



Published in final edited form as:

Evid Based Commun Assess Interv. 2013 ; 7(4): 150–174. doi:10.1080/17489539.2014.917780.

Social communication intervention effects vary by dependent variable type in preschoolers with autism spectrum disorders

Paul J. Yoder¹ [Commentary author], Kristen Bottema-Beutel², Tiffany Woynaroski³, Rameela Chandrasekhar⁴, and Michael Sandbank¹

¹Special Education Department, Vanderbilt University, Nashville, TN, USA

²Lynch School of Education, Boston College, Boston, MA, USA

³Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN, USA

⁴Biostatistics Department, Vanderbilt University, Nashville, TN, USA

Abstract

Individuals with autism spectrum disorders (ASDs) have difficulty communicating in ways that are primarily for initiating and maintaining social relatedness (i.e., social communication). We hypothesized that the way researchers measured social communication would affect whether treatment effects were found. Using a best evidence review method, we found that treatments were shown to improve social communication outcomes approximately 54% of the time. The probability that a treatment affected social communication varied greatly depending on whether social communication was directly targeted (63%) or not (39%). Finally, the probability that a treatment affected social communication also varied greatly depending on whether social communication as measured in (a) contexts very similar to treatment sessions (82%) or (b) contexts that differed from treatment on at least setting, materials, and communication partner (33%). This paper also provides several methodological contributions.

Keywords

Autism spectrum disorders; Early intervention; Research synthesis; Measurement; Systematic review

Difficulty with communication for social purposes is a core feature of autism spectrum disorders (ASDs; American Psychiatric Association, 2000). Children with ASDs exhibit a specific impairment in communication used to relate to others (i.e., social communication). In contrast, communication that is not underpinned by a desire to share with others, but that instead serves a regulatory or requesting function, is relatively spared (Shumway & Wetherby, 2009; Volkmar, Lord, Bailey, Schultz, & Klin, 2004; Wetherby, 2006). This paper synthesizes the best evidence from several decades of intervention research that measures treatment effects on social communication in preschool children with ASDs. This type of synthesis presents several methodological challenges, which we address in hopes of

For correspondence: Paul Yoder, 228 Peabody, Vanderbilt University, Nashville, TN 37212, USA. Paul.Yoder@vanderbilt.edu.

Declaration of interest: The authors have no conflicts of interest and are solely responsible for the content of this article.

providing guidance to the field. The findings and methods have important implications for practitioners and researchers, especially in regards to how we evaluate the effects of treatment on critical outcomes.

RESEARCH SYNTHESIS APPROACH

This review represents a “best evidence” synthesis of treatment studies that have measured social communication outcomes in preschoolers with ASDs over the last several decades. A best evidence synthesis focuses on evidence of treatment effects across only the most internally valid treatment studies (Slavin, 1986). A treatment effect is evidence that treatment is one of the causes of change. Only studies with strong evidence of internal validity were reviewed because weaker studies do not have the necessary control to warrant high confidence in inferring treatment effects (Millar, Light, & Schlosser, 2006). One disadvantage of this approach is that it reduces the number of studies reviewed. However, arguments that one can test whether the presence or size of the treatment effect varies as a function of the internal validity of the study assumes (a) that there is an effect size indicator that identifies the portion of change due to the treatment and (b) that there are sufficient number of studies to provide a sufficiently powered test of the effect. Neither is true in our case. Thus we restrict our review to the studies with the capacity to detect treatment effects with high confidence.

There is some disagreement about the definition of a treatment effect. Informal conversation with some single-case experimental experts indicates that at least some define a functional relation as evidence that the independent variable (e.g., the treatment) is the one and only cause of change in the dependent variable. This logic is best exemplified in withdrawal designs (e.g., ABAB). In contrast, group experimental design seeks to identify the portion of change that is due to treatment and interprets such as the treatment is only one of the causes of change in the dependent variable. A Cohen’s d that quantifies the difference between experimental and control groups of 2.0 corresponds to an R square of .50 (i.e., 50% of the variance in the change in the outcome is accounted for by the treatment group difference). As it is rare to find an internally valid treatment study with effect sizes over 2.0, it is clear that most internally valid treatment studies using the group experimental design logic that claim to have found a treatment effect demonstrate that the treatment accounts for a minority of the variance in change in the dependent variable. In contrast, single-case experimental methodology does not currently provide an agreed-upon way to identify the portion out change due to treatment versus change due to other factors. This may account for the difference in definition of a treatment effect. The important aspect that is shared by all treatment researchers is that at least some of the change in the dependent variable is attributable to the treatment by virtue of controlling for all other explanations for the findings the research design logic uses to infer treatment effects. We used the less restrictive meaning of treatment effect because we wanted to use a definition that addressed this point of agreement among treatment researchers.

We anticipated that many, if not most, of the available internally valid experiments testing treatment effects on social communication would be single-case experimental designs. We

have made the potentially controversial decision not to use an effect size as a continuous index of the degree to which a treatment effect was shown for two reasons.

First, none of the currently proposed single-case “effect size” indicators adequately reflect the degree to which the treatment causes (i.e., is functionally related to) change in the dependent variable (Wolery, Busick, Reichow, & Barton, 2010). A functional relation is marked by a shift in mean, trend, or variability from one design phase to another and by consistent replication of this shift across phases and tiers. The majority of single-case experimental design “effect size” indicators evaluate only one marker of effect (e.g., a shift in level) or are compromised by shifts in trend and thus are inadequate indicators of treatment effects (Campbell & Herzinger, 2010). To be fair, there have been recent attempts to improve our ability to quantify treatment effect sizes in single-case experiments (e.g., Beretvas & Chung, 2008b, for review). Unfortunately, and particularly salient for communication disorders, these attempts usually fail to address a critical single-case experimental design element for studies of treatment effects on nonreversible dependent variables (e.g., language development). Single-case experiments on nonreversible dependent variables most typically require a multiple baseline approach in which a critical criterion for inferring a functional relation is replication of a shift between baseline and treatment phase in dependent variable across tiers (Gast, 2010). Most proposals, including those using mixed-level modeling, quantify the degree of change within a tier or graph (Beretvas & Chung, 2008a, 2008b). One proposal for an effect size index that does attempt to address the “vertical replication” needed for multiple baseline design to demonstrate a functional relation is to use the proportion of attempted replications that were demonstrated (Reichow, Volkmar, & Cicchetti, 2008). However, this does not address other important design elements of the multiple baseline designs, such as the immediacy of change after the onset of the treatment phase (Lieberman, Yoder, Reichow, & Wolery, 2010). One might suggest that we report several effect size indicators for singlecase experiments, but this begs the question of what weight each index should have when aggregating a single index and whether this index would reflect the amount of change due to the treatment, rather than the amount of change that occurred while the participant was treated.

Second, the proposed “effect size” indicators for single-case experiments do not provide comparable information to the commonly used effect sizes for group experiments. Thus, a traditional metaanalytic approach using effect size as the primary metric of treatment effect would have necessitated either that we used unsatisfactory “effect size” indicators for single-case experimental designs, or that we excluded the majority of the evidence on social communicative interventions. For these reasons, it was necessary to identify another informative and systematic method of synthesizing across group design and single-case studies.

We opted to classify the dependent variables within each internally valid study in a dichotomous manner to indicate the presence or absence of strong evidence of a causal effect between the treatment and the dependent variable change. This enabled us to weight evidence from singlecase experimental designs equally with evidence from group experiments.

We also suspected that several studies may include multiple dependent variables that indexed effects of treatment on social communication. Research synthesis approaches often use statistical procedures that assume independence of analysis units to test hypotheses about the probability of treatment effects. This assumption is violated when multiple dependent variables are analyzed from the same study (Lipsey & Wilson, 2001).

