

Antecedent Social Skills Interventions for Individuals With ASD: What Works, for Whom, and Under What Conditions?

Focus on Autism and Other
Developmental Disabilities
2018, Vol. 33(1) 3–13
© Hammill Institute on Disabilities 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1088357616634024
focus.sagepub.com



Jennifer R. Ledford, PhD¹, Seth King, PhD², Emilee R. Harbin, BS¹,
and Kathleen N. Zimmerman, MA¹

Abstract

Social skills interventions designed to increase pro-social interactions for individuals with autism spectrum disorders are critical, but the relative effectiveness of these interventions is not well understood. More than 250 single-case design studies in 113 articles were reviewed and described in terms of participants, settings, arrangements, implementers, social partners, target behaviors, and treatment components. Differential success rates are reported, given the variation in study and participant characteristics (e.g., implementers, treatment components, participant age). Environmental arrangement, social skills training, and prompting were highly successful, and peer training, priming, and video-based interventions were less successful. More evidence is needed, particularly research including older individuals and utilizing indigenous implementers and typical social partners.

Keywords

social skills interventions, evidence-based practices, single-case synthesis

Identified from the earliest reports of the disorder (Kanner, 1943) and maintained in primary diagnostic guides (e.g., *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. [DSM-5]; American Psychological Association, 2013), social deficits have long been one of the primary diagnostic indicators for children with autism spectrum disorder (ASD). Children with ASD frequently exhibit difficulties in communicating, maintaining rapport with peers, and general engagement in social situations (e.g., initiation, reciprocity; Bellini, Peters, Benner, & Hopf, 2007). Moreover, social deficits are unlikely to be ameliorated without targeted intervention, even for individuals in high-quality educational environments and without cognitive impairments (e.g., Gresham, Sugai, & Horner, 2001). Thus, identifying effective interventions to increase the pro-social behaviors of individuals with ASD should be a primary concern for researchers and practitioners.

The recent emphasis on identifying evidence-based practices (EBPs) in psychology and education (e.g., Odom, Collet-Klingenberg, Rogers, & Hatton, 2010) is particularly important for individuals with ASD, given the proliferation of, and support for, interventions unsubstantiated by sound research (Goin-Kochel, Myers, & Mackintosh, 2007). In some cases, popular use of interventions continues even after evidence of *ineffectiveness* is published (e.g., weighted vests; Reichow, Barton, Sewell, Good, & Wolery, 2010).

Research suggests that high-quality intervention grounded in EBPs is likely to result in positive outcomes for young children with ASD (e.g., Wong et al., 2014). Nonetheless, many EBPs have failed to change the behavior of one or more participants in research studies (e.g., Stewart, Benner, Martella, & Marchand-Martella, 2007). Practitioners may need more information about which EBPs are likely to be effective for changing a specific target behavior, given the heterogeneity of the population and variation in intervention contexts.

One approach to stemming the proliferation of fad interventions and treating a heterogeneous population is to teach practitioners to use scientific evidence in implementation decisions (McDonald, Pace, Blue, & Schwartz, 2012). Odom, Hume, Boyd, and Stabel (2012) suggested an individualized approach to selecting EBPs is a desirable practice; Strain, Schwartz, and Barton (2011) asserted that the matching of intervention strategies to target behaviors is essential for intervention providers. Research reviews that

¹Vanderbilt University, Nashville, TN, USA

²Tennessee Technological University, Nashville, USA

Corresponding Author:

Jennifer R. Ledford, Peabody College, Vanderbilt University, Peabody
Box 228, MRL/OMC 416E, Nashville, TN 37203, USA.
Email: Jennifer.ledford@vanderbilt.edu

specifically address the effectiveness of interventions, along with information about the *participants*, *contexts*, and *target behaviors* for which the interventions have been shown effective are needed to facilitate this process.

Despite their apparent value to the field, common threats to internal validity may compromise conclusions from research syntheses. One such threat, the “apples and oranges” problem, refers to the faulty conclusions that arise from reconciling disparate dependent variables, subject populations, and interventions (Sharpe, 1997). The inclusion of studies with methodological issues may also distort conclusions regarding treatment efficacy. Single-case designs with insufficient replications of effect provide inaccurate depictions of treatment efficacy. Likewise, statistical methods used in the analysis of group designs fundamentally differ from analysis in single-case design, which emphasizes the visual determination of consistent and replicated effects, and these methods have the potential to skew data interpretations (e.g., show a large effect even if three replications of effect did not exist; Riley-Tillman & Burns, 2010). The propriety of analyzing common data patterns through the quantification of single-case intervention effects (e.g., percent of nonoverlapping data) remains unclear (e.g., Wolery, Busick, Reichow, & Barton, 2010). More research is needed to evaluate the concurrence of synthesized effect size metrics and visual analysis; these metrics should only be used for studies in which a functional relation may be determined (e.g., three potential demonstrations of effect, few threats to validity). Recent reviews of social skills interventions for individuals with ASD have emphasized the importance of methodologically sound research and used a variety of approaches to analyzing study results. In a best-evidence synthesis of rigorous single-case and group design studies, Reichow and Volkmar (2010) evaluated intervention effects over a relatively short time frame (2001–2008). However, single-case studies with fewer than three replications of experimental effect remained eligible for inclusion. The authors cited evidence for support for all commonly used intervention types (e.g., video modeling) with an average success rate of more than 90% across types. However, for all intervention types and for each age group and “cognitive functioning level,” at least one failure was noted. Goldstein, Lackey, and Schneider (2014) evaluated the methodology and effect sizes of social skill interventions for preschoolers with ASD. Effect sizes for single-case designs were derived from the percentage of nonoverlapping data for studies over 30 years. Few studies identified by Goldstein and colleagues reported evidence of non- or limited effectiveness. Using hierarchical linear modeling, Wang, Parilla, and Cui (2013) calculated effect sizes for 115 single-case studies published between 1994 and 2012. Like Reichow and Volkmar, Wang et al. found that most studies (90%)

successfully changed behavior and found no variance in effect based on participant age.

