Teaching Object Play to Young Children With Disabilities:

A Systematic Review of Methods and Rigor

Erin E. Barton, Ph.D., BCBA-D
Vanderbilt University
Nashville, TN, USA

Rebecca Murray, B.S.
Vanderbilt University
Nashville, TN, USA

Cecelia O'Flaherty, B.S.
Vanderbilt University
Nashville, TN, USA

Erin M. Sweeney M.Ed.
Vanderbilt University
Nashville, TN, USA
Ms. Sweeney is now at Williamson County Schools in Franklin, TN, USA.

Stephanie Gossett, M.Ed.
Vanderbilt University
Nashville, TN, USA
Ms. Gossett is now at Access Family Services in Ashville, NC, USA.

Author note: The authors wish to acknowledge the assistance of Gounah Choi and Evelyn Mauldin in the development of this manuscript.

Corresponding author information: Erin E. Barton, Vanderbilt University, 230 Appleton Place, Peabody 228, Nashville, TN, 37203, erin.e.barton@vanderbilt.edu

Financial disclosure: Preparation of this article was supported in part by the U.S. Department of Education, Office of Special Education and Rehabilitative Services, Grants H325K140110 and H325K140308.

Abstract

In this systematic review, we examined the rigor and outcomes across 27 object play intervention studies using single case research methodology. We focused on studies including children age 5 years or younger and examined several descriptive characteristics including materials, instructional packages, and settings. We also analyzed the facilitation and measurement of generalized play and several methodological features including quality, rigor, and visual analysis procedures. Overall, the identified studies demonstrated positive outcomes, although quality and rigor limited interpretations of the outcomes. Previous reviews also have noted strong outcomes and weak to moderate quality for single case studies. Our results should be interpreted with caution given previous reviews of play intervention studies identified strong outcomes and quality from group design studies. Additional replications testing robust interventions using single case research with strong methodological rigor are warranted.

Teaching Object Play to Young Children With Disabilities: A Systematic Review of Methods and Rigor

Play is a critical early developmental milestone that contributes to the learning and wellbeing of young children (American Academy of Pediatrics [AAP], 2007; Ginsburg, the Committee on Communications, & the Committee on Psychosocial Aspects of Child and Family Health, 2007). Play is particularly important for children with disabilities as they learn important skills within playful interactions with adults and their peers (U.S. Departments of Health and Human Services and Education, 2015). Early experiences that promote creative expression and increasing complexity of play provide a foundation for future development, learning, and academic success. In the past decade, educators, parents, and researchers have expressed concern about the documented decreased time that young children spend engaged in unstructured free play (AAP, 2007; Ginsburg et al., 2007). Recent efforts have been made to reverse this trajectory. Early childhood settings should ensure that all children have multiple opportunities and the needed supports for engaging in sustained play of increasing complexity (National Association for the Education of Young Children, 2009).

Lifter, Mason, and Barton (2011) identified multiple, specific benefits of play. First, play provides a context in which other evidence based and recommended practices can be embedded. Adequate play skills may increase learning opportunities for young children. Second, play predicts and might be a behavioral cusp for other important skills. For example, researchers have documented a relation between play and language development (Lewis, 2003; Vig, 2003), increases in children's vocalizations (Barton & Wolery, 2010; Frey & Kaiser, 2011), and social skills (Freeman, Gulsrud, & Kasari, 2015; Gulsrud, Helleman, Freeman, & Kasari, 2014; Kasari, Gulsrud, Freeman, Paparella, & Helleman, 2012; Toth, Munson, Meltzoff, & Dawson, 2006).

Researchers also have documented relations between complex play skills and academic achievement (Hanline, Milton, & Phelps, 2010; Wolfgang, Stannard, & Jones, 2001). In addition, play has a practical benefit in that it provides a context for meaningful interactions with others across settings, which promotes independent participation and engagement. Also, if children engage in sustained play behaviors, they are less likely to engage in challenging behaviors (Machalicek et al., 2009; Nuzzolo-Gomez, Leonard, Ortiz, Rivera, & Greer, 2002). Play deficits can be particularly debilitating because play is a primary, normalized context for interactions with caregivers and peers, exploring the environment, and learning and practicing skills (Strain, Schwartz, & Bovey, 2008; Yu, Ostrosky, & Fowler, 2014).

What is play?

There are two primary types of play represented across existing play taxonomies: social and object play. In 1931, Parten developed a social play taxonomy that remains relevant and useful today. She defined play relative to the child's interactions with peers across the following categories: (a) onlooker, (b) solitary, (c) parallel, (d) associative, and (e) cooperative. Conversely, there are several different play taxonomies that define play relative to the child's interactions with objects (e.g., toys). These play taxonomies include the following categories of object play: (a) sensori-motor, (b) relational, (c) functional, and (d) symbolic play. Sensori-motor play includes simple exploration of objects (Lifter et al., 2005; Van Berckelaer-Onnes, 2003; Ungerer & Sigman, 1984). Relational play includes the non-functional combinations of two or more objects (Lifter et al., 2005; Ungerer & Sigman, 1984). Functional play is defined as using objects as they were intended (e.g., coloring with crayons, putting a puzzle together; Barton, 2016). Pretend or symbolic play typically includes the following subcategories of play includes non-literal behaviors with some symbolism (Ungerer & Sigman, 1984; Van Berckelaer-Onnes,

2003; Williams et al., 2001). Across play taxonomies, play categories encompass behaviors requiring a range of skills, follow a developmental progression, and are not mutually exclusive (Barton, 2010; Piaget, 1952; Smilansky, 1968; Sherratt & Peter, 2002; Van Berckelaer-Onnes, 2003). Both social and object play skills provide a critical tool and context for engaging with peers and participating in typical early childhood settings. Thus, play is a *critical* intervention goal and should be intentionally taught using evidence-based practices.

Current State of Play Intervention Research

Many children learn to engage in increasingly complex play in high quality early childhood environments (e.g., child care, home, preschool). For many children, providing time for unstructured play and interactions with nurturing, responsive adults and socially competent peers is sufficient for facilitating increasingly complex play. However, research has consistently shown that children with disabilities engage in less complex and fewer play behaviors than their typically developing peers when given the same materials in the same settings (Wilson et al., 2017). Children with autism spectrum disorder (ASD), in particular, consistently demonstrate less varied and frequent pretend play than children with typical development or other disabilities (Charman & Baron-Cohen, 1997; Sigman & Ruskin, 1999). Children with disabilities will require intentional, systematic instruction to learn appropriate play skills (Barton & Wolery, 2008; Thiemann-Bourque, Brady, & Fleming, 2012).

There are numerous, consistent reviews of social play interventions with considerable empirical support demonstrating the efficacy of adult-directed or peer-mediated social skills interventions and emerging evidence for environmental arrangements and novel instructional technologies (Joseph et al., 2016). Likewise, reviews of the object play intervention research have consistently concluded that adult modeling and prompting are related to increased play

behaviors in children with disabilities (Barton & Wolery, 2008) including children with ASD (Jung & Sainato, 2012; Kasari, Chang, & Patterson, 2013; Kasari, Freeman, & Paparella, 2006; Lang et al., 2009; Stahmer, Ingersoll, & Carter, 2003). These reviews consistently identify socially-mediated object play interventions using naturalistic strategies and prompting (including video modeling) as an evidence-based practice. Group design studies included in these reviews were particularly strong in quality, rigor, and outcomes; whereas, single case research studies were characterized with lower quality and rigor (Kasari, Chang, & Patterson, 2013). However, these reviews of object play have focused either on pretend play (Barton, 2010; Barton & Wolery, 2008) or children with ASD exclusively (Jung & Sainato, 2012; Kasari et al., 2006; Lang et al., 2009; Stahmer et al., 2003). Further, there are few overlapping studies across these reviews and few examined the quality or rigor of included research using contemporary research standards (cf. Kasari & Chang, 2014). This might be particularly important for reviews including single case research given the rapidly advancing methodology for synthesizing and conducting single case research (Ganz & Ayers, 2018; Ledford & Gast, 2018).

There is currently no comprehensive review of object play interventions for children with disabilities. A comprehensive review of object play is important to address the existing inconsistencies in object play definitions and ensure balanced information is available to guide future research and practice regarding the object play interventions. We attempted to address this need by conducting a systematic synthesis of interventions focused on teaching any type of object play to children with disabilities using single case methodology. The focus on single case methodology allowed for a nuanced examination of methodological quality and child behavior change. Further, there is no current mechanism for combining or comparing effect sizes across single case and research designs, although research is burgeoning in this area (Shadish, Horner,

Hedges, & Odom, 2015). The following research questions guided our systematic review: (1) Do interventions focused on object play behaviors result in positive and generalized behavior change for young children with disabilities?, (2) What instructional procedures have been used to teach object play?, and (3) To what extent have object play interventions been evaluated using contemporary, rigorous SCD and visual analysis procedures?

Method

We adhered to the guidelines for systematic reviews proposed by Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher, Liberati, Tetzlaff, Altman, & PRISMA Group, 2009). A meta-analysis of these studies was conducted as part of a larger study focused on meta-analytic advancements in single case research (Odom, Barton, Reichow, Pustejovsky, & Swaminathan, 2018).

Eligibility Criteria and Study Identification

To be eligible for inclusion in this review, studies were required to meet several criteria. First, the study evaluated the effects of an intervention on object play skills of children age 5 years or younger. Second, object play skills were measured and reported. Third, the study used an experimental single case design (i.e., at least three opportunities to demonstrate behavior change at three different points in time). We conducted an electronic search of the following databases in March of 2016: Educational Resources Information Center [ERIC], ProQuest, PsycINFO, and PubMed. Key terms used in the electronic search included: *((infant* OR toddler* OR preschool* OR child*)* AND *(teach OR intervention OR prompt OR practice OR strategy OR therapy OR program OR procedure OR instruction) AND (object play OR toy play OR pretend play OR symbolic play OR functional play) and (single case OR single subject OR case study OR withdrawal OR multiple baseline OR multiple probe OR alternating treatment).* Our key terms

were guided by the PICOS framework (Schardt, Adams, Owens, Keitz, & Fontelo, 2007). We used the following limits: peer reviewed and English language only.

