J Autism Dev Disord
DOI 10.1007/s10803-009-0728-1

ORIGINAL PAPER

Group Intervention to Promote Social Skills in School-age Children with Pervasive Developmental Disorders: Reconsidering Efficacy

Kathleen Koenig · Andres De Los Reyes · Domenic Cicchetti · Lawrence Scahill · Ami Klin

Abstract A consistent result in the evaluation of group-delivered intervention to promote social reciprocity in children with PDDs is that outcome data are inconclusive. Lack of robust evidence of efficacy confounds understanding of these interventions and their value to the field. It is conceivable that the construct of impaired social reciprocity in PDD presents unique circumstances that require special consideration when evaluating the evidence base. Social reciprocity and impairment in social functioning are complex constructs, which require a multi-dimensional, multi-method approach to intervention and measurement of gains. The existing paradigm for evaluating the evidence base of intervention may need modification to permit a more intricate analysis of the extant research, and increase the sophistication of future research.

Keywords Autism · Pervasive developmental disorders · Social skills training · Evidence-based treatment · Range of changes · Group therapy

Despite the explosion of recent research testing group interventions to develop social reciprocity skills in school-age children with pervasive developmental disorders (PDDs) including autism, Asperger's Disorder and pervasive developmental disorder—not otherwise specified, progress in the field seems to be one step forward, two steps back. A myriad of group intervention approaches have been tested yet efficacy studies do not yield consistent, robust results (Matson et al. 2007; Rao et al. 2008; White et al. 2007). This state of affairs is particularly problematic because group interventions to promote social skills (commonly known as "social skills training") are ubiquitous in schools and community settings. The lack of firm evidence of efficacy raises the question of whether family and community resources are being used to the best advantage.

With regard to the lack of consistency in outcome, the heterogeneity of the population under study and methodological differences are clearly sources of inconsistency (Lord et al. 2005). Accordingly, if inconsistency in outcome is related to lack of knowledge regarding strategies that work well for teaching skills in this heterogeneous population, then the obvious course is to design and test new strategies. If inconsistency can be attributed to error variance, then greater rigor in research design, execution and data analysis is required. These explanations for the conflicting evidence regarding the efficacy of social skills training (SST) have been identified (Smith et al. 2007; Lord et al. 2005), but perhaps are not the only factors involved.

A key focus of SST intervention is to employ psychological principles to enhance social reciprocity. Inconsistent and inconclusive outcomes may be the result of a *disconnect* in how the construct of impaired social reciprocity in children with PDDs is conceptualized, how change associated with development and intervention is measured, and whether and what changes are observed. In this regard, research on the efficacy of SST for children with PDDs must include acknowledgment of the wide variability in presentation of the impairment in social

K. Koenig (✉) · D. Cicchetti · L. Scahill · A. Klin
Yale Child Study Center, Yale University School of Medicine, 230 South Frontage Road, P.O. Box 207900, New Haven, CT 06520-7900, USA
e-mail: kathy.koenig@yale.edu

A. De Los Reyes
Department of Psychology, University of Maryland, College Park, MD, USA

Published online: 27 March 2009

reciprocity as well as the variability likely to be encountered in measures of change associated with intervention. Attention to the complexity of research design when evaluating intervention delivered in a group format, and the need to consider clinical significance as well as statistical significance are critical to the research process as well. Without attention to these issues, future research on the efficacy of group intervention for teaching social skills to children with PDDs is not likely to provide greater clarity.

The purpose of this paper is to highlight important issues in the evaluation of *group delivered social skills training* (SST) programs for children with PDDs and the interpretation of their outcomes. First, we review the current evidentiary support for SST, including specific strategies and more broad-based intervention approaches and comment on the status of their evidence base. Next, we highlight crucial observations related to inconsistent outcomes identified in well designed and executed studies of SST for children with PDDs and how these inconsistencies are currently interpreted. We discuss recent theoretical and empirical work in the broader intervention literature that sheds light on new ways of both interpreting the inconsistent findings within and between studies and on the testing of psychosocial interventions delivered in a group format. Finally, we apply this work to the study of group SST interventions for PDDs, and consider how current methods for the evaluation of this intervention fit or do not fit the problem. In doing so, we highlight directions for future research and theory on the development and testing of SST interventions for children diagnosed with a PDD.

Intervention for Promoting the Development of Social Skills

Social skills training is a well-established intervention for children with different kinds of social impairments, for example, oppositional behavior or attention deficit hyperactivity disorder (Spence 2003; Quinn et al. 1999; Schneider 1992). Meta-analyses of SST studies show that effect sizes vary widely, from $d = 0.01$ to $d = 1.20$ (with a mean of $\sim d = 0.45$), depending on the population studied, the specific targets of intervention, and the methods used to measure outcome (Beelman et al. 1994). Larger effect sizes are found for intervention focused on specific skills compared to global interventions (Quinn et al. 1999). Further, prior work suggests a reduction in effect size, as time elapsed between treatment administration and outcome measurement (Schneider 1992). Nevertheless, a consensus among researchers and clinicians has emerged supporting the view that small effects can be of clinical importance, and thus, the intervention is worth continuing study and refinement (Lipsey and Wilson 1993; Spence 2003).