Difficulties with interpreting results from these and previous reviews of social skills interventions for individuals with ASD (e.g., McConnell, 2002) stem from (a) the inclusion of single-subject studies with fewer than three replications of effect or other issues that preclude the analysis of a functional relation, (b) attempts to quantify findings from single-case research via statistical analysis alone, and (c) the synthesis of studies with different treatments and dependent measures. In addition to conflating disparate interventions and measures, syntheses of social skills interventions that include studies concerning skills with only an indirect association to social interaction (e.g., imitation; Wang et al., 2013) potentially overstate the general effectiveness of treatment. Moreover, further insight into the range of factors relevant to treatment efficacy (e.g., measurement context) is needed. Wang et al. (2013) noted, “On average, SSIs (social skills interventions) are effective” (p. 1709); nonetheless, important questions remain: *What interventions are effective, for whom, and under what conditions?*

The purpose of this review is to evaluate the characteristics and effectiveness of antecedent social skills interventions that provided sufficient evidence to demonstrate a functional relation. For the purposes of this review, we focus on the antecedent treatment components although consequence components were also coded, and no studies were excluded for failure to include an antecedent component. Questions include the following: (a) How do researchers assess the efficacy of interventions and to what extent do researchers report information about interobserver agreement (IOA), fidelity, and social validity? (b) What are the characteristics of participants? (c) What are the characteristics and components of treatment? (d) How effective are different treatment components? (e) What is the relationship between participant characteristics, treatment components, and outcomes? and (f) How do results obtained through visual analysis compare with results obtained through a nonoverlap measure of effect? We address these questions by analyzing the effects of social interventions for individuals with ASD for approximately the past 20 years; furthermore, we analyze participants, settings, interventionists and social partners, target behaviors, and treatment components in relation to study effects.

Method

Inclusion Criteria

Studies selected for the initial review (a) appeared in English-language, peer-reviewed journals between 1994 and 2014, (b) featured an intervention designed to increase positive or pro-social interactions (i.e., directed at individuals) in students with ASD, and (c) visually presented single-case

designs depicting at least three potential replications of effect. Pro-social behaviors included verbal and non-verbal interactions with others considered by the authors of the original report to be desirable (e.g., sharing, commenting, responding to peer requests). Group comparison designs were excluded due to the lack of an accepted method for comparing results across design types and the prevalence of use of single-case design research among children with ASD (Gast & Ledford, 2014). Also excluded were behaviors designed to manage the environment with no social purpose or variation in responding (e.g., requesting a break by giving a card), behaviors related to social interactions measured outside the context of those interactions (e.g., learning to name emotions), and simple verbal or motor imitation.

A secondary full-text review excluded single-case studies in which methodological issues inhibited the determination of a functional relationship (e.g., fewer than three data points per condition, nonconcurrent multiple baseline design). Studies with data not indicative of functional relations *due to methodological problems* (e.g., behavioral covariation in a multiple baseline design), rather than intervention failures, were excluded in accordance with the single-case design standards of the What Works Clearinghouse (WWC; 2014). Included studies featured a minimum of three data points per phase and initiated treatment following a stable or countertherapeutic baseline trend. In addition, studies without evidence of dependent variable reliability data collection (e.g., IOA data) for at least 20% of study sessions ($n = 7$) were excluded.

Search Procedures

Search procedures consisted of an ancestral examination of literature reviews and an electronic database search. The ancestral search yielded 223 peer-reviewed articles included in previous reviews of social skills interventions (e.g., Reichow & Volkmar, 2010; Wang et al., 2013). PsycINFO and ERIC (Education Resources Information Center) database searches identified 1,275 articles with the following terms in the abstract: *autis** and *social* and with one or more of the following terms: *treatment* or *training* or *intervention*. Of these, 1,149 were not included in other reviews. The database search identified 216 articles meeting initial screening criteria. Thus, 439 articles were retained for full-text review (223 from previous reviews and 216 new articles). Of these, 148 met inclusion criteria. Following the exclusion of studies with methodological issues ($n = 35$), a total of 113 articles with 263 studies were selected for review. The identified articles were originally published in 32 journals, with *Journal of Autism and Developmental Disabilities*, *Journal of Applied Behavior Analysis*, *Journal of Positive Behavior Interventions*, and *Focus on Autism and Other Developmental Disabilities* featuring 10 or more articles. To ensure the adequacy of the search procedures, a

hand search was conducted for recent publications (2009–2014) in the five journals with the highest numbers of included articles. One additional article was located, for a total of 113 articles with 263 studies.

Coding

A team of reviewers comprised of doctoral-level behavioral analysts with experience in single-case design research (first and second authors) assessed the studies across a range of attributes. Codes pertained to study methodology as well as participant, treatment, and outcome characteristics. The extent to which studies demonstrated a functional relation between treatment and outcome variables was also assessed. A second team (first, third, and fourth authors) extracted data from included studies and calculated Tau-*U* values (Parker, Vannest, & Davis, 2011). A detailed description of study codes appears in Table 1.

Methodology. Methodology codes pertained to the design and general methodological features of the identified studies. Combination designs were coded according to the design that allowed for the most demonstrations of effect. IOA, procedural fidelity, and social validity were assessed in accordance with contemporary recommendations (e.g., Ledford & Wolery, 2013; WWC, 2014). Informal reports of social validity (e.g., anecdotal reports) were excluded.