One problematic approach to addressing multiple effects within a study is to select only a single dependent variable that meets criteria for inclusion from each study (Riley, Thompson, & Abrams, 2008). First, there is understandable disagreement among experts regarding the best selection method. Random selection of a dependent variable does not necessarily result in inclusion of an outcome that is “representative” of the set of dependent variables in the study. This is, in part, due to the fact that the set of dependent variables from which to sample is usually small. Selecting dependent variables based on a “favored” way of measuring a construct is also not feasible in many searches because either the preferred method is not represented or there is no consensus on the “best” way to measure a construct. Selecting the single dependent variable with the highest effect size results in positively biased results. Another strategy to address non-independence of effect sizes within a study is to use a single unweighted average across effects within a study. This approach ignores the degree to which effects are correlated within the study. Both approaches reduce the relevant information available for review. For all of these reasons, we opted not to select or create by averaging a single dependent variable from each study in this review. Instead, we utilized a statistical approach that enabled inclusion of multiple dependent variables from the same study while accounting for the intercorrelation among outcome variables within studies (i.e., clustered bootstrapping).

OUTCOME VARIABLE CATEGORIZATION

Within studies that have targeted social communication in young children with ASDs, there is great heterogeneity in both the types of dependent variables measured and the contexts or manner in which they are measured. As part of our research synthesis approach, we evaluated whether the likelihood of finding a treatment effect on a social communication outcome varied according to two attributes of social communication measures.

Each dependent variable was classified according to: (a) the *proximity* of the outcome to the intervention target, and (b) the potential *boundedness* of the outcome to the intervention context. Outcomes that are assessed by items or behaviors that have a high degree of overlap with treatment targets can be considered *proximal* to the treatment. Conversely, outcomes that are assessed by items or behaviors that are broader than what was taught can be thought of as *distal* to the treatment (Gersten, Fuchs, Compton, Coyne, Greenwood, & Innocenti, 2005). Boundedness is the extent to which the occurrence of a behavior possibly depends upon the features of the treatment context. When behaviors are measured with settings, materials, communication partners, or interaction styles that are highly similar to treatment, it is possible that they may be limited to the context of treatment. We classified these outcomes as possibly *context bound*. In contrast, outcomes assessed in situations that differ from the treatment context on multiple dimensions (i.e., setting, activity, materials, person, and interaction style) more clearly reflect generalization of learning outside the treatment

context. These variables were classified as *generalized*. These concepts should not be confused with near- versus far-transfer (Schunk, 2004). Potentially context bound refers to findings that may not transfer past the treatment context at all and thus fall short of Schunk's criteria for even near-transfer. We acknowledge that classifying dependent variables into dichotomous levels oversimplifies reality. We do so as a first step towards synthesizing results from previous work according to these dependent variable types. We believe these classifications are particularly useful for clinical fields such as speech–language pathology, special education, and clinical psychology, as they offer an index of the treatment's ability to impact the child's development.

Table 1 provides a matrix depicting the four possible categorizations that result from classification of each dependent variable along the two dimensions. Most commonly, distal outcomes are highly generalized outcomes. However, to demonstrate that proximity and boundedness are separable dimensions, we highlight here an unusual case of a potential distal effect that is measured in treatment sessions (i.e., the upper right quadrant of the matrix in Table 1). Similarly, proximal outcomes are often measured in potentially context-bound ways, especially in singlecase experiments. However, to reinforce the point that proximity and boundedness are potentially separable, we also highlight here the moderately common case of examining the highly generalized use of a directly taught skill (i.e., the lower left quadrant of the matrix in Table 1).

It stands to reason that we are most likely to effect change on those behaviors that are directly targeted (i.e., are proximal to the treatment) and that are assessed in situations highly similar to, or identical to, how such behaviors were trained (i.e., are potentially context bound). For example, if we target responding to joint attention bids, we are more likely to observe changes in the child's responding to joint attention behavior and less likely to see changes in other developmentally downstream impairments associated with ASDs, such as initiating joint attention. Turning to the boundedness of behavior change assessment, if the child in the aforementioned example developed responding to joint attention skills in the clinic while interacting with an interventionist in a highly structured condition with highly preferred toys, we might expect that the child is more likely to subsequently display responding to joint attention behavior in the same setting with the same communication partner, interaction style, and materials. In contrast, the child may be less likely to spontaneously respond to a joint attention bid in a home or community setting while interacting with a family member in a natural manner in response to a nonpreferred stimulus. We might particularly anticipate this to be the case in children with ASDs, who are often reported to have difficulties with generalization (Fein, Tinder, & Waterhouse, 1979; Horner, Dunlap, & Koegel, 1988; Hume, Loftin, & Lantz, 2009). There is universal agreement that seeing intervention targets utilized in generative, flexible, and highly generalized ways is the eventual goal of early intervention. However, to our knowledge, previous reviews have not differentiated the effects of treatment by whether or not such desirable outcomes have been achieved. The detail provided in this paper regarding classification of outcome variables may improve specification of treatment effects in future studies and, thus, may be considered one of the most important contributions of this paper.

There is already a call for distinguishing treatment outcomes in research syntheses by proximity. In a recent randomized, controlled trial of a naturalistic, parent-mediated communication intervention, Green and colleagues (2010) noted a progressive attenuation of intervention effects as dependent variables became more distal relative to the treatment targets. Although they found positive effects of treatment for behaviors that were explicitly targeted by the treatment, they did not observe a favorable change in general autism symptoms (i.e., severity of broader autism symptoms as indexed by Autism Diagnostic Observation Scale (ADOS) social communication algorithm total score). Of course, global measures are not designed to be sensitive to treatment effects. However, failure to find evidence of treatment effects on distal variables highlights the need to consider different dimensions of dependent variables when quantifying and summarizing the effects of treatments. The authors of the Green et al. (2010) study observed that research reporting positive treatment effects on social communication skills of young children with ASDs has tended not to evaluate distal dependent variables and called for a reanalysis of study results with systematic consideration of this methodological factor.

Less attention has been dedicated to the potential variation of treatment effects by the boundedness of dependent variables. The distinction between generalized and potentially context-bound behavior change is particularly important in evaluating whether intervention has an effect on development. When a newly acquired social communication behavior generalizes outside of the treatment context, a child will have more opportunities to practice the behavior in interactions across settings and communication partners. Ongoing opportunities for practice, even for proximal behaviors, may bootstrap further development so that more distal achievements are realized over time. In contrast, when a behavior change remains context bound, a child has fewer opportunities to practice the newly learned behavior outside of the treatment setting. Thus, context-bound behaviors are less likely to naturally scaffold future development. Furthermore, if a social communication behavior is only elicited by a very specific combination of contextual factors, it is unclear to what extent the behavior reflects a change in the child's intrinsic motivation to communicate for the purpose of sharing experiences with others.

RESEARCH QUESTIONS AND HYPOTHESES

In summary, a review of the extant literature was necessary to evaluate the evidence that existing treatments impact development of social communication in children with ASDs. Past intervention research on social communication in young children with ASDs has varied in measurement of treatment outcomes. There were both logical and empirical reasons to believe that the way that dependent variables had been measured had influenced the likelihood of finding a causal effect of the intervention. Therefore, this synthesis further evaluated whether the likelihood of finding a treatment effect varied according to dependent variable type. Specific research questions were as follows: (a) Across dependent variable types, what is the across-study probability of a treatment effect on social communication in preschoolers with ASD? (b) Does the probability of a treatment effect significantly favor proximal dependent variables compared to distal dependent variables? (c) Does the probability of a treatment effect significantly favor potentially context-bound changes in

social communication skills compared to generalized changes in social communication skills?

METHODS

Locating and initial screening of the sample of relevant studies

This review was limited to children from birth to 71 months of age who had a diagnosis of autistic disorder (AD), pervasive development disorder—not otherwise specified (PDD-NOS), or Asperger’s syndrome (AS) according to criteria from the third edition, fourth edition, or fourth edition, text revision of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-III, DSM-IV, or DSM-IV-TR*; American Psychiatric Association, 1980, 1994, 2000). Additionally, the review was restricted to empirical studies using experimental research designs in which the independent variable was an intentional teaching or therapy method implemented over at least five sessions. Two methods were used to identify the pool of potentially relevant studies: (a) general-purpose databases, and (b) ancestry searching of recent review articles. When the first method was conducted, five databases (PsycInfo, PsycArticles, ERIC, CSA Linguistics and Language Behavior Abstracts, and Pubmed) were searched simultaneously using ProQuest for the period between 1980 and 2012, inclusive. Four sets of descriptor terms were created to describe dependent variables, age, diagnosis, and methodology: (a) communicat*, joint atten*, social adj interact*, turn adj tak*, discourse, convers*, social adj skill*; (b) preschool, early childhood, pre-k, toddler*, infan*; (c) autis*, Asperger*, PDD*, ASD; and (d) interven*, experiment*, treat*, therap*, teach*. The team conducted a preliminary screening by reading titles and abstracts to identify review articles. From the total pool of review articles, the five most recent, relevant reviews were selected (Howlin, Magiati, & Charman, 2009; Marans, Rubin, & Laurent, 2005; McConnell, 2002; Prizant & Wetherby, 2005; Reichow & Volkmar, 2010). The titles of articles in the reference sections of these five reviews were searched to identify articles that appeared potentially relevant (i.e., “ancestry searching”). If not already identified by the databases, these articles were added to the list of potentially relevant studies. Three hundred and ninety-one potentially relevant studies were identified.