We then used a five-step screening procedure to identify studies. First, we conducted a title and abstract screen. Second, we conducted a full-text screen of the studies identified through the initial screening. Two graduate students completed steps one and two independently and had 95% agreement on study identification; disagreements were resolved through consensus. Third, we conducted an ancestral search with all articles identified in Step 2. Finally, we reviewed the reference lists of four recent reviews of play interventions (Barton & Wolery, 2008; Barton, 2010; Jung & Sainato, 2012; Lang et al., 2009; Lifter et al., 2011). Twenty-seven studies within 25 different articles were identified for inclusion in this review. A PRISMA flow diagram (Moher et al., 2009) of study inclusion and exclusion is provided in Appendix A.

Study Coding Procedures

Four concurrent coding procedures were used. Two special education graduate students and one special education undergraduate were trained to code all study variables. First, coders extracted descriptive information from all studies related to the following variables: participant and setting characteristics, target outcomes and measures, materials, instructional procedures and dosage, baseline and maintenance characteristics, and study design. Second, coders extracted data related to specific features of the instructional procedures that facilitate generalization. Third, coders extracted methodological characteristics and used a multi-step, visual analysis process for coding single-case study outcomes. Finally, coders assessed the methodological rigor of all studies using the What Works Clearinghouse (WWC) single case design standards as outlined in the *Procedures and Standards Handbook 4.0* (2017) and the *Single-Case Analysis and Design Framework* (SCARF; Ledford, Lane, Zimmerman, Chazin, & Ayers, 2016).

Descriptive characteristics. Descriptive characteristics were coded for all identified studies using a systematic coding protocol created for this study. All entries were independently coded and discrepancies were discussed until consensus was reached. There were discrepancies on fewer than 2% of the entries across coded variables (agreement range = 98-100%). Coders initially extracted information regarding the number, age, race/ethnicities, disability status, descriptions of participants' functional play repertoires, country in which the study occurred, and instructional settings.

Dependent variables. Five items related specifically to the dependent variables were coded: (a) primary types of play skills targeted, (b) categorical domain of the play skills, (c) operational definitions of object play, (d) other related skills measured, and (e) the measurement system(s) used. The categories for the primary categories of play skills were sensorimotor play, relational play, functional play, functional play with pretense, object substitution, assigning absent attributes, imagining absent objects, and social pretend play. These categories and their definitions were adapted from established play taxonomies which were chosen to ensure a range of object and social play categories demonstrated by young children could be coded (Barton, 2010; Parten, 1931; Smilansky, 1968; Sherratt & Peter, 2002).

Measurement. Two items related specifically to the measurement of the dependent variable. First, we coded the following play-related skills: play sequences, vocalizations, play diversity, and independent play. Second, the measurement procedures for the primary dependent variable were coded as: direct observation or standardized assessments (or both). Studies using direct observation were further coded as using timed event recording, event recording, percentage correct (i.e., accuracy), or an interval system.

Instructional and study procedures. Several items related specifically to the instructional procedures (i.e., independent variable). The name of the instructional procedure was extracted along with the authors' brief description. We also coded the specific instructional types based on previous play intervention reviews (Barton & Wolery, 2008; Jung & Sainato, 2012; e.g., system of least prompts, choice, video modeling) and created categories based on the interventions used in the included studies. We further coded the types of prompts used: physical modeling, full physical prompts, visual cues, and gestures. We also extracted the type of reinforcement used along with the authors' brief description. We coded several items related to the instructional materials used. We coded the following categories: housekeeping, dolls, vehicles, junk toys (i.e., vague objects with no clear function), blocks, figures, toy sets, games, and computers. These material types were adapted from recent reviews of play interventions (Barton & Wolery, 2008; Jung & Sainato, 2012). The coders also extracted information related to the maintenance procedures per study. Initially they evaluated if maintenance was measured. When measured, they further coded the specific procedures used during the maintenance conditions and indicated if the conditions were the same as baseline conditions. We also coded the single-case studies as a withdrawal; multiple baseline across participants, behaviors, or stimuli; multiple probe across participants, behaviors, or stimuli; or alternating treatment design.

Participant disabilities. We also conducted several summative analyses per coded variable to examine differences between studies that included children with ASD exclusively (n = 13), studies that included children with ASD and children with other disabilities (n = 5), and studies that exclusively included children with other disabilities (n = 9). However, we do not report these for two reasons. First, there were no discernable differences between these groups.

Second, the disparate numbers of studies within each group (i.e., 13, 5, and 9) makes the results difficult to interpret. The results are listed in Table 1.

Expanded generalization analysis. To evaluate the extent to which study authors facilitated and measured generalized object play behaviors, the graduate coders extracted information related to the generalized measurement and prompting procedures. The coders used a yes/no scoring system to indicate if the study authors measured object play across: peers in the same classroom, peers in a different classroom, teachers, toys with the same function, toys with a different function, toys with the same form, toys with different forms, settings, and activities or routines. The coders also used a yes/no scoring system to indicate if the implementers facilitated generalized object play by teaching across peers, implementers, materials, activities, settings, common stimuli; using sufficient exemplars; using systematic prompt fading, thinning reinforcement; or using naturally maintaining consequences. Codes were developed based on Stokes and Baer's (1977) conceptualization of generalization.

Methodological rigor and quality. Two tools for evaluating the rigor and quality of single case research (WWC, 2017; SCARF, Ledford et al., 2016) were used to examine the rigor of the object play research. All entries were independently coded by the special education graduate students and discrepancies were discussed with the first author until consensus was reached. There were discrepancies on fewer than 5% of the entries across coded variables (agreement range = 95-100%).

What Works Clearinghouse single case design standards. The coders applied the WWC single case study design standards as outlined in the *WWC Procedures and Standards Handbook 4.0* (2017). The WWC single case design standards were created to assist in identifying EBPs and as a complement to previously created group design standards (Kratochwill et al., 2013). The

tool is based on quality indicators identified by Horner and colleagues (2005) and uses a hierarchal framework with three levels. The first level includes items relating to manipulation of the independent variable, dependent measure reliability, and the number of demonstrations of behavior change. The second level evaluates the number of data points per condition whereby a design *meets standards* with five or more per condition, *meets standards with reservations* with three or four data points in any one condition, and *does not meet standards* with fewer than three data points in any one condition. The third level classifies studies that meet design standards with or without reservations as having no evidence, moderate evidence, or strong evidence of effectiveness (Kratochwill et al., 2013; WWC, 2017).

Single-Case Analysis and Design Framework. SCARF is a synthesis tool used to evaluate SCD using a hierarchical framework (Ledford et al., 2016). Designs are only evaluated if at least three potential demonstrations of effect are present. Scores of 0-4 are possible in 10 categories; these scores are generated via responses to several yes/no questions for 7 categories (data sufficiency, reliability, fidelity, social and ecological validity; condition, participant, and dependent variable descriptions) and are generated via yes/no questions and 0-4 categorical ratings for 3 categories (maintenance, response generalization, and stimulus generalization measurement). The ten categories are divided into rigor and quality/breadth of measurement. Following rigor and quality assessments, outcomes are assessed separately for primary, maintained, and generalized effects, also on a 0-4 scale. Additional information on calculations of study rigor, quality, and outcomes can be found at (http://vkc.mc.vanderbilt.edu/ebip/scarf/).

Visual analysis. The graduate student coders used a multi-step process for coding single-case study outcomes. First, coders used a systematic visual analysis protocol (adapted from Gast & Spriggs, 2014; Kratochwill et al., 2013) to assess the presence or absence of a functional

relation for each case. A *case* was conceptualized as at least 3 opportunities to demonstrate behavior change (i.e., for object play behaviors) at three different points in time within or across participants, target object play behaviors, settings, or materials. Second, coders reviewed and entered the results as described by the study authors. Third, the first author reviewed the coders' assessment of functional relations and the authors' description of results and coded as agreement or disagreement. This multistep coding process was established to reduce the likelihood that author bias impacted their assessment of functional relations. Finally, the coders analyzed the researchers' use of visual analysis terms (i.e., level, trend/slope, variability, stability, immediacy of change, non-overlap, overlap, behavior change, consistency), means/averages, and visual analysis procedures to describe data patterns for the dependent variables. The visual analysis terms selected were based on established procedures (Gast & Spriggs, 2014; WWC, 2013).

Meta-analysis

We also attempted to conduct a meta-analysis of the studies using two between-case standardized effect size approaches that are appropriate for use with single case research data (Pustejovsky, Hedges, & Shadish, 2014; Swaminathan, Rogers, & Horner, 2014). However, only 9 of the 27 studies included in this review met the design criteria for using these approaches. The criteria allows for withdrawal designs with a minimum of three cases and multiple baseline or probe across participants designs. The limited number of studies that met these criteria precludes definitive conclusions regarding the meta-analyses. The specific results and the methodological implications are discussed in detail in a separate manuscript (see Odom, Barton, Reichow, Swaminathan, & Pustejovsky, 2018).

Results

Participants and Settings

The total number of participants within each included study ranged from one to seven. In the aggregate, eighty-three total participants were included; 58 were male and 25 were female. Across all studies, 95% of participants had disabilities. The ages of participants with disabilities ranged from 26 to 117 months with the average participant age of 48.80 months. Race was reported in six studies. For 23 participants for which race was reported, 14 were European American/Caucasian/White and nine were African American/Black. Of the 79 participants with disabilities, ASD was the most commonly reported disability (n = 51, 65%). Eighteen (66.7%) of the 27 studies included children with ASD; 13 of these exclusively included children with ASD and not children with other disabilities (see Table 1). Six participants were at risk for language delays or problem behaviors, seven participants had developmental delays, five participants had Down syndrome, and 10 participants were reported as having other disabilities such as developmental delay, cerebral palsy, intellectual disability, language delay, Hirschsprung disease, communication impairment, and myoclonic epilepsy. Three participants had multiple disabilities. Researchers in 16 of the 27 studies described the functional play repertoire of participants. Twenty-five studies were conducted in the United States, one was conducted in Australia, and one was conducted in New Zealand. Twelve of the 27 studies were conducted in an inclusive classroom setting, twelve studies were conducted in a separate testing or therapy room, six studies were conducted in self-contained classrooms, and three studies were conducted in the participant's home. Five of these studies used a combination of settings.