Participant characteristics. Researchers recorded participant characteristics, including age, gender, school placement, cognitive ability, and author-reported diagnosis (e.g., autism) for each participant. Participants were categorized as having an intellectual disability (ID) based on nominal reports or through the presentation of IQ scores. For participants 8 years of age or younger, if authors did not report an IQ score or ID determination, but reported cognitive assessment scores as an estimate of intellectual ability, this information was used to classify participants.

Contextual variables. The review team assessed contextual aspects of interventions, including the measurement context during which primary data were collected and the instructional arrangement for intervention delivery. Identical codes (e.g., peers, parents, teachers) were used to identify individuals who administered interventions (implementers) and engaged in social interaction with participants (social partners).

Treatment components. Treatment component codes corresponded to the independent variables featured in identified studies. For each study, any number of treatment components could be coded. A full list of definitions can be obtained from the first author. Scripts and written priming interventions—characterized by written or pictorial cues for

Table 1. Coding Categories and Descriptions.

Category	Sub-categories	Description
Methodology	Design	Reported single-case designs including <i>A-B-A-B</i> ; <i>multiple baseline/probe across participants, behaviors, social partners, or contexts</i> ; <i>alternating treatments</i> ; <i>multi-, parallel, and simultaneous treatments</i> ; and <i>other</i> .
	IOA	Separate indicators of whether study (a) reported IOA and (b) met standards of measurement (i.e., collected data in 20% of sessions in each condition, agreement of at least 80%).
	Fidelity	Reported procedural fidelity in each experimental condition.
	Social validity	Reported social validity data through <i>surveys, normative comparisons, or blind coding of videos</i> .
Participants	Demographics	Gender and age (rounded down to the nearest year) of participants.
	Diagnosis	Diagnoses of participants, including <i>autism, Asperger, high-functioning autism, and ASD</i> .
	Cognitive disabilities	Indicated whether authors reported ID among participants either nominally or through the presentation of IQ scores. For participants 8 years of age or younger, ID was inferred from a mental-chronological age deficit exceeding 12 months.
	School placement	Reported school placement of participants classified as <i>general</i> (i.e., $\geq 50\%$ in a general education setting) or <i>special education</i> (i.e., $\leq 50\%$ in a general education setting).
Contextual variables	Measurement context	Data collection venues, including <i>home, clinic, and teacher- or child-directed school activities</i> .
	Arrangement	Instructional groupings including <i>1:1</i> (implementer and child), <i>small group</i> (2–5 children), <i>large group</i> (6 or more children), or <i>undefined</i> (number of children unconstrained or not reported).
	Implementer/social partner	Individuals responsible for (a) delivering the intervention and (b) social interaction with participants, including <i>researchers, peers</i> (e.g., siblings), <i>teachers</i> (e.g., therapists or instructional assistants), or multiple types.
Treatment components	Antecedent	Antecedent components implemented, including <i>academic-based groups, environmental arrangement, modeling, peer training, prompting, responsive interactions, scripts, social skills training, video, and written priming interventions</i> .
	Consequence	General category for interventions consisting of the contingent presentation of stimuli (e.g., in-vivo reinforcement).
Outcome variables	Target behavior	Reported dependent variables classified as either <i>general</i> or comprising a variety of behaviors within or across response classes, and <i>specific</i> , or consisting of a single response topography (e.g., “Hello”).
	Results	Indicated whether study demonstrated a functional relation through the consistent and replicated change in behavior between identical study conditions.

Note. IOA = interobserver agreement; ASD = autism spectrum disorder; ID = intellectual disability.

a social behavior—were differentiated by their presence (scripts) or absence (written priming) in the measurement context. Modeling was differentiated from prompting as a demonstration of a desirable behavior without additional assistance to perform the behavior. Interventions that *only* featured demonstrations of the target behavior were coded as modeling. Prompting hierarchies that included modeling among other cues that ensured correct responding were coded as prompting.

Outcome variables. Reviewers assessed two variables related to study outcomes. *Target behavior* codes assessed the breadth of response topographies evaluated in each of

the studies. Researchers coded target behaviors as either *general* or *specific*. General behaviors included a variety of responses within or across response classes (e.g., initiations, joint attention [JA]). Specific behaviors were those defined as having a single topography (e.g., saying “Hello”).

Results were coded to indicate the presence or absence of a functional relation for each study. Reviewers initially coded results for each study as a binary judgment based on visual analysis of a functional relation, defined as a consistent and replicated change in behavior (level, trend, or variability; WWC, 2014) between identical experimental conditions. Results of studies comparing multiple treatments were assessed as separate comparisons if the data

supported the analysis (e.g., alternating treatments design (ATD) with a control condition). “Yes” was coded if a functional relation existed; “no” was coded if no evidence of a functional relation was present. Overall success rates were calculated based on the percentage of studies where the independent variable had consistent positive effects. Success rates were not calculated for any variations or contexts with fewer than five studies, as this would grant considerable impact on effectiveness ratings to a limited number of studies. For the purposes of describing relationships between treatment, outcomes, and participant characteristics, we collapsed continuous variables (e.g., age) into categorical variables. Treatments with a minimum success rate of 80% supported by a minimum of 20 studies across three independently researched articles (WWC, 2014) were considered to possess sufficient evidence of effectiveness.

Tau-*U* values and 90% confidence intervals (Parker et al., 2011) were calculated for each study. Tau-*U* was selected due to (a) the frequent use nonoverlap methods in meta-analyses of single-case research, (b) the applicability of nonoverlap techniques across a range of single-case designs, and (c) the ability of Tau-*U* to adjust for trends in the baseline condition (Parker et al., 2011). Graphed values were extracted using an online application (Huwaldt, 2014). Effect sizes were calculated using a web-based program (Vannest, Parker, & Gonen, 2011). Specific cut points were adapted from the nonoverlap of all pairs [NAP] metric: 0.92 for a large effect, 0.66 for a medium effect, and 0.5 for a weak effect (Parker & Vannest, 2009). Values lower than 0.5 were considered non-effects, which are a modification of usual reporting procedures but consistent with NAP interpretation. All reported effects were weighted across replications for each study (i.e., rather than single A-B comparisons). Alternate cut points featured in recent research (small, <0.5; medium, 0.5–0.7; large, >0.71; Crutchfield, Mason, Chambers, Wills, & Mason, 2015) were compared with the NAP-based values.