Criteria for inclusion and exclusion

A secondary screening process was conducted to exclude empirical studies that clearly did not meet the selection criterion. Initial exclusion was based on: (a) not meeting participant inclusion criteria; (b) having no, or an insufficient number of, treatment sessions; (c) using a nonexperimental research design (e.g., nonrandomized group experiments, correlational designs, pre–post treatment studies without randomized control groups, AB single-case studies, single-case experiments with insufficient tiers or design phases to afford sufficient replication, single-case experiments with insufficient data within a phase to afford visual analysis of trend or variability; see Table 2 for details); or (d) not written in English. In some instances articles were excluded at this stage based on absence of a social communication outcome, but most often these judgments were deferred for in-depth analysis regarding whether the dependent variables met our definition for social communication. Two hundred and forty-nine reports were excluded during this secondary screening, leaving one hundred

and forty-two potentially relevant studies. See Figure 1 for an illustration of the exclusion process.

We based our definition of social communication on the specific deficits that are seen in children with ASDs—communication to relate to others. Our definition of social communication includes behavior that is at least in part: (a) intentionally communicative (i.e., directed toward another person and about a topic); (b) primarily used for the purpose of sharing affect or interest; (c) not an immediate imitation of a model with no intervening activity; and (d) not exclusively a single response form to a single prompt or question form. A range of dependent variables, such as initiations and responses to peers, initiating and responding to joint attention, the ADOS-G Social Communication Algorithm, and the Vineland Socialization subscale score met these criteria for inclusion (a comprehensive list of variables that met inclusion criteria is delineated in Table 3 along with study characteristics). We excluded dependent variables that only measured (a) spoken language that was not explicitly defined as communicative; or (b) communication that was not, at least in part, for the purposes of relating to others (e.g., only requesting, protesting, directing). We excluded these dependent variables because we suspected that they might be more readily facilitated by treatment in young children with ASDs. Additionally, noncommunicative language and instrumental communication are not the core deficits of children with ASDs.

Studies in this review included outcomes that either (a) clearly and exclusively met this narrow definition of social communication, or (b) met this definition at least in part, but also aggregated social communication according to our definition with communication for nonsocial purposes (i.e., was in part, but not solely, for the purpose of sharing affect or interest). To address potential concerns about this decision, additional analyses were conducted with such aggregated variables excluded (see Table 3).

The final screening involved thorough reading and coding of all remaining potentially relevant studies. A detailed coding manual (available from the first author upon request) guided decisions at this stage. The primary screening and coding judgments were made by four readers (one full professor in Special Education, one postdoctoral fellow in Special Education, one PhD student in Special Education, and one PhD student in Speech–Language Pathology). These readers met weekly for three months to develop and master the coding system. The training method involved teams of two readers reviewing two to five articles independently and discussing discrepancies each week. Disagreements among paired readers were discussed within the full group and resolved through a consensus process. In the event that a consensus was not reached within paired or whole-group discussion, the content lead (the full professor in Special Education) settled the dispute.

See Tables 2 and 4 for criteria used to determine whether a study had the design control for alternative explanations to findings used to infer a causal relation. These decisions were paramount because only types of studies can rule out alternative explanations for results that might otherwise appear to show treatment effects. For group research designs, only randomized between-group experiments were accepted for review. The quality indicators widely accepted for group experiments were required (Gersten et al., 2005, Table 4). For

single-case experimental designs, best practice principles were used to identify essential quality indicators using a widely accepted guide (Kratochwill et al., 2010, Table 2).

One exception to the usual rules for selecting internally valid studies was made in this review: The examiner or reporter could or did have knowledge of timing or membership of the treatment condition. This exception was necessary because, in single-case experimental designs, the dependent variable is most often measured during the treatment session, preventing blindness to treatment condition.

This final screening resulted in the exclusion of an additional 118 articles. Thus, only 24 articles were selected as “best evidence” for evaluating potential effects of a treatment on social communication dependent variables in young children with ASDs.

Data extraction

The process to classify social communication dependent variable is illustrated in Figures 2 and 3. These variables were categorized in two ways. First, variables were distinguished by whether they were proximal (i.e., directly taught by the treatment) or distal (i.e., beyond what was directly taught) to the treatment targets. The second distinction involved determining whether change due to intervention was potentially context bound (i.e., possibly limited to the treatment context) or more clearly generalized (i.e., measured in a manner that reflected generalization of learning to contexts that differed from treatment).

We subsequently coded each dependent variable as warranting “high” or “other” confidence of a treatment effect on the basis of criteria in Tables 2 and 4. For group research designs, only statistically significant and clinically important (defined as having a Cohen’s d greater than 0.5) between-group differences were coded as having a “high” degree of evidence of a causal effect. For single-case experimental designs, the decision process was more complicated. Achieving a “high” rating of causal effect in a single-case experimental study required: (a) that the dependent variable showed a minimum of three basic effects involving a change in variability, level, and/or trend; and (b) that the majority of planned replications showed an effect (Kratochwill et al., 2010). Additionally, for multiple baseline studies, the change in the dependent variable had to begin within one month of treatment onset (Lieberman et al., 2010), and there could be no change in the dependent variable during baseline when the prior tier’s treatment phase began. Withdrawal designs had to demonstrate a countertherapeutic trend or an 80% reduction from the treatment phase median in the withdrawal phase. For alternating treatments designs, there had to be at least five data points with a therapeutic separation of level, trend, or variability across contrasting conditions (Kratochwill et al., 2010).

Agreement on the major coding decisions

The process described above was implemented by consensus teams, which consisted of pairs of trained coders. One half (12) of the studies was independently coded by a second team of coders blind to the previous coding decisions. Agreement between the first and second teams was estimated using agreements/(agreements + disagreements). For the other half of the studies, the second team referred to the first consensus team’s decisions, then either

confirmed or changed the previously determined decision (i.e., verification). In this way, all major decisions were checked by all four judges.

The percentage agreement was calculated for key coding dimensions: (a) proportion of agreed upon social communication dependent variables with agreed-upon level of proximity to what was taught (i.e., proximal vs. distal) = .90; (b) proportion of agreed-upon social communication dependent variables with agreed-upon level of boundedness (possibly context bound versus generalized) = .90; and (c) proportion of agreed-upon social communication dependent variables with agreed-upon decision regarding whether a treatment effect was demonstrated = .90. This set of procedures allowed us to estimate agreement, while ensuring that studies were coded as accurately as possible.

Coding minor elements of the selected studies

Participant information and additional treatment descriptors were also coded. The age, IQ, developmental level, and diagnosis of participants were summarized, if available. This coding was conducted by two graduate students and a research staff member. Training involved a three-way consensus to decide on the description of 10 studies. Once agreement reached .8 among all three coders; a primary coder made the sole decisions for the remaining 14 studies. Twenty-one percent of the studies were randomly selected and independently coded by a second observer with the following results: (a) proportion of participants with a particular autism spectrum disorder (i.e., AD, PDD-NOS, AS) = 1.0; (b) mean chronological age = .8; (c) mean IQ = .8; (d) mean expressive language age = 1.0. Table 3 includes details about the participants, research designs, and treatments represented by the 24 studies that met criteria for inclusion in the present review.

Statistical procedures

As we suspected, multiple dependent variables were often nested within individual studies. In fact, only eight studies included a single social communication dependent variable. One study included a total of 22 dependent variables meeting our definition of social communication. Furthermore, two reports included in our review presented results on the same participants (Kasari, Freeman, & Paparella, 2006; Kasari, Paparella, Freeman, & Jahromi, 2008). The clustered bootstrapping approach that we utilized allows for dependence among variables, but requires that *studies* are independent. Thus, these two reports were treated as one, reducing the number of “studies” for the sake of our analyses to 23.