Target Outcomes and Measures

All of the 27 studies measured object play as a primary or secondary variable as per the inclusion criterion. Twenty-five (96%) of the studies included operational definitions of object play. The primary target skills of the articles ranged from toy play—which generally involved

the manipulation of toys—to imitating play, and social and pretend play. Fifteen of the studies measured some form of pretend play with objects. Of those, 12 addressed multiple types of pretend play. Four of the 15 studies that addressed pretend play measured sensory motor play, four measured relational play, and nine measured functional play. Seventeen measured functional play with pretense, eight measured object substitution, twelve measured assigning absent attributes, seven measured imagining absent objects, and five measured at social pretend play. Researchers in 7 of the 27 studies measured sequences of object play, 4 measured diversity of play, and 17 measured and defined vocalizations related to play. Researchers in 14 of the 27 studies used event recording or timed event recording to measure object play. Researchers in 9 of the 27 studies used partial interval recording. Others used momentary time sampling, percentage correct play actions, or highest play level observed.

Instructional Procedures

Researchers in all 27 studies described their instructional procedures. Ten studies used the system of least prompts (SLP). For example, Lang and colleagues (2014) used SLP to teach object play skills and used gestural, model, verbal, and physical prompts. Researchers in 6 of the 27 studies used video modeling and one showed the participants a play script on an iPad (Murdock, Ganz, & Crittendon, 2013). For example, Dupere and colleagues (2013) used video modeling with substitutable loops to increase diversity of play. Researchers in 3 of the 27 studies delivered the intervention within small groups or used peer mediated instruction (PMI). For example, Craig-Unkeffer and Kaiser (2002) taught object play and communication skills to dyads of children in which both children were considered at-risk for behavior problems. In four studies, researchers described using systematic modeling and prompting without a specific hierarchy or procedure (Ballard & Medland, 1986; DiCarlo, Reid, & Stricklin, 2003; Friman,

1990; Taylor & Iacono, 2003). In two studies, researchers used pivotal response training (PRT; Stahmer, 1995; Thorp et al., 1996). Frey and Kaiser (2011) used play expansions and contingent imitation to increase levels of object play. Researchers in 15 studies used a physical model prompt, 13 used hand over hand prompts, two used a visual prompt, and three used choice prompts. 13 studies used other specific prompts, which included verbal prompts, verbal models, and gestural prompts. Researchers in 20 of the 27 studies used reinforcement or feedback. Researchers in 19 studies used verbal praise and one used verbal mapping and imitation. Eight of those studies also used edibles. Two studies also used related toys as a reinforcer and one used touch. Eighteen studies delivered the reinforcement directly after each target behavior. Barton (2015) and Lang and colleagues (2014) gradually thinned reinforcement during intervention. Researchers in 2 of the 27 studies delivered reinforcement on a fixed interval.

Materials

Researchers in all 27 studies provided at least a partial description of their materials. Researchers in 11 studies included housekeeping toys (e.g., plates, cups, mops, brooms), 14 had dolls, 19 had vehicles, 6 had junk toys, 12 had blocks, 13 had other figures, 13 used themed sets, 3 used computers, 0 used games, and 19 had other toys (e.g., puzzles, figures, balls, & books). Researchers in 6 of the 27 studies measured play during centers or free play in the classroom and used toys that were already available in the classroom such as dramatic play centers and books.

Maintenance

Maintenance (i.e., post intervention) was measured in 16 studies. Thirteen of these 16 studies described maintenance conditions with replicable precision. Researchers in 10 of these 13 studies used procedures identical to baseline conditions.

Study Designs

Study design was reported for all 27 studies. Eight of these studies used a multiple baseline across participants design, six used multiple probe across participants, four used multiple probe across behaviors, two studies used a multiple baseline across behaviors design, and one used an A-B-A-B design. Two studies used combination designs: MacDonald and colleagues (2010) used a multiple probe across behaviors with an embedded A-B-C-A-B-C-A-C-B-C design and Lang and colleagues (2014) used a multiple baseline across behaviors with an embedded A-B-A-C design. Four studies used other designs including multiple baseline across toy sets (n = 1), multiple probe across toys (n = 2), and an alternating treatment design (n = 1). In eight out of 27 studies, procedural fidelity was reported for at least 20% of baseline and intervention conditions across participants, with results above 90%.

Baseline Procedures

Researchers in 26 of the 27 of the studies included a replicable description of baseline or initial probe conditions. Seven studies included prompts in baseline, 3 included reinforcement (and all praised correct responses), and 23 included responding to appropriate social initiations. Researchers across all 27 studies provided time for the children to play with the instructional materials (e.g., toys) without the intervention during baseline or initial probe conditions. Researchers in 3 of the 27 studies prompted the participant to choose a classroom center during baseline conditions. Researchers in 5 of the 27 studies told the implementers to play with the child and use prompts as they typically would during baseline conditions (Barton et al., 2013; DiCarlo et al., 2003). Researchers in 2 of the 27 studies prompted the children to play by modeling target behaviors or identifying a peer engaged in the target behavior, labeling the behavior, and looking expectantly at the child. Researchers in 4 of the 27 studies told

implementers to talk about the participants play behaviors; in 2 of these studies the implementers also contingently imitated the participants' play (Barton, 2015; Frey & Kaiser, 2011).

Expanded Generalization Analysis

Measuring generalization. Response generalization was measured in 17 of the 27 studies. Sixteen of these studies measured object play generalization across peers from the same classroom as the target child, and one study measured generalization across peers from a different classroom. Seventeen of these studies measured generalization across adults who were the target child's teachers, and six studies measured across adults familiar to the target child. Fifteen studies measured generalization across toys and materials that were different from the primary intervention toys and materials, but had the same general form and function. Eight studies measured generalization across toys and materials that varied in form, but not function, while five studies measured generalization across toys and materials that differed in both form and function from the primary intervention toys and materials. Eleven studies measured generalization across settings within the child's classroom, such as across different centers, and seven studies measured across different classroom settings in the same school. Six studies measured generalization across therapy or clinic room settings in the same school, and three studies measured across settings in the child's home. Six studies measured generalization across different activities, such as dramatic play and circle time.

Facilitating generalization. Stimulus generalization was facilitated across all 27 studies. Nine studies facilitated stimulus generalization across different peers, and seven studies facilitated generalization across different implementers. Twenty-three studies facilitated generalization across different toys and materials, and 6 facilitated generalization across different activities. Ten studies facilitated generalization across different settings, and all 27 studies

facilitated generalization across common stimuli. Eighteen studies facilitated generalization using sufficient exemplars. Seven studies facilitated stimulus generalization by using systematic prompt fading (facilitating play in contexts with fewer and no prompts), and 1 facilitated generalization using variable contingencies. Twenty-three studies facilitated stimulus generalization by using natural maintaining contingencies. Two studies facilitated stimulus generalization using systematic reinforcement thinning.

Study Rigor and Quality

WWC Standards. Eighteen of the 27 studies met WWC single case design standards. Eleven of these 18 met with reservations and 7 met without reservations. There were 47 designs (opportunities to identify a functional relation) across the 27 studies related to increasing object play behaviors. Thirty-one of the opportunities occurred within studies that met design standards with or without reservations and were further coded for strength of the evidence. Thirteen of these 31 opportunities provided strong evidence, 6 provided moderate evidence, and 12 provided no evidence. Overall, moderate or strong effects occurred for 19 cases across 11 studies and 8 different research groups, which very closely approaches the WWC summative criterion for evidence-based practices (i.e., 20 cases, 5 studies, 3 research groups; Horner et al., 2005).

SCARF summaries. Forty-seven designs were analyzed using the SCARF framework across the 27 studies (Ledford et al., 2016). SCARF includes an analysis of quality and rigor and outcomes on a 0-4 scale. An intervention is classified as evidence-based if the majority of studies are high quality with evidence of positive effects (top right quadrant of the scatterplot) and no high quality studies that show evidence of negative or null effects are present (lower right quadrant of the scatterplot). The average quality/rigor score across all studies was a 1.6 (standard deviation = 0.6, range = 0.0 – 2.9) indicating low quality and rigor on average across this

research (see Table 3 and Figure 2). The primary outcome and quality rigor scores across all 47 designs are depicted in Figure 2. Overall, 33 of the 47 designs had low ratings for quality and rigor (a score lower than 2 and upper and lower left quadrants). Twenty eight (60%) of the 47 designs yielded strong primary outcomes, which is indicated by scores in the top quadrants. However, only 13 (28%) designs had high quality evidence of positive effects (strong outcome and quality/rigor ratings), which is indicated by scores in the upper right quadrant. The remaining 15 designs with strong primary outcomes had low quality rigor ratings (upper left quadrant). There was one design with strong evidence of minimal or non-effects, which is indicated in the lower, right quadrant (Frey & Kasier, 2011). Conversely, there were 14 designs with low quality evidence of positive effects (upper left quadrant) and 20 designs with low quality evidence of minimal or non-effects (lower left quadrant), which limits confidence and interpretations of the results. The results for generalized and maintained outcomes demonstrated similar patterns with most designs having low quality and rigor (left quadrants of middle and lower panels in Figure 2). Fifteen designs had moderate to strong rigor and quality for generalized outcomes (middle panel, right quadrants of Figure 2), but few (n = 5) of these had moderate to strong effects for generalization outcomes (upper right quadrant). Six designs had moderate to strong rigor and quality for maintained outcomes (see lower panel of Figure 2), but only two of these had strong maintained outcomes (upper right quadrant). The remaining designs had strong maintained outcomes but low quality and rigor (n = 22, upper left quadrant), or no maintained outcomes and low quality and rigor (n = 25, lower left quadrant).

Visual Analysis

Object play outcomes. The subsequent analyses were conducted using visual analysis procedures. There were 47 opportunities to identify a functional relation for object play and

related behaviors across the 27 studies. The coders identified a functional relation for 23 (49%) opportunities across 15 studies using visual analysis. This analysis did not match with the study authors' reported results for 9 of the 47 opportunities. We aggregated visual analysis results across specific intervention types when possible. The system of least prompts was used in ten studies and functional relations were identified in seven of these; six of these seven met WWC standards with or without reservations. Video modeling was the next most frequently reported intervention with six studies and functional relations were identified in just one of these studies; this study did not meet WWC standards. PMI and PRT were used in three and two studies, respectively; however, functional relations were not identified for any cases across these five studies. Seven studies with functional relations specifically reported using reinforcement; in all of these studies, reinforcement was used with SLP.