IOA

The first author coded all articles, and the second author coded 23 randomly selected articles (20%) for the purposes of conducting reliability assessments. IOA was 95.4% across codes, including visual analysis (range = 89%–100%). Codes with the lowest agreement were arrangement (89%) and educational placement (91%). Agreement for data extraction and Tau-*U* was 97% (calculated for 45 studies; 17%).

Results

Methodology

Authors used a variety of designs to assess outcomes across studies; the most common designs were multiple baseline

across participants ($n = 109$) or behaviors ($n = 60$) and withdrawal ($n = 38$) designs. Although collecting reliability data for 20% of sessions with at least 80% fidelity was required for inclusion, few studies (32%; $n = 86$) met contemporary standards for IOA collection and reporting for 20% of sessions in *each condition*. Procedural fidelity data were reported for all primary comparison conditions (e.g., baseline and intervention) in only 65 studies (22%). Social validity data were collected in 123 studies (46%) and was most frequently assessed via satisfaction or acceptability surveys or interviews ($n = 66$); blind raters ($n = 55$) and normative comparisons ($n = 32$) appeared less frequently.

Participant Characteristics

The reviewed studies included 409 participants with ASD; average participant age was 7 years. Age was not reported for 19 participants; authors reported information such as age range or grade level. The majority of participants with reported ages were 3 to 6 years old ($n = 206$; 53%). When divided into 4-year blocks, most participants were pre-school (2–5 years; $n = 178$) or early elementary school aged (6–9 years; $n = 135$). Relatively few participants were upper-elementary to middle school aged (10–13 years; $n = 51$), fewer were high school aged (14–17 years; $n = 22$), and only four were 18 years of age or older. Few participants were female ($n = 55$; 14%). Almost all participants were identified as having autism ($n = 278$); others were identified as having Asperger syndrome ($n = 22$), pervasive developmental disorder ($n = 30$), ASD ($n = 44$), and high-functioning autism ($n = 18$). School placement was reported for 240 participants; when reported, a general education or inclusive placement was the primary or only placement for 149 participants (62%); a self-contained placement was reported for 91 students (38%). Preschool-aged students had a lower percentage of self-contained placements ($n = 23$; 27%), middle and high school students had higher percentages (36% and 41%, respectively), and 6- to 9-year-olds had the highest percentage of self-contained placements ($n = 41$; 47%).

Authors reported intellectual ability for only 24% of participants ($n = 98$). Of these, 49% ($n = 48$) were identified as having an ID or had a reported IQ score of below 70 points. Cognitive estimates were reported for 32% ($n = 60$) of participants less than 7 years of age without an IQ evaluation. Of these, 35 (58%) were reported to have significant impairments and 25 (42%) were reported to not have such impairments. Overall, 83 participants (20%) were reported to have ID or cognitive impairments, 75 (18%) were reported not to have these impairments, and no information was provided for the remaining participants ($n = 251$; 61%). Few studies included data on cognitive ability and placement ($n = 83$), but available data suggest that those with ID were more likely to have a placement in a self-contained setting: 56%

of participants with ID were in a self-contained setting while only 12% of those reported as having no ID were in self-contained settings.

Contextual Variables

Setting and arrangement. Most studies were conducted in school settings, with 132 (49%) conducted in the context of student-led school activities (e.g., recess) and 48 (18%) conducted in the context of teacher-led activities (e.g., direct instruction). Other common settings included homes ($n = 29$; 11%) and clinics ($n = 34$; 13%). Remaining studies were conducted in multiple ($n = 18$; 7%) or unreported ($n = 2$; 1%) settings. Measurement was often conducted in 1:1 arrangements (adult implementer; $n = 148$; 56%), with other individuals in small group arrangements ($n = 82$; 31%), in whole group arrangements ($n = 2$; 1%), or arrangements with no identified group size ($n = 31$; 12%; for example, typical play settings with any number of peers).

Implementers and social partners. Researchers implemented instruction in 140 studies (53%). Other implementers included teachers, therapists, or assistants ($n = 37$; 14%), peers ($n = 27$; 10%), and parents ($n = 17$; 6%). In 32 studies (12%), the role of the implementer was not identified; 10 studies (4%) featured multiple implementers. Peers ($n = 117$; 44%) and researchers ($n = 82$; 31%) commonly appeared as social partners. Parents ($n = 14$; 5%) and teachers ($n = 17$; 6%) appeared less often. Some studies used a variety of social partners ($n = 23$; 9%), and a small number did not report social partner type ($n = 10$; 4%). The most common combination of implementer–social partner was researcher–researcher; the researcher implemented the intervention and served as the social partner. As unspecified implementers and social partners were likely researchers, 35% of the studies may fit into this researcher–researcher category.

Treatment Components

Most studies ($n = 161$; 61%) used more than one treatment component. The most frequently used stand-alone antecedent interventions were prompting ($n = 27$; 10%), peer training ($n = 22$; 8%), and social skills training ($n = 17$; 6%). Priming ($n = 15$; 6%), environmental arrangement ($n = 14$; 5%), and video-based interventions ($n = 13$; 5%) were also used alone; no other intervention component was used in isolation at least 10 times. The most common intervention combinations (used at least 10 times; see Table 2) were (a) prompting plus social skills training, (b) prompting plus environmental arrangement, and (c) prompting plus scripts. When considering uses across a number of intervention combinations, prompting ($n = 112$; 43%), social skills training ($n = 69$; 26%), and peer training ($n = 57$; 22%) frequently

appeared as treatment components. Environmental arrangement ($n = 38$; 14%), scripts ($n = 54$; 21%), video ($n = 29$; 11%), priming ($n = 24$; 9%), responsive interactions ($n = 16$; 6%), academic-based groups ($n = 7$; 3%), and modeling ($n = 4$; 2%) appeared less frequently. In nine studies (3%), interventions categorized as “other” were used; these included sensory-based interventions, self-management training, imitation training, and training the use of alternative communication methods; these were usually combined with other intervention components.