We used a clustered bootstrapping method (Feng, McLerran, & Grizzle, 1996) to account for the intercorrelation among dependent variables within studies. Two statistics of interest were derived: (a) the probability of a treatment effect across studies (i.e., the number of dependent variables showing a treatment effect divided by the total number of dependent variables in sample) and (b) the difference in probability of finding a treatment effect according to our two dichotomized dependent variable types. Two types of difference scores were computed: (a) one for the proximal versus distal contrast, and (b) one for the context-bound versus generalized contrast. The number of studies resampled was 23 (i.e., equal to the 23 independent studies we reviewed). The lower and upper limits of the 95% confidence

interval for the statistic of interest were identified using the proportion of bootstrapped estimates in tails of the empirical bootstrap distribution of estimates correcting this for bias (Efron, 1981). Confidence intervals need to be adjusted for bias because the empirical distributions generated by the resampling process are often asymmetrical. More details on the analysis process are available from the corresponding author. Simulation studies have shown that bootstrapping in this way retains the intercorrelation among dependent variables without biasing the probability of an effect or its confidence interval (Harden, 2011).

RESULTS

In the 23 independent samples, 60 social communication dependent variables were evaluated for treatment effects. These are described in Table 3. Of the social communication dependent variables included in our review, 62% (37/60) were classified as measuring change in a manner that was highly generalized relative to the treatment context. Additionally, 35% (21/60) of the dependent variables were classified as measuring behaviors distal to the treatment targets. Thus, there was sufficient opportunity to demonstrate treatment effects on the most informative outcomes, assuming that evidence of a treatment effect on distal and generalized social communication development is the ultimate goal of intervention in young children with ASDs.

Probability of strong evidence of a treatment effect

Our first research question relates to the average probability of a treatment effect on social communication dependent variables in preschoolers with ASD. Strong evidence of a treatment effect on social communication was demonstrated in 54% of the dependent variables. The 95% confidence interval around this estimate is 38% to 77%. Thus, the probability of finding a treatment effect on social communication was greater than chance (i.e., the confidence interval does not include 0). Very similar findings occurred when aggregate dependent variables were excluded (53% probability of a treatment effect, 95% CI [34%, 71%]). This finding does not classify the probability of finding a treatment effect by type of dependent variable.

The probability of a treatment effect varied by proximity of the outcome

The treatment effect difference was 24%, favoring proximal dependent variables. This difference was significantly greater than expected by chance; 95% CI [1%, 62%]. The probability of strong evidence of a treatment effect on proximal dependent variables was 63%; 95% CI [43%, 81%]. In contrast, the probability of a treatment effect for distal social communication outcomes was estimated at only 39%; 95% CI [14%, 67%]. When aggregate dependent variables were excluded, the difference in probability of a treatment effect was 14% favoring proximal outcomes, 95% [7%, 29%].

The probability of a treatment effect varied by boundedness of the outcome

The treatment effect difference was 50%, favoring context-bound dependent variables. This difference was also greater than expected by chance; 95% CI [21%, 77%]. The probability of strong evidence of a treatment effect on potentially context-bound measures of social communication was estimated at 82%; 95% CI [60%, 95%]. In contrast, the probability of a

treatment effect for highly generalized social communication was estimated at only 33%; 95% CI [14%, 60%]. When aggregate dependent variables were excluded, the difference in probability of a treatment effect was 56% favoring potentially context bound outcomes, 95% CI [44%, 75%].

DISCUSSION

This quantitative synthesis evaluated treatment effects on social communication in well-designed studies across several decades of treatment literature involving young children with ASDs. Our findings allow us to draw two primary conclusions from the most rigorous research on treatment of social communication in preschoolers with ASDs. First, the majority of the evidence from well-designed studies suggests that, as a whole, established interventions can impact critical social communication skills in young children with ASDs, even when we exclude variables that exclusively measure instrumental communication or outcomes that include language use that may not be communicative. Second, the likelihood of finding effects of treatment on social communication depends on how social communication is measured.

Effects for social communication vary according to dependent variable type

It is encouraging to find that just over one half (54%) of the social communication outcomes from well-designed studies representing a wide range of treatment approaches (see Table 3) were influenced by treatment. However, it is also clear that we are more likely to conclude that treatments improve social communication in young children with ASDs when we measure change (a) in behaviors that are directly targeted, or (b) in conditions that approximate the intervention context. The differences in the probability of finding a treatment effect were quite striking along both of our dichotomized dimensions. As we suspected, there was a greater likelihood of finding a treatment effect for proximal versus distal outcomes (63% versus 39%, respectively) and for potentially context-bound versus highly generalized variables (82% versus 33%, respectively). It may be tempting to dismiss the results of this review because it might seem we “already knew” that results vary by boundedness and proximity of the dependent variable. Indeed, on some level, most clinicians and researchers probably do have some sense that these factors influence the likelihood of finding an effect for a given treatment. However, to our knowledge, this synthesis provides the first empirical support for our feeling that how we measure change matters. Likewise, it is important that the developmental implications of how change is measured is made clear in future intervention research syntheses.

Implications for research and clinical practice

Thus, these findings have important implications for both research and clinical practice. From a researcher’s standpoint, it should be obvious that treatment effects will more readily be observed for proximal and potentially context-bound variables. However, one should consider that effects on these types of variables provide less convincing evidence of the efficacy of social communication treatments than effects on distal and highly generalized outcomes. Thus, to demonstrate that treatment impacts children’s general propensity to socially communicate, researchers must include assessment procedures that differ from the

treatment procedures on multiple dimensions. Likewise, to truly conclude that treatment has impacted downstream social communication development, researchers must index change in behaviors beyond the treatment targets (i.e., in distal dependent variables).

However, proximal and potentially context bound outcomes still have their place in our research toolkit. Proximal outcomes that are measured in a context that is very similar to, or identical to, the treatment context are more likely to show more rapid and larger changes than are distal or generalized outcomes. Learning and developmental theorists from widely different camps posit that practicing a skill under highly scaffolded conditions can enable independent, generalized use and eventual integration with other social skills (e.g., Skinner, 1953; Vygotsky, 1978). Whether, and to what extent, proximal and context-bound outcomes will lead to distal and generalized outcomes must be empirically tested by using appropriate research designs and not taken as an assumption.

From a clinical perspective, this synthesis provides a useful guide for evaluating a treatment's potential to impact social communication in young children with ASDs. When considering a particular treatment approach, a clinician should reference well-designed studies that have measured effects of the treatment on social communication (Table 3 provides a current, but not necessarily exhaustive, list of the well-designed studies that met our definition for social communication). Clinicians should consider the types of dependent variables included. We recommend placing more confidence on treatments when researchers demonstrate that the effects were seen on aspects of development that extend beyond what was directly taught and were measured in context that differed from the treatment on all of the primary dimensions of stimulus generalization (i.e., setting, communication partners, interaction styles, and materials). Clinicians should then consider whether they can be confident that there was a therapeutic effect of the treatment achieved on social communication outcomes given the evidence provided (Tables 2 and 4 provides criteria for judging evidence of a treatment effect for both single-case and group designs). These considerations will all help the clinician evaluate whether they are likely to achieve distal and highly generalized effects on a child's social communication with a given intervention.

Additional methodological contributions

In addition to illustrating the importance of distinguishing the class of dependent variables on which treatment effects are shown, two other methodological contributions were made. First, we illustrated a logic that allowed a quantitative synthesis of effects across both single-case and group experimental designs. Many have struggled unsuccessfully with this challenge because of the issues surrounding the use of "effect size" indicators for single-case experimental designs (Wolery et al., 2010). Second, we applied a statistical approach that enabled the inclusion of multiple dependent variables from individual studies (clustered bootstrapping). These methods collectively allowed us to consider a much larger amount of the existing evidence for effects of treatment on social communication in young children with ASDs.