Terms and procedures. There were a total of 69 reported dependent variables across the 27 studies. Researchers reported the *mean, average,* or *median* of data for 30 of the 69 variables (43%) across 15 of the 27 studies. The same number, 30 of the 69 dependent variables (43%), described the *level* of the data, which represents 16 different studies. Researchers reported the *trend/slope* for 18 of the 69 dependent variables (26%) across 13 of the 27 studies. Researchers reported the data *stability* for 9 of the 69 dependent variables (13%) and the data *variability* for 16 of the 60 dependent variables (23%). *Immediacy of change* was used to describe 10 (14%) of the dependent variables, *overlap/non-overlap* was used to describe 8 (12%) of the dependent variables, *consistency* was used to describe 5 (9%) of the dependent variables, and *behavior change* was used to describe 16 (23%) of the dependent variables. Barton (2015) reported the most visual analysis terms—using 8 out of 10 visual analysis characteristics for two dependent variables; Murdock and colleagues used 6 of the 10 visual analysis terms. For those researchers

who provided research questions or hypotheses, ten of the 27 included the expected data pattern in their primary research question or hypotheses. Researchers reported evaluating data patterns within conditions for 56 (81%) of the 69 dependent variables and between conditions for 62 (90%) of the 69 dependent variables. Researchers described conducting a vertical analysis for 21 (36%) of the 59 dependent variables within a time-lagged or tiered single case research design. Finally, only three of the authors provided a reference for their visual analysis procedures (Barton, 2015; Barton et al., 2013; Murdock et al., 2013). The references included single case text books (Gast, 2010; Kazdin, 1982), a technical report (Kratochwill et al., 2010), and a peer-reviewed publication (Parker, Cryer, & Byrns, 2006).

Meta-analysis

The limited number of studies included in the meta-analysis limit the results. Overall, the effect sizes were positive and five were statistically distinguishable from zero (i.e., Barton et al., 2013; Frey & Kaiser, 2011; Lang et al., 2014; Nuzzolo-Gomez et al. 2002; Stanton-Chapman & Brown, 2015). The studies with evidence of a functional relation had relatively larger between case effect size estimates (see Odom et al., 2018 for more details regarding these results).

Discussion

Overall, this research supports the use of adult modeling and prompting to support improvements in object play in young children with disabilities of whom almost two thirds (65%) had ASD. There is strong, emerging support for the use of the system of least prompts to teach object play, which has been documented as effective in previous reviews (Barton & Wolery, 2008; Kasari & Chang, 2014). Further, there were no discernable differences across studies that included children with ASD and those that did not. However, the authors of included studies documented improvements in object play with varying magnitudes of behavior change,

few studies with strong rigor also demonstrated maintained play outcomes, and fewer than half of the studies met contemporary single case design standards.

Overall, a majority of researchers used intervention packages with components that are known to facilitate generalized behavior change (e.g., training sufficient exemplars [multiple play behaviors], programming common stimuli [using a variety of common toys], and using naturally maintaining contingencies [e.g., the adult comments on play and played with the child]) (Stokes & Baer, 1977; Stokes & Ones, 1989). The materials used were typically found in early childhood classrooms (e.g., dolls, vehicles, blocks). However, an inadequate proportion of studies had high rigor *and* evidence for generalization and maintenance. Also, authors across these studies used descriptive statistics (i.e., mean, average) as often as visual analysis terms to describe their single case data. Thus, additional research is needed to develop more robust interventions that produce lasting, generalized change using rigorous, quality single case research and visual analysis procedures.

The results of this systematic review are significant for several reasons. In particular, these studies affirm the importance of play, the limited play skills demonstrated by children with disabilities, and the efficacy of *direct teaching* of play. Given the critical importance of play—perhaps as its own developmental domain (Lifter et al., 2011)—for children with and without disabilities, play can and should be taught to children who have delays in play. Although play is often defined as spontaneous or "internally driven," all children are in fact taught to play. Most children acquire play skills (e.g., how to engage in meaningful ways in their environments with objects, adults, and peers) in high quality environments with responsive adults and competent peers. However, some children need systematic instruction to learn to play. Similar to skills across other domains (e.g., communication, social initiations), children who are learning to play

might not emit unprompted play at the same rates and levels as their peers. This review highlighted that adult modeling and prompting, particularly in the context of a least to most prompting hierarchy, is effective for teaching play to children with heterogeneous characteristics. Future research should continue to build on the efficacy of these studies to develop more robust, effective interventions that result in frequent, generalized play.

The majority of the object play studies included in this review taught some type of pretend play. No identified studies specified how play targets were chosen. For example, Lifter and colleagues (1993, 2005) used the Developmental Play Assessment (Lifter, 2000) to select play target levels and activities; however, these studies were not included in the review due to a lack of an experimental design. That is, although pretend play is developmentally appropriate for children with typical development, pretend play might not be developmentally appropriate for children with disabilities given the mismatch between their chronological age and the acquisition of some skills. Individual play targets might produce more efficient and effective learning for young children with disabilities and should be examined in future, rigorous research. Further, there is lack of research on the emergence of play skills for children with disabilities; additional research is needed examining the scope and sequence of play instruction.

Procedural Outcomes

The system of least prompts or a least to most prompting procedure was the most common and successful instructional procedure used across these studies. The system of least prompts might be particularly effective for teaching play because it allows the adult an opportunity observe the child's play and deliver a prompt related to what the child is doing. This might increase the likelihood that adults prompt preferred play behaviors, which might increase the child's attention or motivation thereby enhancing learning. In this review, authors in all 27

studies facilitated generalization during instruction to some extent. This is an important strength of this body of research and critical for developing a robust intervention package (Stokes & Baer, 1977). However, fewer studies actually measured generalized outcomes, which limits conclusions regarding the efficacy of the interventions. The instructional materials used were appropriate and all authors used developmentally appropriate toys that are likely to be present in typical early childhood settings, which strengthens the ecological validity of this research. However, only three of the authors of included studies conducted preference assessments to ascertain child preferences for toys or objects prior to starting play instruction (DiCarlo et al., 2003; Lang et al., 2009; Reinhartsen et al., 2002) and only two of these completed preference assessments pre-baseline. Child preferences for objects or toys might enhance motivation and improve outcomes. Future replications are needed with a focus on examining child preferences to select materials and comparing interventions using preferred and non-preferred toys.

Methodological Outcomes

Overall, the low quality and rigor of these studies highlights the need for quality research in this area. For example, the lack of maintenance measurements makes comparisons to baseline conditions difficult. Using the SCARF tool, several studies did not meet minimum standards for procedural fidelity, description of the measurement system, or sufficiency of data (specific data are available from the first author via email). Future research should be designed to meet and exceed contemporary design standards. Likewise, across these studies, authors generally did not use visual analysis terms to analyze results. This is particularly concerning given all studies included in this review used SCR and visual analysis is currently the only way to identify functional relations. Contemporary SC design standards include systematic visual analysis guidelines (Kratochwill et al., 2013); further, visual analysis is the primary mechanism for

analyzing SCR data. Several authors in this review reported descriptive statistics to report data, which might have impacted their summative analysis of results and identification of functional relations. Further, few authors described conducting a vertical analyses, which are required for tiered designs (e.g., multiple baseline and probe). We conducted independent visual analyses and disagreed with study authors for 9 design decisions, which might be linked to differences in visual analysis procedures. Guidelines for visual analyses should be widely disseminated, ecologically valid, and consider the range of professionals and researchers who conduct SCR.

Limitations

There are several important limitations to note. First, we limited our search criteria to published single case studies. Although this allowed for a focus on specific methodological issues, this limited the scope of our findings. We did not calculate an effect size given that would have been outside the scope of the current review. Further, there are equivocal findings regarding effect sizes for SCR, many of the studies in this review would not fit current analyses, and controversy exists regarding the comparison of effect sizes from group studies and single case research (cf. Odom, Barton, Reichow, Swaminathan, & Pustejovsky, 2018; Shaddish, Hedges, Horner, & Odom, 2015; Wolery, 2013). However, the exclusion of group studies might have biased or altered the results. For example, several researchers have identified effective play interventions using group research designs (e.g., Bernard-Opitz, Ing, & Kong, 2004; Kasari et al., 2006, 2012). Further, the exclusion of unpublished documents (e.g., dissertations, theses) indicates risk of publication bias might be high. The SCARF analyses support the presence of publication bias in this review. The upper panel of Figure 2 shows the quality and rigor of studies for primary outcomes; there are very few studies in the lower right quadrant, which represents studies with high rigor but weak effects and indicates the potential for missing studies. These

biases should be considered when interpreting our results. Also, we did not conduct a forward search, which might have limited the search and resulted in missed studies.

There also were issues related to coding. The manner in which generalization was coded did not allow for rating the extent to which generalization was systematically facilitated or if effective generalization procedures were used. Also, we did not code issues related to the ecological validity of the studies (e.g., implementer's role), which might be useful for informing practice. We used WWC and SCARF to capture the rigor and quality of the included studies; however, the WWC is known to be overly conservative and not appropriate for SCR (Wood, 2014). Further, neither the WWC nor the SCARF have been validated. Finally, the first author also was an author of 3 of the 27 studies, which might have biased results for these studies.

Future Research

Participant diagnostic and other characteristics. Two-thirds of the studies (n = 18) in this review included children with ASD. Researchers in 13 of these 18 studies included children with ASD exclusively and researchers in the remaining 5 studies included children with other disabilities as well (e.g., Down syndrome, non-specified developmental delay). Although we sought to provide a comprehensive understanding of the play intervention research across disabilities, there is not evidence to indicate that the etiology or types of play deficits are the same across disability groups. Future research should be conducted to make these explicit comparisons and identify potential associations between play deficits within and across disability types and participant characteristics (e.g., language repertoire, adaptive functioning).

Robust intervention package. Robust intervention packages focusing on increasing generalized play complexity should continue to be developed and tested. Current technologies have shown limited increases in generalized play complexity in particular, which is critical for

ensuring sustained outcomes. Specific prompt types, prompt hierarchies, reinforcement contingencies, and instructional intensity should be considered in the development of effective and efficient play interventions.