Outcome Variables

The majority of studies ($n = 139$) measured specific behaviors; remaining studies ($n = 124$; 47%) measured general behaviors. Among studies that measured general behaviors, six studies measured conversation (5%), 36 measured interactions (30%), 31 measured initiating only (24%), 15 measured responding only (12%), nine measured JA behaviors (8%), and 22 measured social engagement (18%). Studies measuring specific behaviors reported a variety of target behaviors, including responses to certain affective stimuli, answering or asking targeted questions, and engaging in trained helping or sympathetic behaviors.

Results by participant characteristics. In 263 studies, there were 195 demonstrations of a functional relation (74%) as judged by visual analysis using consistency of effect and the presence of adequate replications of effect as criteria. In terms of cognitive disability across studies, 47 reported only participants without ID (18%), 34 reported only participants with ID (13%), and 15 studies reported including participants with and without ID (6%). The remaining studies ($n = 167$; 63%) did not report ID status for at least one participant. Success rates for studies with participants with ID were 82%, success rates for participants without ID were 56%, and the success rate in studies with both types of participants was 64%. Success rates of interventions by ID status of participants are shown in Table 2.

We characterized results by participant age by using nominal categories described earlier (i.e., 2–5, 6–9, 10–13, 14–17, 18+ years); some studies included participants from multiple categories. Study effects varied based on participant age groups. Many studies with adults (18+ years) and high school students (14–17 years) had high success rates (12 of 13 studies; 92%), but there are too few studies to make confident conclusions about these results. Studies conducted with 10- to 13-year-olds had the lowest success rates (22 of 37 studies; 59%), studies with 6- to 9-year-old participants had high success rates (56 of 65 studies; 86%), and studies with 2- to 5-year-olds had moderate success rates (58 of 84 studies; 69%). Studies with mixed age groups had success rates that were variable (e.g., studies with 2- to 5- and 6- to 9-year-olds had a

Table 2. Percentage Success Rates (and Total Numbers of Studies) by Participant, Behavior Characteristics, Implementer, and Social Partner.

Tx	n	SRt	Rate by behavior type		Rate by cognitive ability ^a		Rate by age (in years) ^b		Rate by implementer			Rate by social partner		
			S	G	ID	No ID	2–10	11–21	R	T	O	R	P	O
PM + SC	34	91	83 (29)	100 (5)	— (2)	— (1)	94 (18)	77 (13)	88 (26)	100 (5)	— (3)	86 (21)	— (4)	100 (9)
PM	27	89	90 (20)	86 (7)	80 (10)	— (0)	85 (20)	100 (5)	72 (11)	100 (9)	100 (7)	70 (10)	— (3)	100 (14)
PT	22	41	— (3)	37 (19)	— (3)	22 (9)	81 (11)	22 (9)	— (0)	— (0)	— (0)	— (0)	41 (22)	— (0)
PT + SST	20	80	88 (17)	— (3)	100 (5)	67 (6)	73 (15)	— (4)	83 (12)	— (2)	100 (6)	— (0)	79 (19)	— (1)
EA	18	100	— (0)	100 (18)	— (0)	— (0)	100 (12)	— (0)	100 (6)	— (0)	100 (10)	100 (6)	100 (9)	— (1)
SST	17	94	91 (12)	100 (5)	— (2)	100 (7)	92 (12)	— (3)	92 (13)	— (2)	— (4)	90 (10)	— (3)	100 (6)
PR	15	47	43 (7)	50 (8)	— (0)	36 (11)	67 (6)	— (3)	55 (9)	— (1)	40 (5)	33 (6)	50 (8)	— (1)
VB	13	54	67 (9)	— (4)	— (2)	— (3)	63 (8)	— (0)	71 (7)	— (0)	33 (6)	71 (7)	33 (6)	— (0)
PM + EA	11	100	100 (5)	100 (6)	— (0)	— (0)	100 (11)	— (0)	100 (6)	100 (5)	— (0)	100 (5)	— (1)	100 (5)
PM + SST	10	90	— (3)	86 (7)	— (2)	— (0)	90 (10)	— (0)	— (2)	— (0)	100 (8)	— (0)	86 (7)	— (3)
≥3 comp	27	70	69 (13)	71 (14)	100 (5)	— (3)	64 (22)	— (4)	67 (18)	100 (5)	— (4)	— (2)	71 (17)	63 (8)

Note. Tx = treatment; SRt = success rate; S = specific; G = general; ID = intellectual disability; R = researcher; T = teacher; O = other; P = peer; PM = prompting; SC = scripting; PT = peer training; SST = social skills training; EA = environmental arrangement; PR = priming; VB = video-based; ≥3 comp = 3 or more treatment components.

^aStudies including participants with and without ID or did not report ID were excluded. ^bStudies including participants including both age ranges were excluded.

success rate of 59%; studies with 6- to 9- and 10- to 13-year-olds had a success rate of 85%). Success rates for younger (2–10 years) and older (11 years and older) age groups for specific intervention components are shown in Table 2.