Limitations

There are limitations to this review. First, limited resources prevented back and forward searching all selected articles and prevented finding unpublished studies. Thus, this review might not be exhaustive, and the extent to which the pool of potentially relevant studies is representative of the universe of currently available relevant studies is unknown. However, exclusion of studies was conducted strictly by the indicated criteria. And the lack of an exhaustive review does not negate the value of the points made in this review.

Second, postdiscussion consensus was used to ultimately decide which articles were selected. This process has been appropriately criticized because it reduces replicability of the process (Bornstein, Hedges, Higgins, & Rothstein, 2009).

Third, there are caveats to our decision to dichotomize the evidence for a treatment effect as either “high” or “other” (i.e., moderate or under). This “vote counting” procedure ignores treatment effect size beyond that required to be rated as a high effect in group designs. While this approach allowed us to consider both single-case and group design studies using an approach that utilizes the respective design elements to infer a treatment effect, there has been criticism of this approach (Lipsey & Wilson, 2001). Critics point out that (a) small studies are given as much weight as large studies, and (b) treatment effects that some would call trivial are given as much weight as treatment effects that all would consider substantial. Once the field has a useful and comparable effect size indicator for single-case experiments, the latter concern can be addressed. While group experimental effect sizes have confidence intervals that are informed by the sample size of the study, there may continue to be disagreement regarding whether single-case experiments should be given as much weight as group experiments. However, the currently illustrated synthesis approach does attend to the issue of “clinically important effect” by: (a) requiring at least a moderate effect size for large group studies; (b) including only statistically significant effects in small group studies (which will only leave large effects); and (c) distinguishing between effects on dependent variables that vary in the degree to which an effect on development is demonstrated.

Fourth, the same can be said for our dichotomization of the proximity and boundedness of dependent variables. Each construct is probably more accurately represented as a continuum. However, we do not currently have valid and reliable means to place variables along such continua for these dimensions. The current paper provided decision trees for the dichotomous decision as a starting point to making these important distinctions in future quantitative syntheses.

Finally, our review did not uncover a sufficient number of internally valid studies for each type of treatment approach (e.g., naturalistic, developmental, traditional behavioral, other) to evaluate whether the likelihood of finding effects for different types of outcomes (e.g., distal or generalized outcomes) varied by treatment type. This is due, in large part, to the fact that some types of approaches tended to be tested by one specific type of dependent variable, rather than to be represented by a wide range of dependent variable types. For example, effects of traditional behavioral approaches were most often tested by proximal and context-bound dependent variables, and effects of “other” approaches (i.e., not naturalistic, developmental, or traditional behavioral) tended to be represented by distal and generalized

outcomes. Based on the current data, possible conclusions about the best treatment approach for generalized outcomes, for example, would be biased by this covariance of type of treatment with type of dependent variable. Our ability to evaluate the potential for different treatment approaches to impact social communication in ASDs would be bolstered by future research that employs a broader variety of dependent variable types.

CONCLUSION

In conclusion, this synthesis provides encouraging evidence that treatments can have an effect on the most difficult aspects of communication in preschoolers with ASD. However, researchers should be aware that the manner in which they measure social communication will impact not only the likelihood that treatment effects will be detected, but also the degree to which treatment effects are demonstrated to affect social communicative development. Additional work needs to be done to improve our understanding of how interventions impact social communication outcomes in young children with ASDs.

Acknowledgments

Source of funding: This work was supported by training grants from the following: Institute of Educational Sciences [grant number R324B080005]; U.S. Department of Education [grant number ED H325D080075]; and Office of Special Education Programs [grant number H325D100034A].

We would like to acknowledge Rebecca Frantz, George Castle, and Ariel Schwartz for their helpful work on this project.

References

*Studies included in systematic review

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 3. Washington, DC: American Psychiatric Association; 1980.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: American Psychiatric Association; 1994.
- American Psychiatric Association. Diagnostic and statistical manual of psychiatric disorders. 4. Washington, DC: American Psychiatric Association; 2000. text rev
- Beretvas SN, Chung H. An evaluation of modified R^2 -change effect size indices for single-subject experimental designs. *Evidence-Based Communication Assessment and Intervention*. 2008a; 2:120–129.
- Beretvas SN, Chung H. A review of meta-analyses of single-subject experimental designs: Methodological issues and practices. *Evidence- Based Communication Assessment and Intervention*. 2008b; 2:142–151.
- Bornstein, M.; Hedges, L.; Higgins, J.; Rothstein, H. *Introduction to meta-analysis*. Chichester: Wiley; 2009.
- *. Buffington DM, Krantz PJ, McClannahan LE, Poulson CL. Procedures for teaching appropriate gestural communication skills to children with autism. *Journal of Autism and Developmental Disorders*. 1998; 28:535–545. [PubMed: 9932240]
- Bush, HM. *Biostatistics: An applied introduction for the public health practitioner*. Clifton Park, NY: Delmar; 2011.
- Campbell, JM.; Herzinger, CV. Statistics and single subject research methodology. In: Gast, DL., editor. *Single subject research methodology in behavioral sciences*. New York, NY: Routledge; 2010. p. 417-453.

- *. Dawson G, Rogers S, Munson J, Smith M, Winter J, Greenson J, Donaldson A, et al. Randomized, controlled trial of an intervention for toddlers with autism: The early start Denver model. *Pediatrics*. 2010; 125:e17–e23. [PubMed: 19948568]
 - *. Drew A, Baird G, Baron-Cohen S, Cox A, Slonims V, Wheelwright S, Swettenham J, et al. A pilot randomised control trial of a parent training intervention for pre-school children with autism: Preliminary findings and methodological challenges. *European Child & Adolescent Psychiatry*. 2002; 11:266–272. [PubMed: 12541005]
 - *. Dykstra JR, Boyd BA, Watson LR, Crais ER, Baranek GT. The impact of the advancing social communication and play (ASAP) intervention on preschoolers with autism spectrum disorder. *Autism*. 2012; 16:27–44. [PubMed: 21788256]
- Efron B. Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics*. 1981; 9:139–158.
- Fein D, Tindler P, Waterhouse L. Stimulus generalization in autistic and normal children. *Journal of Child Psychology and Psychiatry*. 1979; 20:325–335. [PubMed: 511961]
- Feng Z, McLerran D, Grizzle J. A comparison of statistical methods for clustered data analysis with Gaussian error. *Statistics in Medicine*. 1996; 15:1793–1806. [PubMed: 8870161]
- *. Field T, Lasko D, Mundy P, Henteleff T, et al. Brief report: Autistic children's attentiveness and responsivity improve after touch therapy. *Journal of Autism and Developmental Disorders*. 1997; 27:333–338. [PubMed: 9229263]
- Gast, D. *Single subject research methodology in behavioral sciences*. New York, NY: Routledge; 2010.
- *. Gena A. The effects of prompting and social reinforcement on establishing social interactions with peers during the inclusion of four children with autism in preschool. *International Journal of Psychology*. 2006; 41:541–554.
- Gersten R, Fuchs L, Compton D, Coyne M, Greenwood C, Innocenti M. Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*. 2005; 71:149–164.
- *. Green J, Charman T, McConachie H, Aldred C, Slonims V, Howlin P, Le Couteur A, et al. Parent-mediated communication-focused treatment in children with autism (PACT): A randomised controlled trial. *The Lancet*. 2010; 375:2152–2160.
- Harden JJ. A bootstrap method for conducting statistical inference with clustered data. *State Politics & Policy Quarterly*. 2011; 11(2):223–246.
- Horner, R.; Dunlap, G.; Koegel, R. *Generalization and maintenance: Life-style changes in applied settings*. Baltimore, MD: Paul H. Brookes; 1988.
- Howlin P, Magiati I, Charman TA. Systematic review of early intensive behavioral interventions for children with autism. *American Journal of Intellectual and Developmental Disabilities*. 2009; 114:23–41.
- Hume K, Loftin R, Lantz J. Increasing independence in autism spectrum disorders: A review of three focused interventions. *Journal of Autism and Developmental Disorders*. 2009; 39:1329–1338. [PubMed: 19430897]
- *. Ingersoll B. The differential effect of three naturalistic language interventions on language use in children with autism. *Journal of Positive Behavior Interventions*. 2011; 13:109–118.
 - *. Ingersoll B, Schreibman L. Teaching reciprocal imitation skills to young children with autism using a naturalistic behavioral approach: Effects on language, pretend play, and joint attention. *Journal of Autism and Developmental Disorders*. 2006; 36:487–505. [PubMed: 16568355]
 - *. Kaale A, Smith L, Sponheim E. A randomized controlled trial of preschool-based joint attention intervention for children with autism. *Journal of Child Psychology and Psychiatry*. 2012; 53:97–105. [PubMed: 21883204]
 - *. Kasari C, Freeman S, Paparella T. Joint attention and symbolic play in young children with autism: A randomized controlled intervention study. *Journal of Child Psychology and Psychiatry*. 2006; 47:611–620. [PubMed: 16712638]
 - *. Kasari C, Gulsrud AC, Wong C, Kwon S, Locke J. Randomized controlled caregiver mediated joint engagement intervention for toddlers with autism. *Journal of Autism and Developmental Disorders*. 2010; 40:1045–1056. [PubMed: 20145986]