Individualized adaptations. Across multiple studies individualized adaptations (e.g., reinforcer, stimulus adaptations) resulted in behavioral improvements, which supports the use of individualized instruction when teaching play. Although individualized adaptations are a recommended practice (Division for Early Childhood, 2014), these idiosyncratic findings are often overlooked given they are unlikely to allow for experimental control. Unfortunately, the disproportionate publication of studies with immediate and robust behavior change, threatens the validity of research, future replications, and the examination of studies with complex variables (Cook & Therrien, 2017; Tincani & Travers, 2018). The idiosyncratic nature of play and its measurement might adversely impact replications or result in multifarious results. Future research should focus on publishing studies based on their methodological rigor rather than the magnitude of behavior change to ensure comprehensive research is accessible to advance research and practice (Cook & Therrien, 2017).

Rigorous studies. The studies included in this review were characterized as lacking in rigor and quality, which was apparent across both the WWC and SCARF quality ratings. Further, 14 (52%) of the 27 studies were published after Horner and colleagues (2005) published quality indicators for SCR. Four of these 14 studies did not meet WWC design standards and five met design standards with or without reservations but had no evidence of a functional relation. Thus, the lack of rigor is not limited to studies published prior to the development of contemporary rigor standards. The overall lack of rigor limits interpretations of the results and highlights the need to plan and conduct SCR that meets contemporary design standards.

Measurement of generalization. Although the generalization of play skills to the classroom (e.g., free play, center times) should be a primary outcome, this type of measurement might be difficult given the likelihood of competing activities, which might make other activities and materials more interesting to the participants. Only one study measured the generalization of play in free play contexts, but they used toys that were similar in form and function to the instructional toys (Barton, 2015). Future research should develop interventions that result in sustained, generalized play behaviors with a variety of toys and materials.

Positive affect. One element of play that was rarely directly measured in this research is positive affect or behavioral indicators of happiness or enjoyment. Although descriptive definitions of play within play taxonomies often include positive affect (Barton, 2014), in the current review, Lang and colleagues (2009) and Reinhartsen and colleagues (2009) were the only authors who measured child affect during or post-intervention. Research has advanced to support play as a critical intervention goal, suggesting overt enjoyment or positive affect is not a critical element. That is, changes in affect during academic or adaptive interventions are not generally considered relevant and its role in play research is thus far understudied. Descriptive studies showing changes in affect during play are needed before positive affect should be considered a behavioral marker of play. However, future research could identify methods of measuring changes in positive affect following the acquisition of play skills (e.g., counting smiles and laughter). In doing so, both the child's and the adult's (or peer's) affect should be measured to examine changes in shared enjoyment. Further, changes in positive affect or enjoyment could be used to monitor or evaluate child preferences and capitalize on their motivation to engage in object play, which might be a more useful proxy measure of enjoyment.

Conclusions

A distinct finding from this research is that object play behaviors are amenable to change. Given the clear and replicated research highlighting the benefits of play, research and practice should continue to focus on teaching play to children who do not play in the same manner as their peers. In the aggregate, the studies included in this review demonstrated variable change in object play behaviors particularly in maintenance and generalization conditions across intervention types and participants. The variability might be related to the need for more robust interventions or the limited methodological quality. Overall, there is a strong, albeit emerging evidence base for teaching play and continued replications are warranted.

References

American Academy of Pediatrics. (2007). Position statement: Principles and guidelines for early hearing detection and intervention programs. *Pediatrics, 120,* 898–921.

*Ballard, K. D., & Medland, J. L. (1986). Collateral effects from teaching attention, imitation and toy interaction behaviors to a developmentally handicapped child. *Child & Family Behavior Therapy, 7,* 47–60.

Barton, E. E. (2010). Development of a taxonomy of pretend play for children with disabilities. *Infants and Young Children, 23,* 247–261.

*Barton, E. E., Chen, C., Pribble, L., Pomes, M., & Kim, Y.A. (2013). Coaching pre-service teachers to embed prompting procedures in inclusive preschool classrooms. *Teacher Education and Special Education, 36,* 330–349.

Barton, E. E. (2014). Assessing social competence and play skills. In M. McLean, M. L. Hemmeter, & P. Synder (Eds.). *Essential elements for assessing infants and young children with special needs* (pp. 195-241). Boston, MA: Pearson.

*Barton, E. E. (2015). Teaching generalized pretend play and related behaviors to young children with disabilities. *Exceptional Children, 81,* 489–506.

Barton, E. E., & Wolery, M. (2008). Teaching pretend play to children with disabilities: A review of the literature. *Topics in Early Child Education, 28,* 109–125.

*Barton, E. E., & Wolery, M. (2010). Training teachers to promote pretend play in young children with disabilities. *Exceptional Children, 77,* 85–106.

Belsky, J., & Most, R. (1981). From exploration to play: A cross sectional study of infant free play behavior. *Developmental Psychology, 17,* 630–639.

Bernard-Opitz, V., Ing, S., & Kong, T. Y. (2004). Comparison of behavioural and natural play interventions for young children with autism. *Autism*, *8*(3), 319-333.

*Buggey, T., & Ogle, L. (2012). Video self-modeling. *Psychology in the Schools*, *49*, 52–70.

Charman, T., & Baron-Cohen, S. (1997). Brief report: Prompted pretend play in autism. *Journal of Autism and Developmental Disorders, 27*, 325–332.

*Colozzi, G. A., Ward, L. W., & Crotty, K. E. (2008). Comparison of simultaneous prompting procedure in 1: 1 and small group instruction to teach play skills to preschool students with pervasive developmental disorder and developmental disabilities. *Education and Training in Developmental Disabilities, 43*, 226–248.

Cook, B. G., & Therrien, W. J. (2017). Null effects and publication bias in special education research. *Behavioral Disorders, 42*, 149-158.

*Craig-Unkefer, L. A., & Kaiser, A. P. (2003). Increasing peer-directed social-communication skills of children enrolled in head start. *Journal of Early Intervention, 25*, 229–247.

Division for Exceptional Children (2014). *DEC Recommended Practices*. Retrieved from http://www.dec-sped.org/recommendedpractices

*DiCarlo, C. F., Reid, D. H., & Stricklin, S. B. (2003). Increasing toy play among toddlers with multiple disabilities in an inclusive classroom: a more-to-less, child-directed intervention continuum. *Research in Developmental Disabilities*, *24*, 195–209.

*Dupere, S., MacDonald, R. P., & Ahearn, W. H. (2013). Using video modeling with substitutable loops to teach varied play to children with autism. *Journal of Applied Behavior Analysis, 46*, 662–668.

*Fox, L., & Hanline, M. F. (1993). A preliminary evaluation of learning within developmentally appropriate early childhood settings. *Topics in Early Childhood Special Education, 13,* 308–327.

Freeman, S. F., Gulsrud, A., & Kasari, C. (2015). Brief report: Linking early joint attention and play abilities to later reports of friendships for children with ASD. *Journal of Autism and Developmental Disorders*. *Advance online publication.* doi: 10.1007/s10803-015-2369-x

*Frey, J. R., & Kaiser, A. P. (2011). The use of play expansions to increase the diversity and complexity of object play in children with disabilities. *Topics in Early Childhood Special Education, 31,* 99–111. doi:10.1177/0271121410378758

*Friman, P. C., & Altman, K. (1990). Parent use of DRI on high rate disruptive behavior: Direct and collateral benefits. *Research in Developmental Disabilities, 11,* 249–254.

Gast, D. L., & Ledford, J.R. (2014). *Single case research methodology: Applications in special education and behavioral sciences*. New York, NY: Routledge.

Gast, D. L., & Spriggs, A. D. (2014). Visual analysis of graphic data. In D. L. Gast & J. R. Ledford (Eds.) *Single case research methodology: Applications in special education and behavioral sciences*. (pp. 199-233). New York, NY: Routledge.

Ginsburg, K. R., the Committee on Communications, & the Committee on Psychosocial Aspects of Child and Family Health. (2007). The importance of play in promoting healthy child development and maintaining strong parent–child bonds. *Pediatrics, 119,* 182–191.

Gulsrud, A., Hellemann, G., Freeman, S., & Kasari, C. (2014). Two to ten years: developmental trajectories of joint attention in children with ASD who received targeted social communication interventions. *Autism Research, 7,* 207–215.

Hanline, M. F., Milton, S., & Phelps, P. C. (2010). The relationship between preschool block play and reading and math abilities in early elementary school: A longitudinal study of children with and without disabilities. *Early Child Development and Care, 180,* 1005–1017.

*Hine, J. F., & Wolery, M. (2006). Using point-of-view video modeling to teach play to preschoolers with autism. *Topics in Early Childhood Special Education, 26,* 83–93.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71,* 165–179.

Joseph, J. D., Strain, P., Olszewski, A., & Goldstein, H. (2016). A consumer reports-like review of the empirical literature specific to preschool children's peer-related social skills. In B. Reichow, B. Boyd, E. E. Barton, & S. Odom (Eds). *Handbook of Early Childhood Special Education* (pp. 179-197). London: Springer.

Jung, S., & Sainato, D. (2013). Teaching play skills to young children with autism. *Journal of Intellectual and Developmental Disability, 38,* 74–90.

Kasari, C., & Chang, Y. C. (2014). Play development in children with autism spectrum disorders: Skills, object play, and interventions. In F. Volkmar, S. Rogers, R. Paul, & K.A. Pelphrey (Eds.), *Handbook of autism and pervasive developmental disorders* (4th ed., pp. 263–277). Hoboken, NJ: John Wiley & Sons.

Kasari, C., Chang, Y. C., & Patterson, S. (2013). Pretending to play or playing to pretend: The case of autism. *American Journal of Play, 6,* 124–135.

Kasari, C., Freeman, S., & Paparella, T. (2006). Joint attention and symbolic play in young children with autism: A randomized controlled intervention study. *Journal of Child Psychology and Psychiatry, 47*, 611–620.

Kasari, C., Gulsrud, A., Freeman, S., Paparella, T., & Hellemann, G. (2012). Longitudinal follow-up of children with autism receiving targeted interventions on joint attention and play. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*, 487–495.

Kazdin, A. E. (1982). *Single-case experimental designs: Strategies for studying behavior change*. New York: Oxford University Press.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M & Shadish, W. R. (2010). *Single case designs technical documentation.* In What Works Clearinghouse: Procedures and standards handbook (version 2.0). Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_procedures_v2_standards_handbook.pdf

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*, 26–38.

*Lang, R., Machalicek, W., Rispoli, M., O'Reilly, M., Sigafoos, J., Lancioni, G., ... & Didden, R. (2014). Play skills taught via behavioral intervention generalize, maintain, and persist in the absence of socially mediated reinforcement in children with autism. *Research in Autism Spectrum Disorders, 8*, 860–872.