Results by contextual variables. We evaluated success rates for studies that reported a single setting, including teacher-directed school activities, child-directed school activities, clinics, and homes ($n = 205$). Success rates were not markedly different, although studies conducted in homes had the lowest success rates (67%), and those conducted in clinics (79%) and schools (73% child directed, 83% teacher directed) had higher success rates. In terms of instructional arrangement, success rates were lowest (68%) for studies that delivered social instruction without a specified number of peers present; they were higher (74%) within a 1:1 instructional arrangement and were highest (80%) in small group arrangements. Success rates varied by implementer with lower rates for parents (60%) and peers (52%) compared with teachers (84%) and researchers (81%; these data include 224 studies in which a single implementer was named). Interestingly, parents as social partners were also associated with low rates of success (58%) while studies with peers as social partners had higher success rates (71%), nearly equal to that of researchers (78%). Teachers as social partners were associated with 100% success rates, but there were relatively few studies in this category ($n = 17$).

Results by treatment component. The percentage of demonstrated functional relations (e.g., success rate) for all intervention combinations used more than 10 times appears in Table 2. Several intervention combinations exhibited success rates of 100% (environmental arrangement alone, prompting plus environmental arrangement). Several other interventions had high success rates, including prompting and social skills training alone and in combination and prompting plus the use of scripts. Interventions with markedly lower success rates included peer training, video-based interventions, and priming. Interventions including three or more components had average success rates.

Success rates with teachers implementing prompting (alone or in combination with environmental arrangement or scripts) exceeded the average success rate (74%). Interventions involving peer social partners exhibited success rates more than 10% below average when combined with peer training, priming, or video-based interventions. Peer social partner interventions involving environmental arrangement or that combined peer training or prompting with social skills training exhibited average to above-average success rates. Interventions including children aged 2 to 10 years exhibited average to above-average success across common interventions excluding combined peer and social skills training packages, which exhibited below average success rates. Interventions for children aged 11 to 21 years involving prompting (alone or with scripts) demonstrated average to above-average success rates; peer training had lower than average success.

Because many authors did not report ID status, few conclusions can be made about the effect this variable might have on intervention effectiveness. For students with ID, prompting and combinations of peer and social skills training were generally successful interventions. Prompting interventions had a high success rate for students with ID regardless of age group (i.e., age 2–10 years, 11+ years). Of the interventions commonly featured in the research, only peer training (alone or with social skills training), priming, and social skills training have sufficient evidence for individuals without ID. Social skills training alone exceeded the average success rate. Moreover, the evidence of effectiveness for interventions for older students without ID is insufficient. Peer training alone had less than average success when peers were target social partners, but rates were improved when peer training was combined with social skills training for the target participant. For older individuals and those with ID, many interventions do not have sufficient evidence of effectiveness. Because many authors did not report ID status, few conclusions can be made about the effect this variable might have on intervention effectiveness. Only prompting and social skills training had sufficient evidence of high success rates for specific groups (e.g., older individuals with ID).

Results by target behavior. The success rate across studies, including all intervention types, was 74%. A higher success rate was observed among studies designed to increase specific (80%) rather than general (70%) behaviors. For specific intervention types, no differential effectiveness appeared based on target behavior type (see Table 2), although this may be a function of the relatively small number of studies in each category. Insufficient evidence exists for studies of peer training and combinations of prompting and social skills training on specific behaviors. Video-based interventions and combinations of peer and social skills training yielded above-average success rates with specific behaviors. However, a paucity of evidence was observed for interventions involving generalized behaviors. With these exceptions—and priming, which exhibited low success rates for both behavior types—success rates were average to above average across all common interventions for specific and general behaviors.

Tau-U estimates. For 263 studies, Tau-U estimates ranged from 0.0 to 1.0 with 114 strong effects, 79 medium effects, 27 weak effects, and 43 non-effects when typical cut points were used (see Table 3). Thus, Tau-U estimates identified 193 medium or strong effects (73%). Only 22 studies (all identified with non-effects via both visual analysis and Tau-U values of less than 0.5) had confidence intervals within a single category (e.g., small effect). More than half of the confidence intervals spanned three or four of four categories (57%; that is, non-effects to medium effects, small effects to large effects, or non-effects to large effects).

Table 3. Agreement and Disagreement Types With Traditional and Alternative Tau-U Values.

Traditional Tau-U values	Visual analysis conclusion	
	No	Yes
Large effect (>0.92)	1	113
Medium effect (0.66–0.92)	12	67
Small effect (0.50–0.65)	14	13
Non effect (<0.50)	41	2
Alternative Tau-U values		
Large effect (>0.92)	1	113
Medium effect (0.71–0.92)	5	60
Small effect (0.50–0.70)	21	20
Non effect (<0.50)	41	2

Note. Agreements are shown in shaded cells. Traditional values are based on those specified for nonoverlap of all pairs (Parker & Vannest, 2009). Alternative values are modified based on cut points used by other research teams (e.g., Crutchfield, Mason, Chambers, Wills, & Mason, 2015).

Table 4. Potential Sources of Disagreement.

Conclusion			
VA	Tau-U	Data pattern (design type)	n
No	Yes	Inconsistent delay in behavior change across tiers (multiple baseline)	5
No	Yes	Inconsistent effects (multiple baseline)	3
No	Yes	No return to baseline (similar patterns between B1 and A2; A-B-A-B)	2
No	Yes	Similarity between baseline and intervention trends, with low overlap (multiple baseline)	3
Yes	No	Separation in data paths over time (ATD)	1
Yes	No	Considerable overlap but changing trends between conditions (A-B-A-B, multiple baseline)	3
Yes	No	Initial overlap with improving data in intervention conditions (multiple baseline)	8
Yes	No	Small but consistent effects across tiers (multiple baseline)	3

Note. VA = visual analysis.