- *. Kasari C, Paparella T, Freeman S, Jahromi LB. Language outcome in autism: Randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology*. 2008; 76:125–137. [PubMed: 18229990]
- *. Kohler FW, Greteman C, Raschke D, Highnam C. Using a buddy skills package to increase the social interactions between a preschooler with autism and her peers. *Topics in Early Childhood Special Education*. 2007; 27:155–163.
- Kratochwill, TR.; Hitchcock, J.; Horner, RH.; Levin, JR.; Odom, SL.; Rindskopf, DM.; Shadish, WR. Single-case designs technical documentation. 2010. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf
- *. Kroeger KA, Schultz JR, Newsom C. A comparison of two group-delivered social skills programs for young children with autism. *Journal of Autism and Developmental Disorders*. 2007; 37:808–817. [PubMed: 16927011]
- *. Landa RJ, Holman KC, O’Neill AH, Stuart EA. Intervention targeting development of socially synchronous engagement in toddlers with autism spectrum disorder: A randomized controlled trial. *Journal of Child Psychology and Psychiatry*. 2011; 52:13–21. [PubMed: 21126245]
- Lieberman RG, Yoder PJ, Reichow B, Wolery M. Visual analysis of multiple baseline across participants graphs when change is delayed. *School Psychology Quarterly*. 2010; 25:28–44.
- Lipsey, MW.; Wilson, DB. Practical metaanalysis. Thousand Oaks, CA: Sage; 2001.
- Marans, WD.; Rubin, E.; Laurent, A. Addressing social communication skills in individuals with high-functioning autism and Asperger syndrome: Critical priorities in educational programing. In: Volkmar, FR.; Paul, R.; Klin, A.; Cohen, D., editors. *Handbook of autism and pervasive developmental disorders*. 5. Hoboken, NJ: Wiley and Sons; 2005. p. 977-1002.
- *. Martins MP, Harris SL. Teaching children with autism to respond to joint attention initiations. *Child and Family Behavior Therapy*. 2006; 28:51–68.
- McConnell S. 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*. 2002; 32:351–372. [PubMed: 12463515]
- *. McGee GG, Almeida MC, Sulzer-Alzaroff B, Feldman RS. Promoting reciprocal interactions via peer incidental teaching. *Journal of Applied Behavior Analysis*. 1992; 25:117–126. [PubMed: 1582961]
- Millar DC, Light JC, Schlosser RW. The impact of augmentative and alternative communication intervention on the speech production of individuals with developmental disabilities: A research review. *Journal of Speech, Language & Hearing Research*. 2006; 49(2):248–264.10.1044/1092-4388(2006/021
- *. Odom SL, Strain PS. A comparison of peer-initiation and teacher-antecedent interventions for promoting reciprocal social interaction of autistic preschoolers. *Journal of Applied Behavior Analysis*. 1986; 19:59–71. [PubMed: 3710949]
- *. Odom SL, Watts E. Reducing teacher prompts in peer-initiation interventions through visual feedback and correspondence training. *Journal of Special Education*. 1991; 25:26–43.
- Prizant, BM.; Wetherby, AM. Critical issues in enhancing communication abilities for persons with autism spectrum disorders. In: Volkmar, FR.; Paul, R.; Klin, A.; Cohen, D., editors. *Handbook of autism and pervasive developmental disorders*. 5. Hoboken, NJ: Wiley and Sons; 2005. p. 925-946.
- Reichow B, Volkmar FR. Social skills interventions for individuals with autism: Evaluation for evidence-based practices within a best evidence synthesis framework. *Journal of Autism and Developmental Disorders*. 2010; 40:149–166. [PubMed: 19655240]
- Reichow B, Volkmar FR, Cicchetti DV. Development of the evaluative method for evaluating and determining evidence-based practices in autism. *Journal of Autism and Developmental Disorders*. 2008; 38:1311–1319. [PubMed: 18095149]
- Riley RD, Thompson JR, Abrams KR. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*. 2008; 9:172–186. [PubMed: 17626226]
- *. Roberts J, Williams K, Carter M, Evans D, Parmenter T, Silove N, Clark T, et al. A randomised controlled trial of two early intervention programs for young children with autism: Centre-based

with parent program and home-based. *Research in Autism Spectrum Disorders*. 2011; 5:1553–1566.

Schunk, D. *Learning theories: An educational perspective*. 4. Upper Saddle River, NJ: Pearson; 2004.

Shumway S, Wetherby A. Communicative acts of children with autism spectrum disorders in the second year of life. *Journal of Speech, Language, and Hearing Research*. 2009; 52:1139–1156.

Skinner, BF. *Science and human behavior*. New York, NY: Macmillan; 1953.

Slavin RE. Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*. 1986; 15:5–11.

*. Smith T, Groen AD, Wynn JW. Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal of Mental Retardation*. 2000; 105:269–285. [PubMed: 10934569]

*. Strain PS, Danko CD, Kohler FW. Activity engagement and social interaction development in young children with autism: An examination of “free” intervention effects. *Journal of Emotional and Behavioral Disorders*. 1995; 3:108–123.

*. Vismara LA, Lyons GL. Using perseverative interests to elicit joint attention behaviors in young children with autism: Theoretical and clinical implications for understanding motivation. *Journal of Positive Behavior Interventions*. 2007; 9:214–228.

Volkmar F, Lord C, Bailey A, Schultz R, Klin A. Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry*. 2004; 45:135–170. [PubMed: 14959806]

Vygotsky, LS. *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press; 1978.

Wetherby, AM. Understanding and measuring social communication in children with autism spectrum disorders. In: Charman, T.; Stone, W., editors. *Social and communication development in autism spectrum disorders: Early identification, diagnosis, and intervention*. New York, NY: The Guilford Press; 2006. p. 3-34.

Wolery M, Busick M, Reichow B, Barton EE. Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education*. 2010; 44:18–28.

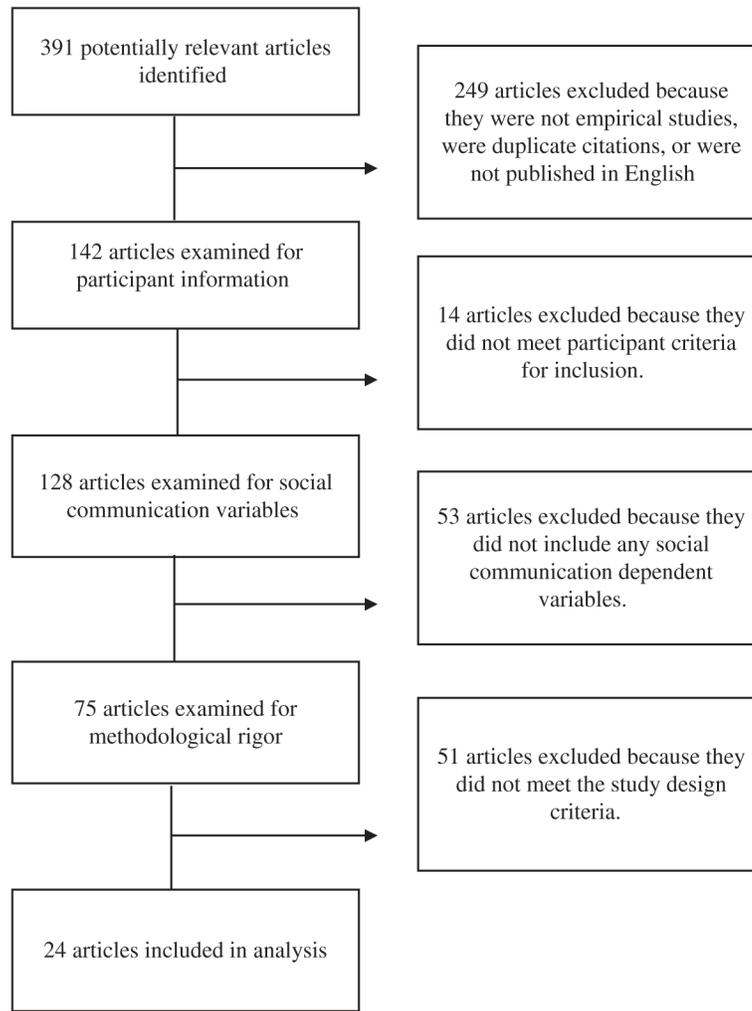


Figure 1.
Study exclusion process.