Lang, R., O'Reilly, M., Shorogren, K., Machalicek, W., Sigafoos, J., & Regester, A. (2009). Review of intervention to increase functional and symbolic play in children with autism. *Education and Training in Developmental Disabilities, 44*, 481–492.

Ledford, J. R., Lane, J. D., Zimmerman, K. N., Chazin, K. T., & Ayres, K. A. (2016, April). *Single case analysis and review framework* (SCARF). Retrieved from: http://vkc.mc.vanderbilt.edu/ebip/scarf/

Lewis, V. (2003). Play and language in children with autism. *Autism, 7,* 391–399.

Lifter, K. (2000). Linking assessment to intervention for children with developmental disabilities or at-risk for developmental delay: The developmental play assessment (DPA) instrument. In K. Gitlin-Weiner, A. Sandgrund, & C. Schafer (Eds.), *Play diagnosis and assessment* (2nd ed., pp. 228-261). New York: John Wiley and Sons.

Lifter, K., Ellis, J., Cannon, B., & Anderson, S. R. (2005). Developmental specificity in targeting and teaching play activities to children with pervasive developmental disorders. *Journal of Early Intervention, 27,* 247–267.

Lifter, K., Foster–Sanda, S., Arzamarski, C. A., Briesch, J., & McClure, E. (2011a). Overview of play: Its uses and importance in early intervention/early childhood special education. *Infants & Young Children, 24,* 225–245.

Lifter, K., Mason, E. J., & Barton, E. E. (2011b). Children's play: Where we have been and where we could go. *Journal of Early Intervention, 33,* 281–297.

Lifter, K., Sulzer-Azaroff, B., Anderson, S., & Cowdery, G. E. (1993). Teaching play activities to preschool children with disabilities: The importance of developmental considerations. *Journal of Early Intervention, 17,* 139–159.

MacDonald, R., Clark, M., Garrigan, E., & Vangala, M. (2005). Using video modeling to teach pretend play to children with autism. *Behavioral Interventions, 20,* 225–238.

Machalicek, W., Shogren, K., Lang, R., Rispoli, M., O'Reilly, M. F., Franco, J. H., & Sigafoos, J. (2009). Increasing play and decreasing the challenging behavior of children with

autism during recess with activity schedules and task correspondence training. *Research in Autism Spectrum Disorders, 3,* 547–555.

*MacManus, C., MacDonald, R., & Ahearn, W. H. (2015). Teaching and generalizing pretend play in children with autism using video modeling and matrix training. *Behavioral Interventions, 30,* 191–218.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine, 6*(7), e1000097.

*Murdock, L. C., Ganz, J., & Crittendon, J. (2013). Use of an iPad play story to increase play dialogue of preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43,* 2174–2189.

National Association for the Education of Young Children. (2009). Developmentally appropriate practice in early childhood programs serving children from birth through age 8: A position statement of the National Association for the Education of Young children. In S. Bredekamp & C. Copple (Eds.), *Developmentally appropriate practice in early childhood programs* (pp. 1–32). Washington, DC: Author.

*Nuzzolo-Gomez, R., Leonard, M. A., Ortiz, E., Rivera, C. M., & Greer, R. D. (2002). Teaching children with autism to prefer books or toys over stereotypy or passivity. *Journal of Positive Behavior Interventions, 4*, 80–87.

Odom, S. L., Barton, E. E., & Reichow, B. R., Pustejovsky, J. P., & Swaminthan, R. (2018). Between-case standardized effect size analysis of single case design: Examination of the two methods. *Research in Developmental Disabilities, 79,* 88-96.

*Palechka, G., & MacDonald, R. (2010). A comparison of the acquisition of play skills using instructor-created video models and commercially available videos. *Education and Treatment of Children, 33,* 457–474.

Parten, M. (1932). Social participation among preschool children. *Journal of Abnormal and Social Psychology, 27,* 243–269.

Parker, R. I., Cryer, J., & Byrns, G. (2006). Controlling baseline trend in single-case research. *School Psychology Quarterly, 21,* 418–444.

Piaget, J. (1952). *The child's conception of number.* New York: Norton.

Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics, 39,* 368–393.

*Reinhartsen, D. B., Garfinkle, A. N., & Wolery, M. (2002). Engagement with toys in two-year-old children with autism: Teacher selection versus child choice. *Research and Practice for Persons with Severe Disabilities, 27,* 175–187.

Schardt, C., Adams, M. B., Owens, T., Keitz, S., & Fontelo, P. (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Medical Informatics and Decision Making, 7,* 16–22.

*Scheflen, S. C., Freeman, S. F., & Paparella, T. (2012). Using video modeling to teach young children with autism developmentally appropriate play and connected speech. *Education and Training in Autism and Developmental Disabilities, 47,* 302–318.

Sherratt, D., & Peter, M. (2002). *Developing play and drama in children with autistic spectrum disorders*. New York, NY: Routledge.

Shadish, W. R., Hedges, L. V., Horner, R. H., & Odom, S. L. (2015). The Role of Between-Case Effect Size in Conducting, Interpreting, and Summarizing Single-Case Research. NCER 2015-002. *National Center for Education Research*.

Sherratt, D., & Peter, M. (2002). *Developing play and drama in children with autistic spectrum disorders.* New York, NY: Routledge.

Sigman, M., & Ruskin, E. (1999). Continuity and change in the social competence of children with autism, Down Syndrome, and developmental delays. *Monographs of the Society for Research in Child Development, 64,* 1–114.

Smilansky, S. (1968). *The effects of sociodramatic play on disadvantaged children: preschool children.* New York, NY: Wiley.

*Stahmer, A. C. (1995). Teaching symbolic play skills to children with autism using pivotal response training. *Journal of Autism and Developmental Disorders*, *25*, 123–141.

Stahmer, A. C., Ingersoll, B., & Carter, C. (2003). Behavioral approaches to promoting play. *Autism, 7*, 401–413.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, *10*, 349–367.

Stokes, T. F., & Osnes, P. G. (1989). An operant pursuit of generalization. *Behavior Therapy*, *20*, 337-355.

*Stanton-Chapman, T. L., & Brown, T. S. (2015). A strategy to increase the social interactions of 3-year-old children with disabilities in an inclusive classroom. *Topics in Early Childhood Special Education*, *35*, 4–14.

Strain, P. S., Schwartz, I. S., & Bovey, E. H. (2008). Social competence interventions for young children with autism. In W. H. Brown, S. L. Odom, & S. R. McConnell (Eds.), *Social*

competence of young children: Risk, disability, and intervention (pp. 253–272). Baltimore: Paul H. Brookes.

Swaminathan, H., Rogers, H. J., Horner, R. H., Sugai, G., & Smolkowski, K. (2014). Regression models and effect size measures for single case designs. *Neuropsychological Rehabilitation, 24,* 554-571

*Taylor, R., & Iacono, T. (2003). AAC and scripting activities to facilitate communication and play. *Advances in Speech Language Pathology, 5*, 79–93.

Thiemann–Bourque, K. S., Brady, N. C., & Fleming, K. K. (2012). Symbolic play of preschoolers with severe communication impairments with autism and other developmental delays: More similarities than differences. *Journal of Autism and Developmental Disorders, 42,* 863–873.

*Thorp, D. M., Stahmer, A. C., & Schreibman, L. (1995). Effects of sociodramatic play training on children with autism. *Journal of Autism and Developmental Disorders*, *25,* 265–282.

Tincani, M., & Travers, J. (2018). Publishing single-case research design studies that do not demonstrate experimental control. *Remedial and Special Education, 39,* 118-128.

Toth, K., Munson, J., Meltzoff, A., & Dawson, G. (2006). Early predictors of communication development in young children with autism spectrum disorder: Joint attention, imitation, and toy play. *Journal of Autism Developmental Disorders, 36,* 993–1005.

Ungerer, J. A., & Sigman, M. (1984). The relation of play and sensorimotor behavior to language in the second year. *Child Development*, *55,* 1448-1455.

U.S. Department of Health and Human Services and U.S. Department of Education. (2015). Policy statement on the inclusion of children with disabilities in early childhood

programs. Retrieved from: http://www2.ed.gov/about/inits/ed/earlylearning/inclusion/index.html

Van Berckelaer-Onnes, I. A. (2003). Promoting early play. *Autism, 7*, 415–423.

*VanDerHeyden, A. M., Snyder, P., DiCarlo, C. F., Stricklin, S. B., & Vagianos, L. A. (2002). Comparison of within-stimulus and extra-stimulus prompts to increase targeted play behaviors in an inclusive early intervention program. *The Behavior Analyst Today, 3*, 189–198.

Vig, S. (2007). Young children's object play: A window on development. *Journal of Developmental and Physical Disabilities, 19,* 201–215.

What Works Clearinghouse. (2017). Procedures and Standards Handbook (v. 3). Washington, D.C.: Institute for Education Sciences, U.S. Department of Education.

Williams, E., Reddy, V., & Costall, A. (2001). Taking a closer look at functional play in children with autism. *Journal of Autism and Developmental Disorders, 31*, 67-77.

Yu, S., Ostrosky, M. M., & Fowler, S. A. (2015). The relationship between preschoolers' attitudes and play behaviors toward classmates with disabilities. *Topics in Early Childhood Special Education, 35*, 40–51.

Table 1.