Agreement between visual analysis and Tau-U. For the purposes of comparison, strong and medium effects were considered agreements with “yes” judgments for visual analysis and weak and non-effects were considered agreements with “no” judgments. Overall, agreement between visual analysis and

Tau-*U* effect size categories was 89.4%, with disagreements occurring for data patterns commonly accepted as problematic for overlap-based methods (Table 4). The NAP-based effect cut points identified a relatively higher number of studies as having medium to high effects ($n = 13$) with no functional relation judged via visual analysis; when alternative values were used, this issue was minimized while total error rates remained stable.

Discussion

Our review examined the characteristics and outcomes of single-case studies of social skills interventions for children with ASD. Fewer than half of studies met standards for reporting IOA, fidelity, or social validity data. Most participants were young children and roughly 50% had ID or cognitive impairments (although many authors did not report ID status). About 25% of studies included no indigenous implementer *and* no typical social partner (e.g., researcher served as both). Approximately 75% of studies featured effective treatments for individuals with ASD. Lower success rates were noted when parents or peers implemented interventions and when studies were conducted outside of schools or clinics. Among interventions, prompting (with and without additional components), environmental arrangement, and social skills training resulted in the highest success rates.

Our findings differ from previous syntheses, which found success rates of 90% or better by evaluating only one comparison (Wang et al., 2013) or separating results by participants or behaviors (Reichow & Volkmar, 2010). The results differ from a review of studies of interventions for preschoolers with ASD, which showed large differences in success rates for generalized (33%) versus specific (82%) behaviors (Yoder, Bottema-Beutel, Woynarowski, Chandrasekhar, & Sandbank, 2014). Given that few single-case studies in the earlier review measured generalized variables, this finding may primarily pertain to group studies. Previous studies attributed some degree of success to a study if a change occurred but was not replicated. We believe our metric to be more conservative, given our consideration of studies as a whole and exclusive focuses on studies with sufficient experimental control to demonstrate change with appropriate replications.

Visual analysis of outcomes generally corresponded with Tau-*U* values. Disagreement regarding the presence of positive effects was minimized when alternatives to NAP cut points were used. Additional research is needed to determine the propriety of nonoverlap techniques, the accuracy of various cut point values, and specific values associated with *non-effects* (e.g., no intervention could be identified as ineffective or associated with lack of behavior change). The relatively high-agreement percentages between visual analysis conclusions and Tau-*U* values may be due to the inclusion of only methodologically rigorous studies (e.g., with adequate numbers of data points and potential replications

of effect) and the comparison of weighted Tau-*U* values for three or more demonstrations of effect with overall determination of function relations (rather than A-B comparisons).

Limitations

This review has several notable limitations. The extent to which the identified studies encompass the totality of research regarding individuals with ASD is uncertain, as resources did not permit the evaluation of “grey literature” (e.g., unpublished studies). Nonetheless, we feel our emphasis on published single-case design studies with sufficient rigor to demonstrate a treatment effect (i.e., functional relation) represents a unique contribution to the literature.

We did not perform a meta-analysis of Tau-*U* estimates providing an overall estimate of the effects of all potential interventions across all variations, as our aim was to identify effectiveness of social skills interventions under *given* conditions. Aggregating disparate interventions and variables does not provide precise information for informing research and practice. Between-study comparisons using overlap methods may be inaccurate when measurement differs across studies (e.g., duration measurement vs. interval-based measurement; cf. Ledford, Ayres, Lane, & Lam, 2015; Pustejovsky, 2015). Differences in overlap-based metrics do not accurately reflect differences in behavior change between conditions (e.g., consistent small effects may be superior to inconsistent effects, but when overlap metrics are used, this difference is not apparent in the size of the metric). Thus, analyses of results and study characteristics were conducted using visual analysis outcomes, which were largely in agreement with Tau-*U* values.

Implications for Research

Given the number of included studies, it is clear that considerable research exists related to increasing pro-social behaviors for individuals with ASD. However, there is little support for isolated or combined treatment components. More research is needed, particularly for older individuals with ASD. That interventions were found to be most successful for elementary-aged participants and less successful for preschool and middle school participants may suggest that (a) some interventions are more successful for individuals who have developed school-readiness behaviors, and (b) we know less about what components are effective for older individuals. Additional research is needed for treatments that are conducted by indigenous implementers (e.g., teachers, parents, peers), even for relatively well-researched interventions with high success rates, to ensure that these interventions are not only successful when implemented by researchers, but that they are also feasible and effective when used in typical contexts.

Peer training, priming, and video-based interventions appeared less successful than other often-used techniques; more research is needed to determine whether adding other

specific treatment components or targeting specific skills results in increased efficacy (e.g., adding social skills training to peer training). Relatively few conclusions can be drawn about intervention for students with ID due to the limited number of studies reporting ID status. It may nonetheless be reasonable to expect some interventions to be more effective for individuals with higher level cognitive skills. For example, social skills training depends on the ability to learn and retain information about behaviors to be performed later. This does not suggest that specific procedures should not be used with students who have intellectual or cognitive impairments; rather, it is an argument that researchers should consider and report participant characteristics so the field can determine *what is effective, for whom, and under what conditions*.

Although we purposefully excluded studies that inhibited decision making about the presence of a functional relation, we found that multiple baseline across participants designs resulted in less favorable ratings than other designs. This may be due to differences in participant characteristics across tiers; as we argue elsewhere in this article, researchers should consider that interventions may be differentially effective across participants. Use of within-participant designs (e.g., multiple baseline across social partners, multiple probe across contexts) may be prudent if researchers are interested in testing intervention effects for participants with varying characteristics.