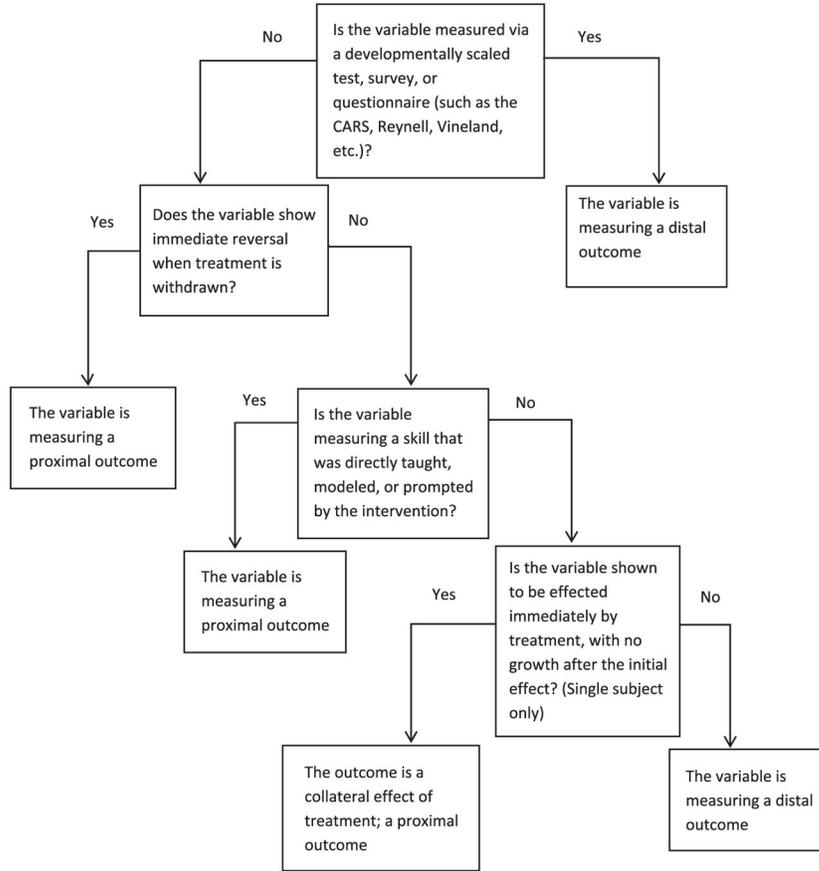


Figure 2. Variable proximity decision tree.

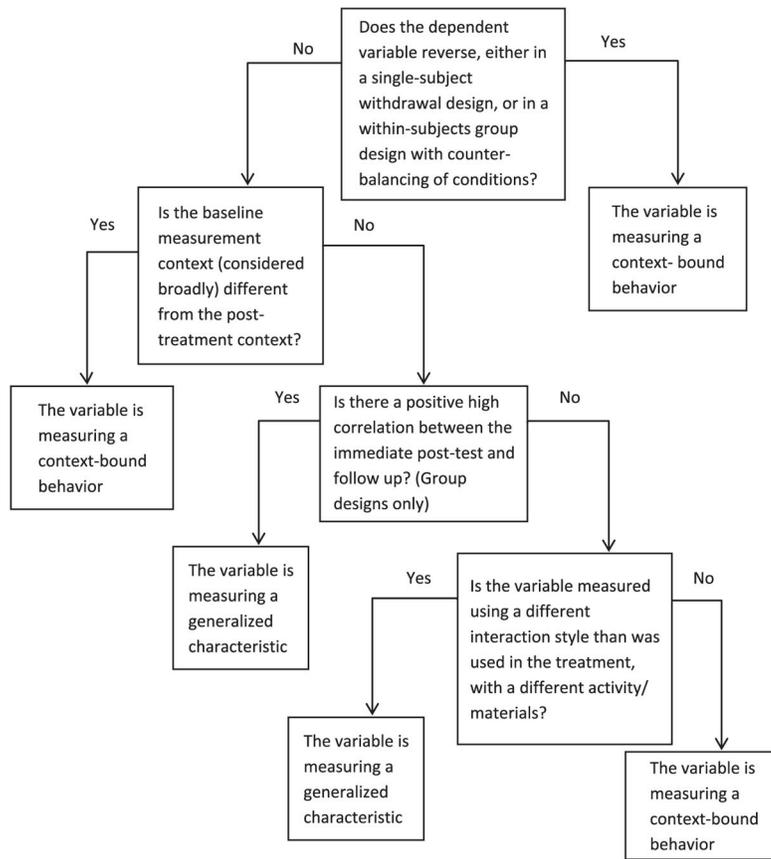


Figure 3. Variable boundedness decision tree.

Table 1

Matrix of dependent variable types

	Proximal	Distal
Context bound	Variables that are directly taught by the intervention and are only experimentally measured within the intervention or similar context. Example: Responding to joint attention measured within an intervention that targets responding to joint attention.	Variables that are not directly taught by the intervention and are only measured within the intervention or similar context. Example: Initiating joint attention for declaratives measured within an intervention session that targets responding to joint attention.
Generalized	Variables that are directly taught by the intervention and are measured in a context other than the intervention context. Example: Responding to joint attention measured in a parent–child session after a researcher-led intervention that targets responding to joint attention.	Social communication that is not directly taught by the intervention and is measured in a context other than the intervention context. Example: Initiating joint attention for declaratives measured in a parent–child session using materials not used in treatment after a researcher-led intervention that targets responding to joint attention.

Table 2

Basis for judging strong evidence of a treatment effect for single-case studies

Criteria	Rationale	Source
Must show evidence of reliability among coders, measured across phases. Gross agreement must be $>.9$ OR point by point agreement must be $>.8$ OR Cohen's kappa is $>.6$ OR coders are blind to phase	Protects against researcher bias favoring the treatment group in outcome scores	Kratochwill et al., 2010
Must have a minimum of 3 data points per phase	Allows for a reasonably accurate visual analysis of intervention effects (e.g., variability can be observed with a minimum of 3 data points)	Kratochwill et al., 2010
Design must include >3 planned replications of a basic effect	Provides adequate opportunity to determine if effects are due to the treatment and not to chance occurrence or the effect of an unobserved variable	Kratochwill et al., 2010
Must employ variant-specific design elements: Multiple baseline <ul style="list-style-type: none"> • The baseline trend is generally flat or countertherapeutic, or there is an immediate, large shift in level or trend • Treatment phase begins after change in dependent variable is seen in immediately preceding tier • >3 tiers (do not consider repetitions with <3) • Withdrawal of treatment • >4 phases/case (don't consider cases with <4) • >3 adjacent phase replications • Alternating treatments • Context-bound dependent variables are functionally independent • Goals in each condition are equally difficult relative to participant's readiness to learn (adapted only) • A baseline is present in designs where treatments may have interfered with one another 	Ensures that the research design is implemented adequately so that there is a possibility of demonstrating a functional relationship between the treatment and dependent variable	Kratochwill et al., 2010