Baseline and Instructional Procedures and Reported Participant Disability Descriptions

Reference	Baseline Procedure(s)	Instructional Procedures				Participant Disabilities
		Procedure Type	Prompt Type	Reinforcement	Reinforcement Type	
Ballard (1985)	RSI	AP, R	PM, HOH	Differential	Edible, Praise	Myoclonic epilepsy
Barton (2015)	*R, RSI, C, CI*	*SLP, R*	*PM, HOH*	--	*Edible, Praise, Physical*	*ASD, DS*
Barton (2010)	*P, RSI*	*SLP, R, CI*	*PM, HOH, CP*	*Differential*	*Praise, Physical*	*ASD, DD,* C/LD
Barton (2013)	*RSI*	*SLP, R*	*PM, HOH*	--	*Praise*	*ASD, DD,* C/LD
Buggey (2012)	*RSI*	*VM*	--	--	--	*ASD*
Colozzi (2008)	*R, RSI*	*SIM, R*	*HOH*	*Multiple*	*Edible, Praise*	*ASD, DS, Severe ID Hirschsprung disease*
Craig-Unkefer (2002)	RSI	R, PMI	--	Scheduled	Praise	At-risk for language delay or problem behaviors
DiCarlo (2003)	P, RSI	AP, R	HOH, CP	Differential	Praise	DD
Dupere (2013)	*RSI*	*VM*	--	--	--	*ASD*
Fox (1993)-1	P, RSI	SLP, R	PM, HOH	Differential	Praise	DS
Fox (1993)- 2	P, RSI	SLP, R	PM, HOH	Differential	Praise	DD
Frey (2011)	R, RSI, C, CI	R, PE, CI	PM	Scheduled	Praise	DS, DD, C/LD
Friman (1990)	RSI, C	AP, R	--	Differential	Edible, Praise	Severe ID
Hine (2006)	*RSI*	*R, VM*	--	--	*Edible, Praise*	*ASD*
Lang (2014)	--	*SLP, R*	*PM, HOH*	*Multiple*	*Edible, Praise*	*ASD*
MacManus (2015)	--	*VM*	--	--	--	*ASD*
Murdock (2013)	*RSI*	*SC*	*VP*	--	--	*ASD*
Nuzzolo-Gomez (2002)-1	*RSI*	*SLP, R*	*PM, HOH*	*Multiple*	*Edible, Praise*	*ASD*
Nuzzolo-Gomez (2002)-2	*RSI*	*SLP, R*	*PM, HOH*	*Multiple*	*Edible, Praise*	*ASD*
Palechka (2010)	--	*VM*	--	--	--	*ASD*

Reinhartsen (2002)	--	*SLP*	*PM, HOH, CP*	--	--	*ASD*
Scheflen (2012)	*RSI*	*R, VM*	--	*Differential*	*Praise*	*ASD*
Stahmer (1995)	*P, RSI*	*R, PMI, PRT*	*PM*	*Differential*	*Praise, Toys*	*ASD*
Stanton-Chapman (2015)	RSI	R, PMI	VP	Differential	Praise	DD, C/LD
Taylor (2003)	RSI, C	AP, SC	PM	--	--	Mild ID, C/LD
Thorp (1995)	*P, RSI*	*R, PRT*	*PM*	*Differential*	*Praise*	*ASD,*
VanDerHyden (2002)	*P, RSI*	*SLP, R*	*PM, HOH*	--	*Praise, Physical*	*ASD, DS*

Note. Italics identify studies that included participating children with ASD. AP = adult prompting not otherwise specified; VM = video model; SIM = simultaneous prompting; R = reinforcement; PRT = Pivotal Response Training; PMI = peer-mediated instruction; CI = contingent imitation; SLP = system of least prompts; P = prompt; PE = play expansions; SC = play scripts; RSI = responding to appropriate social interactions; C = implementer commentary; PM = physical model; CP = choice; HOH = hand over hand, full physical; VP = visual prompt; ASD = autism spectrum disorder; DD = developmental delay; DS = Down syndrome; C/LD = communication/language delays; ID = intellectual delay

Table 2.
Measurement and Materials

Reference	Meas. System[a]	Materials[b]	Play Types[c]	Play-Related Outcomes[c]	Generalization Measurement[d]
Ballard (1985)	ER	D, V, J	*FPP*, SM, FP	UP, VR	S
Barton (2015) *	TER	HK, D, B, J	*FPP, IAO, OS, AAA, SPP*	UP, VR, SP, DP	P, T, S, O
Barton (2010) *	TER	HK, D, V, J, B, O, J	*FPP, OS, IAO, AAA*	UP, VR, SP, DP	T, I
Barton (2013) *	TER	HK, D, B	*FPP, AAA, IAO, OS*	UP	P, I, O
Buggey (2012) *	ER	TS, C	--	VR	P
Colozzi (2008) *	ER	HK, D, V	*FPP*	--	P, T, S, I
Craig-Unkefer (2002)	PIR	HK, V, B, TS	*FPP, SPP, AAA*	VR	P
DiCarlo (2003)	PIR	V, B	SM, FP	UP	--
Dupere (2013)*	ER	V, TS, O	*FPP, AAA*	VR	T, S, I
Fox (1993)-1	ER	D, V, B, O	FP	UP	P
Fox (1993)- 2	PC	D, V, B, O	--	UP, VR	P, S, I
Frey (2011)	PIR	HK, D, V, TS, O	*FPP*, SM	UP, VR, DP, SP	T
Friman (1990)	PIR	O	RP	--	--
Hine (2006) *	ER	HK, TS, O	*FPP, OS*	DP, I	P, T, S
Lang (2014) *	PIR	V, TS, O	*FPP*	UP, SP	P, T, S
MacManus (2015) *	PC	V, TS, O	*FPP, AAA*, FP	VR, SP	O
Murdock (2013) *	ER	V, TS, O	--	VR	P, I
Nuzzolo-Gomez (2002)-1 *	PIR	O	--	UP	P, O
Nuzzolo-Gomez (2002)-2 *	PIR	D, V, B, O	*FPP*	--	P, O
Palechka (2010) *	ER/PC	V, TS, O	*FPP, AAA*	VR	--
Reinhartsen (2002) *	MTS	D, V, B	*FPP*, FP, RP	--	P
Scheflen (2012) *	PL	HK, D, V, B, TS, O, C	*FPP, OS, IAO, AAA*, SM, FP	VR, SP	P, S
Stahmer (1995) *	ER	HK, D, V, J, TS, O	*OS, IAO, AAA, SPP*	SP	P, T, S, I
Stanton-Chapman (2015)	ER	D, V, TS	*FPP, AAA, SPP*, FP, RP	VR	P
Taylor (2003)	ER	V, B, TS, O	*FPP, OS, IAO, AAA*, FP	UP, VR	T
Thorp (1995) *	PIR	HK, D, J, TS	*OS, IAO, AAA, SPP*	VR, SP	T, S
VanDerHyden (2002) *	PIR	HK, C	FP	VR	P

Note. * Asterisks identify studies that included participating children with ASD.
[a]Measurement system: PTS = Partial Interval Recording; MTS = Momentary Time Sampling; TER = Timed Event Recording; ER = Event Recording; PC = Percentage Correct; PL = Play Level
[b]Materials: HK = Housekeeping Toys; D = Dolls; V = Vehicles; J = Junk Toys; B = Blocks; O = Other Toys; TS = Themed Toy Sets; C = Computers
[c]Play types and outcomes: SM = Sensory-Motor Play; RP = Relational Play; FP = Functional Play; FPP = Functional Play with Pretense; SP = Symbolic Play; OS = Object Substitution; IAO = Imagining Absent Objects; AAA = Assigning Absent Attributes; SPP = Social Pretend Play; UP = Unprompted Play; VR = Vocalization Related to Play; SP = Sequences of Play; DP = Diversity of Play; I = Imitation
[d]Generalization: P=peers, T=toys, S=setting, I=interventionist, O=Other

Table 3.
Study Results

Reference	VA	WWC		SCARF						Pustejovsky Effect Size (CI)[f]	Swaminathan Effect Size (CI)[f]
				Primary Outcomes		Gener.		Maint.			
	FR /Opps[a]	Design Standards	Evidence Criteria	Q/R	Out.[b]	Q/R	Out.[c]	Q/R	Out.[d]		
Ballard (1985)	1/1	No	-	0.9	3.0	0.0	0.0	0.0	0		
Barton (2010) *	3/4	WR	Strong*	1.9	4.0	3.0	2.0	1.0	4		
			Moderate	2.1	4.0	3.0	4.0	1.0	4		
			Strong	2.1	3.0	3.0	2.0	1.0	4		
			No evidence	2.0	4.0	3.0	2.0	0.0	0		
Barton (2013) * [e]	1/1	WR	Strong*	2.4	4.0	4.0	2.0	1.0	4	2.86 (1.54, 4.18)	1.385 (0.33, 2.44)
Barton (2015) *	4/4	WR	Strong*	2.4	3.0	4.0	1.0	1.0	4		
			Strong*	2.4	4.0	4.0	2.0	1.0	4		
			Moderate*	1.9	1.0	4.0	2.0	1.0	2		
			Moderate*	1.7	3.0	0.0	0.0	0.0	0		
Buggey (2012) *	0/1	WR	No evidence	1.6	0.0	0.0	0.0	1.0	4		
Colozzi (2008) *	0/4	No	-	1.6	4.0	2.0	2.0	1.0	3		
			-	1.6	4.0	2.0	2.0	1.0	3		
			-	1.6	0.0	2.0	2.0	1.0	3		
			-	1.6	0.0	2.0	3.0	1.0	3		
Craig-Unkefer (2002)	0/1	Meets	No evidence	1.7	2.0	0.0	0.0	0.0	0		
DiCarlo (2003) [e]	0/1	Meets	Strong	1.4	0.0	0.0	0.0	0.0	0	0.51 (-0.84, 1.86)	0.590 (0.23, 0.96)
Dupere (2013) *	1/3	Meets	Strong*	0.9	4.0	2.0	2.0	1.0	2		
			No evidence	1.6	3.0	2.0	2.0	1.0	2		
			No evidence	1.4	1.0	2.0	3.0	1.0	4		
Fox (1993)-1	1/1	No	-	2.6	4.0	0.0	0.0	1.0	4		
Fox (1993)- 2	0/1	WR	Moderate	1.6	4.0	3.0	4.0	1.0	4		
Frey (2011) [e]	0/1	WR	No evidence	2.0	1.0	3.0	1.0	4.0	1	1.13 (0.04, 2.22)	0.51 (-0.84, 1.86)
Friman (1990)	1/1	WR	Moderate*	1.1	3.0	0.0	0.0	0.0	0		
Hine (2006) *	0/1	WR	No evidence	0.0	0.0	0.0	0.0	0.0	0		
Lang (2014) * [e]	1/1	Meets	Strong*	1.5	1.0	0.0	0.0	0.0	0	5.72 (2.93, 8.51)	1.714 (1.13, 2.10)
McManus (2015) * [e]	1/1	No	-	2.1	3.0	2.0	2.0	1.0	4	5.02 (-0.25, 10.3)	0.839 (0.36, 1.45)
Murdock (2013) *	0/1	WR	No evidence	2.0	4.0	2.0	3.0	4.0	3		
Nuzzolo-Gomez (2002)-1 *	1/1	No	-	0.9	4.0	3.0	2.0	1.0	4		