Implications for Practice

Although many of the included treatment components are considered EBPs for individuals with ASD (Wong et al., 2014), practitioners should consider the extent to which evidence for the practice has been accumulated *for the behaviors they encounter, for the individuals they serve, in contexts similar to theirs*. Careful author reporting of participant characteristics, IOA, and fidelity are critical to building a practical literature base. Given that complex interventions (i.e., three or more components) did not result in above-average success rates, practitioners may consider using parsimonious interventions rather than complex ones. Only two treatment components (prompting, social skills training) had sufficient evidence of high success rates in any treatment group. Clearly, more research is needed. The current review highlights the likelihood that even interventions with high success rates are unlikely to work in every context, for all individuals. This underscores the need to not only identify EBPs but also train and support practitioners in the use of data-based decision making in modifying or changing interventions if sufficient behavior change does not occur.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article

Funding

Support for this research was provided by the Institute of Education Sciences Grant #R32B130014.

References

- American Psychological Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.
- Bellini, S., Peters, J. K., Benner, L., & Hopf, A. (2007). A meta-analysis of school-based social skills interventions for children with autism spectrum disorders. *Remedial and Special Education, 28*, 153–162.
- Crutchfield, S. A., Mason, R. A., Chambers, A., Wills, H. P., & Mason, B. A. (2015). Use of a self-monitoring application to reduce stereotypic behavior in adolescents with autism: A preliminary investigation of I-connect. *Journal of Autism and Developmental Disorders, 45*, 1146–1155.
- Gast, D. L., & Ledford, J. R. (2014). *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed.). New York, NY: Routledge.
- Goin-Kochel, R. P., Myers, B. J., & Mackintosh, V. H. (2007). Parental reports on the use of treatments and therapies for children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 1*, 195–209.
- Goldstein, H., Lackey, K. C., & Schneider, J. B. (2014). A new framework for systematic reviews: Applications to social skills interventions for preschoolers with autism. *Exceptional Children, 80*, 262–286.
- Gresham, F. M., Sugai, G., & Horner, R. H. (2001). Interpreting outcomes of social skills training for students with high-incidence disabilities. *Exceptional Children, 67*, 331–344.
- Huwaldt, J. (2014). Plot digitizer (Version 2.6.6). Available from <http://plotdigitizer.sourceforge.net>
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2*, 217–250.
- Ledford, J. R., Ayres, K. A., Lane, J. D., & Lam, M. F. (2015). Accuracy of interval-based measurement systems in single case research. *Journal of Special Education, 49*, 104–117. doi:10.1177/0022466915568975
- Ledford, J. R., & Wolery, M. (2013). Procedural fidelity: An analysis of measurement and reporting practices. *Journal of Early Intervention, 35*, 173–193.
- McConnell, S. R. (2002). Interventions to facilitate social interaction for young children with autism: Review of available research and recommendations for educational intervention and future research. *Journal of Autism and Developmental Disorders, 32*, 351–372.
- McDonald, M. E., Pace, D., Blue, E., & Schwartz, D. (2012). Critical issues in causation and treatment of autism: Why fads continue to flourish. *Child & Family Behavior Therapy, 34*, 290–304.
- Odom, S. L., Collet-Klingenberg, L., Rogers, S. J., & Hatton, D. D. (2010). Evidence-based practices in interventions for children and youth with autism spectrum disorders. *Preventing School Failure, 54*, 275–282.
- Odom, S. L., Hume, K., Boyd, B., & Stabel, A. (2012). Moving beyond the intensive behavior treatment versus eclectic dichotomy: Evidence-based and individualized programs for learners with ASD. *Behavior Modification, 36*, 270–297.

- Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 40*, 357–367.
- Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification, 35*, 302–322.
- Pustejovsky, J. E. (2015, April). *Operational sensitivities of non-overlap effect sizes for single case designs*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Reichow, B., Barton, E. E., Sewell, J. N., Good, L., & Wolery, M. (2010). Effects of weighted vests on the engagement of children with developmental disabilities. *Focus on Autism and Developmental Disabilities, 25*, 3–11.
- Reichow, B., & Volkmar, F. R. (2010). Social skills interventions for individuals with autism: Evaluation for evidence-based practices within a best evidence synthesis framework. *Journal of Autism and Developmental Disorders, 40*, 149–166.
- Riley-Tillman, T. C., & Burns, M. K. (2010). *Evaluating educational interventions: Single-case design for measuring response to intervention*. New York, NY: Guilford Press.
- Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review, 17*, 881–901.
- Stewart, R. M., Benner, G. J., Martella, R. C., & Marchand-Martella, N. E. (2007). Three-tier models of reading and behavior: A research review. *Journal of Positive Behavior Interventions, 9*, 239–252.
- Strain, P. S., Schwartz, I. S., & Barton, E. E. (2011). Providing interventions for young children with autism spectrum disorders: What we still need to accomplish. *Journal of Early Intervention, 33*, 321–332.
- Vannest, K. J., Parker, R. I., & Gonen, O. (2011). *Single case research: Web based calculators for SCR analysis* (Version 1.0). College Station: Texas A&M University. Available from <http://www.singlecaseresearch.org/>
- Wang, S., Parilla, R., & Cui, Y. (2013). Meta-analysis of social skills interventions of single-case research for individuals with autism spectrum disorders: Results from three-level HLM. *Journal of Autism and Developmental Disorders, 43*, 1701–1716.
- What Works Clearinghouse. (2014). *Procedures and standards handbook* (Version 3.0). Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf
- Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education, 44*, 18–28.
- Wong, C., Odom, S. L., Hume, K., Cox, A. W., Fettig, A., Kucharczyk, S., . . . Schultz, T. R. (2014). *Evidence-based practices for children, youth, and young adults with autism spectrum disorders*. Retrieved from http://autismpdc.fpg.unc.edu/sites/autismpdc.fpg.unc.edu/files/2014_EBP_Report.pdf
- Yoder, P. J., Bottema-Beutel, K., Woynaroski, T., Chandrasekhar, R., & Sandbank, M. (2013). Social communication interventions effects vary by dependent variable type in pre-schoolers with autism spectrum disorders. *Evidence-Based Communication Assessment and Intervention, 7*, 150-174. doi:10.1080/17489539.2014.917780