Table 3

Study and variable information

Citation	Participant means (ages in months)				Design	Intervention name	Dependent variables	Variable type		
	N	CA	IQ/MA	ELA				Proximity	Boundedness	Effect rating
Buffington et al. (1998)	2	53	NR	32	SCD-MBL	Behavior gestural communication training	Percentage of trials with appropriate gestural and verbal response for attention, affect, and reference	P	CB	High
Dawson et al. (2010)	48	23.5	60.25/NR	25.3	RCT	ESDM	VABS Socialization subscale at Year 1 ^a	D	GC	Other
Drew et al. (2002)	24	22.5	78.55/NR	NR	RCT	Social pragmatic parent training program	VABS Socialization subscale score at Year 2 ^a	D	GC	Other
Dykstra et al. (2012)	3	50.3	69.1/NR	32.5	SCD-MBL	ASAP	ADJ-RSI subscale score	D	GC	Other
Field et al. (1997)	22	54	91/NR	16	RCT	Touch therapy	Percentage of intervals with interaction, IBR and IJA ^a	P	CB	Other
Gena (2006)	4	49.75	NR	NR	SCD-MBL	Classroom-based social behavior intervention	ESCS Joint Attention subscale change score	D	GC	High
Green et al. (2010)	152	45	NR	NR	RCT	PACT	ESCS Social subscale change score	D	GC	High
Ingersoll (2011)	2	41	NR/28	23.5	SCD-ATD	Responsive interaction/milieu teaching	ESCS Initiating subscale change score ^a	D	GC	High
Ingersoll & Schreibman (2006)	5	36.6	NR/19.8	21.3	SCD-MBL	Reciprocal imitation therapy	Number of initiations to peers, ^a P CB High Percentage of responses to peer initiations	P	CB	High
							ADOS-G Social Communication Algorithm total score	D	GC	Other
							Number of child initiations in parent-child interaction ^b	P	CB	Other
							CSBS-DP social composite change score	D	GC	High
							Percentage of 10-s intervals with spontaneous language for initiation ^a	P	CB	Other
							Percentage of intervals with CJA	P	CB	High

Citation	Participant means (ages in months)				Intervention name	Dependent variables	Variable type			
	N	CA	IQ/MA	ELA			Design	Proximity	Boundedness	Effect rating
Kasari et al. (2012)	61	48.8	NR/27.7	21.5	RCT	Joint attention intervention for caregivers	Frequency of IJA during ESCS	P	GC	Other
							Frequency of IJA during teacher-child play	P	CB	High
							Frequency of IJA during mother-child play	P	GC	Other
Kasari et al. (2006)	58	42.6	NR/24.4	20.6	RCT	Joint attention intervention	Frequency of shows during ESCS	P	GC	High
							Frequency of CJA during ESCS	P	GC	Other
							Frequency of pointing during ESCS	P	GC	Other
							Frequency of gives during ESCS	P	GC	Other
							Frequency of RJA during ESCS	P	GC	High
							Frequency of CJA during mother-child interaction	P	GC	High
							Frequency of points during mother-child interaction	P	GC	Other
							Frequency of gives during mother-child interaction	P	GC	Other
							Frequency of shows during mother-child interaction	P	GC	Other
							Frequency of shows during ESCS	D	GC	High
							Frequency of CJA during ESCS	D	GC	Other
Play intervention							Frequency of pointing during ESCS	D	GC	Other
							Frequency of gives during ESCS	D	GC	Other
							Frequency of RJA during ESCS	D	GC	Other
							Frequency of CJA during mother-child interaction	D	GC	High
							Frequency of points during mother-child interaction	D	GC	Other
							Frequency of gives during mother-child interaction	D	GC	Other
							Frequency of shows during mother-child interaction	D	GC	Other
							Frequency of shows during ESCS	D	GC	High
							Frequency of CJA during ESCS	D	GC	Other
							Frequency of gives during ESCS	D	GC	Other
							Frequency of RJA during ESCS	D	GC	Other

Citation	Participant means (ages in months)				Intervention name	Design	Variable type			
	N	CA	IQ/MA	ELA			Dependent variables	Proximity	Boundedness	Effect rating
Kasari et al. (2010)	38	30.8	NR/19.2	NR	Joint attention intervention for caregivers	RCT	Frequency of IJA during child-caregiver interaction	P	CB	Other
Kasari et al. (2008)	46	42.6	NR/24.4	20.6	Joint attention intervention	RCT	Frequency of RJA during child-caregiver interaction Average frequency of IJA in ESCS and mother-child interaction at follow-up Frequency of RJA in ESCS at followup	P P P	CB GC GC	High High Other
Kohler et al. (2007)	1	57	NR/NR	NR	Play intervention	SCD-MBL	Buddy skills package for typical peers	P	CB	High
Kroeger et al. (2007)	25	63.3	NR/NR	NR	Direct teaching of social skills group	RCT	Number of initiations of social interaction with peers ^a	P	CB	High
Landa et al. (2011)	48	28.8	NR/NR	NR	Interpersonal synchrony	RCT	Number of responses to social interactions with peers Interacting behaviors ^a IJA in CSBS	P P P	CB CB GC	High High Other
Martins & Harris (2006)	3	52.3	47.2/NR	NR	Joint attention training	SCD-MBL	Shared positive affect in CSBS	P	GC	Other
McGee et al. (1992)	3	54.3	NR/NR	NR	Peer incidental teaching	SCD-MBL	Percentage of responses to joint attention Percent of intervals with positive reciprocal interaction	P P	CB CB	High High
Odom & Strain (1986)	3	48	NR/NR	NR	Teacher antecedent	SCD-ATD	Mean length of interactions with peer ^a	P	CB	High
Odom & Watts (1991)	3	54	NR/NR	19	Peer initiation CTVF	SCD-ABAB	Total positive initiations to peer ^a Total positive responses to peer Frequency of positive social interactions in CTVF sessions ^a Frequency of positive social interactions in peer initiation sessions ^a	P P P P	CB CB CB CB	High High High High

Citation	Participant means (ages in months)				Intervention name	Dependent variables	Variable type		
	N	CA	IQ/MA	ELA			Design	Proximity	Boundedness
Roberts et al. (2011)	82	42	NR/NR	NR	Building blocks program	VABS Socialization subscale ^a	D	GC	High
Smith et al. (2000)	28	35.9	12.43/21.7	15.7	Intensive early intervention	The Pragmatics Profile VABS Socialization subscale score ^a	D	GC	Other Other
Strain et al. (1995)	4	54	NR/NR	NR	Active engagement & social engagement intervention	Percentage of intervals with social interaction ^a	P	CB	High
Vismara & Lyons (2007)	3	32.6	NR/NR	NR	Pivotal response treatment	Frequency of IJA	P	CB	High

Note. ABAB = withdrawal design; ADI-RSI = Autism Diagnostic Interview–Reciprocal Social Interaction; ADOS-G = Autism Diagnostic Observation Scales–General; ASAP = Advanced social communication and play; ATD = alternating treatment design; CA = chronological age; CB = context-bound; CJA = coordinated joint attention; CSBS-DP = Communication and Symbolic Behavior Scales–Developmental Profile; CTVF = correspondence training/visual feedback; D = distal; ELA = expressive language age; ESCS = Early Social Communication Scales; ESDM = early start Denver model; IBR = initiating behavioral request; IJA = initiating joint attention; GC = generalized characteristic; MA = mental age; MBL = multiple baseline; NR = not reported; P = proximal; PACT = Preschool Autism Communication Trial; RCT = randomized controlled trial; RJA = responding to joint attention; SCD = single-case design; VABS = Vineland Adaptive Behavior Scales.

^aVariable was an aggregate that included intentional forms of communication other than social communication. Additional analyses were conducted with these variables excluded. Conceptually, results were the same.

Table 4

Basis for judging strong evidence of a treatment effect for group design studies

	Rationale	Source
Subjects must be randomly assigned to groups of >10 subjects, with at least three variables thought to correlate with the outcome shown to have no significant differences between groups at pretreatment	Provides evidence of equality between control and treatment groups	Gersten et al., 2005
Attrition must be under 20% or intent to treat analysis must be used	Ensures that events that occur after treatment assignment are unlikely to affect between group comparability on all relevant variables	Gersten et al., 2005
The unit of randomization must be the unit of analysis	Ensures that statistical tests match design features	Gersten et al., 2005
If an observational measure was used, the coders must be blind to group assignment	Protects against researcher bias favoring the treatment group in outcome scores	Gersten et al., 2005