Nuzzolo-Gomez (2002)-2 *[e]	0/1	Meets	No evidence	1.2	2.0	1.0	2.0	3.0	4	1.88 (1.07, 2.69)	1.446 (.0.24, 2.57)
Palechka (2010) *	0/4	No	-	1.0	1.0	0.0	0.0	1.0	4		
			-	1.0	0.0	0.0	0.0	1.0	4		
			-	1.4	4.0	0.0	0.0	1.0	4		
			-	1.9	4.0	0.0	0.0	1.0	4		
Reinhartsen (2002) *	3/3	Meets	Strong*	2.1	4.0	0.0	0.0	0.0	0		
			Strong*	2.1	3.0	0.0	0.0	0.0	0		
			Strong*	2.1	4.0	0.0	0.0	0.0	0		
Scheflen (2012) * [e]	0/1	No	-	1.2	0.0	1.0	0.0	1.0	0	1.35 (-1.47, 4.18)	-0.068 (-.60, 0.40)
Stahmer (1995) * [e]	1/1	No	-	1.1	3.0	3.0	4.0	1.0	2	1.61 (-1.51, 4.73)	1.439 (0.46, 2.59)
Stanton-Chapman (2015) [e]	1/1	Meets	Strong*	2.9	3.0	0.0	0.0	0.0	0	5.00 (2.35, 7.64)	0.966 (0.54, 1.50)
Taylor (2003)	0/2	WR	No evidence	1.0	0.0	0.0	0.0	0.0	0		
			No evidence	1.0	3.0	0.0	0.0	0.0	0		
Thorp (1995) *	2/3	WR	Strong*	1.3	0.0	3.0	3.0	4.0	2		
			Moderate	1.3	0.0	3.0	2.0	4.0	2		
			No evidence	1.3	0.0	3.0	3.0	4.0	2		
VanDerHyden (2002) *	1/2	No	-	1.9	4.0	0.0	0.0	0.0	0		
			-	1.1	0.0	1.0	0.0	0.0	0		

Note. * Asterisks identify studies that included participating children with ASD. WR = With reservations

[a] Functional relations/opportunities

[b] Graphed on Figure 2

[c] Generalization outcomes (graphed on Figure 3)

[d] Maintenance outcomes (graphed on Figure 4)

[e] Studies that met criteria to be included in the meta-analysis.

[f] Effect sizes were originally published in Odom et al. (2018).

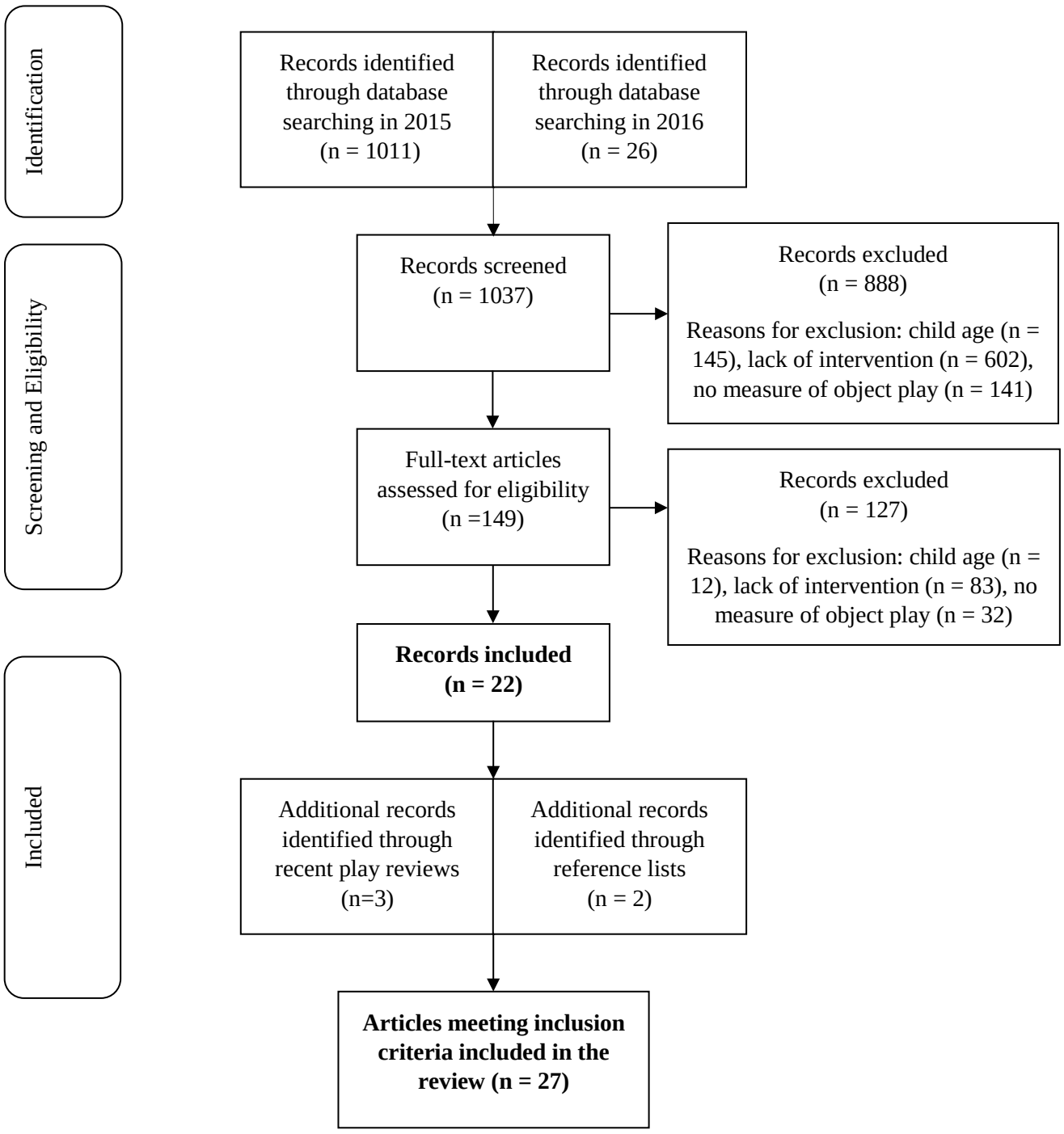

Figure 1. Article inclusion decision tree (adapted PRISMA flow diagram)

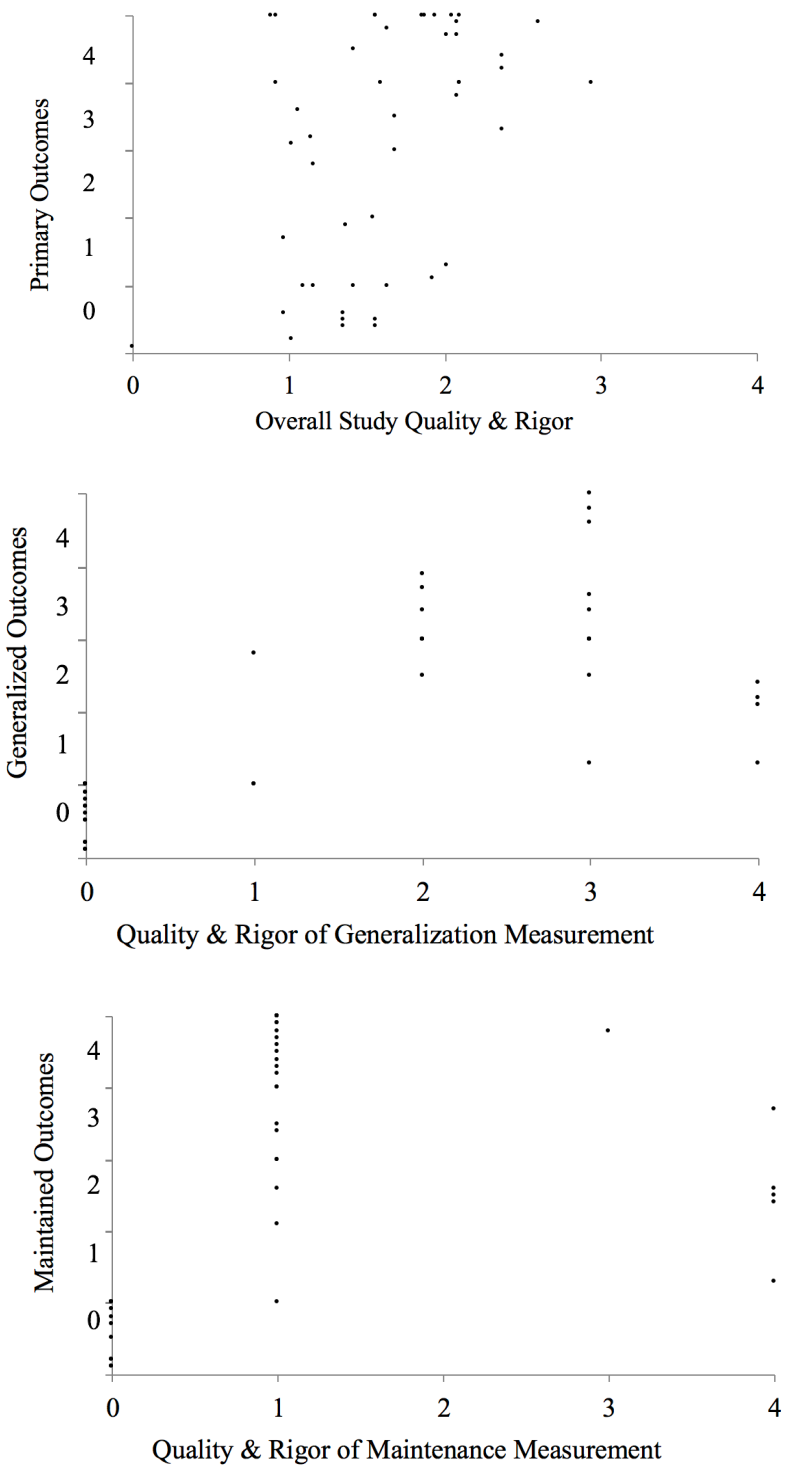

Figure 2. The top panel shows the primary outcomes using the SCARF framework. The middle panel shows the generalized outcomes using the SCARF framework. The lower panel shows the maintained outcomes using the SCARF framework.