the electronic databases search due to the search terms used (e.g., the term “PDD-NOS” was not used when conducting the electronic databases search). Bibliographic searches of the included articles or archival hand searches of professional journals to identify articles related to behavioral interventions, early childhood, and authors who typically publish on these topics were not conducted. Therefore, this literature review may not be exhaustive.

A limitation of using the CEC’s evidence-based standards to determine whether or not the reviewed interventions could be considered EBPs is that these standards require studies to address all of the indicators in order to be classified as methodologically sound. This strict requirement could limit the consideration of studies conducted before quality indicators for single-case design research were developed and emphasized in published studies (CEC, 2014). Reviewers must evaluate the methodological quality of studies based on the information reported in manuscripts, which often do not fully contain all aspects of studies (Cook et al., 2014). Therefore, the application of these evidence-based standards is limited by accurate and complete reporting of research methods in published articles.

Conclusion

Challenging behaviors (e.g., aggression or stereotypic behaviors) constitute substantial obstacles to the participation of children with ASD in inclusive settings, such as general education classrooms or community activities (Dunlap et al., 2010). Challenging behaviors interfere with the ability of children with ASD
to interact socially with peers; therefore, a high priority for enabling children with ASD to benefit from inclusive settings is to implement school-based interventions that are effective in decreasing challenging behaviors. The findings of this review suggest that young children with ASD exhibiting challenging behaviors can benefit from positive behavioral interventions implemented in school settings. However, the research base in this area is small and firm conclusions about the effectiveness of the reviewed interventions in decreasing challenging behaviors exhibited by young children with ASD in school settings cannot be drawn at this time.

Based on the findings of this review, it is cautiously recommended that teachers implement the reviewed interventions in their classrooms as a means to decrease challenging behaviors exhibited by their students with ASD. However, when implementing these interventions, teachers are encouraged to use formative data collection in order to determine whether or not the interventions are effective for their students’ target behaviors.

References

*References marked with an asterisk were included in the review.


Rakap, S. (2015). Effect sizes as result interpretation aids in single-subject experimental research: Description and application of four nonoverlap
methods. *British Journal of Special Education, 42*(1), 11–33. doi: 10.1111/1467-8578.12091


Received: 1 July 2015
Initial Acceptance: 1 September 2015
Final Acceptance: 24 November 2015