Social Skills Training in PDD

Adapting SST methods to children with PDDs is a particularly vexing problem because the social impairment in PDD is *qualitatively* different from that encountered in other childhood psychiatric disorders (Rao et al. 2008). Often, a distinction is made between acquisition deficits (the child has not acquired the skill) or performance deficits (the child possesses the skill but does not apply it appropriately), and treatment methods are guided by this distinction (Spence 2003). For example, a child may know how to brush her teeth (the skill has been acquired), but never do so unless prompted (a deficit in performance). In PDDs, acquisition and performance deficits in social skills are compounded over the course of development. An affected child presents with an intricate set of interpersonal difficulties based not just on current functional limitations but also on the paucity of a rich historical knowledge regarding relationships and their complexities. The impairment is further complicated by multiple factors, including the child's level of cognitive functioning, the presence of behavioral rigidity, the presence of anxiety or other co-morbid conditions, the degree of receptive and expressive language impairment, and the degree and severity of stereotypic or repetitive behaviors (Volkmar et al. 2005). Features of social functioning deemed affected in PDD include: (1) impairment in the ability to identify faces and facial expressions (Schultz et al. 2003); (2) impairment in understanding prosodic elements of speech, the nuances of language, the pragmatics of communication and the interpretation of gesture (Tager-Flusberg et al. 2003); (3) poor interpretation of contextual elements of the social environment (Klin 2000); (4) poor ability to regulate affect (Konstantareas and Stewart 2006); (5) poor insight into the emotional components of relationships (Begeer et al. 2008); (6) lack of ability to take the perspective of the other (Baron-Cohen et al. 2000); and (7) poor ability to self-monitor behavior (Koegel and Koegel 1995). Further, a hallmark of effective social functioning noticeably absent in PDD is the *fluid application of one's knowledge and behavior to reciprocal interaction with others* (Klin et al. 2003). This requires a sense of timing and rapid recognition of what is emotionally salient in a particular context, two elements of social interaction that are not well understood. Moreover, lack of motivation to learn these skills—a frequent but not inevitable component of the clinical picture—adds another layer of complexity to the process of intervention and measurement of gains (Koegel and Koegel 1995). While individual skills can, of course, be delineated as discrete targets for intervention and measured with reasonable accuracy, testing the efficacy of a broad based group intervention is much more daunting task.

Table 1 Quality indicators in clinical trials on psychosocial intervention for individuals with ASD

1. Random assignment of participants to intervention and control groups
2. Manuals for all groups
3. A recruitment plan to obtain a representative sample
4. Clearly stated inclusion and exclusion criteria
5. Careful characterization of participants at entry into the study (e.g. diagnosis, symptom severity, and level of functioning)
6. Systematic monitoring of intervention fidelity
7. Clear rationale for the choice of outcome measures and, especially in studies of comprehensive intervention packages, inclusion of measures that assess core features of autism such as reciprocal social interaction.
8. Use of outcome measures collected blind to intervention group.
9. Appropriate statistical analysis of differences between groups after intervention, effect size and clinical significance of differences, and variables that may influence outcomes (i.e. mediators and moderators)

Reprinted from Smith et al. (2007)

With regard to understanding how well the current standards for examining the evidence based of group delivered SST work, it is informative to consider the work of the National Institute of Mental Health (NIMH) task force charged with establishing guidelines for the conduct and evaluation of efficacy research in psychosocial intervention for PDDs (Smith et al. 2007). Based on discussions regarding the state of the science and the available methodology for evaluating intervention research, this task force targeted the methodological challenges, gaps in evidence-based treatments, and the dearth of reliable information about long-term outcomes (Smith et al. 2007). Recommendations were that specific strategies and techniques be established as effective using case studies and case series; strategies should be manualized in order to promote consistent application of the intervention by all researchers; clinical trials should be conducted incorporating certain critical components (Table 1) and ultimately outcomes should be assessed in community settings.

Research Outcomes

White et al. (2007) used the recommendations from the NIMH working group as a framework for reviewing the extant research on SST for children with PDDs. Briefly, an examination of case studies and case series showed that direct instruction using behavioral (operant) strategies increased social behaviors such as greeting others, making eye contact and responding to others. The most successful interventions used multiple training sessions, modeling, and naturally occurring and direct reinforcement. Strong support for these approaches is not surprising in that these results are entirely consistent with the extremely robust literature on behaviorally based intervention in PDDs (Schriebman 2000).

Moving to the evaluation of research on *group delivered* SST, the focus of this paper, those interventions that incorporated strategies and techniques established as effective in case studies and case series proved to be effective. These included a cognitive-behavioral approach, role-play and practice, the use of peer models and structured methods for teaching and reinforcement (Bauminger 2002; Barry et al. 2003; Koenig et al. in press; Morrison et al. 2001; Tse et al. 2007; Yang et al. 2003; Ozonoff and Miller; 1995).

On the whole, studies evaluating group delivered SST were difficult to compare and synthesize because of varying experimental designs and methods of measurement. Sample sizes ranged from 3 to 45 children, with an approximate mean of ten subjects and an approximate median of nine subjects per study. Different approaches to intervention were used; length of treatment varied from several weeks to a several months. Only three studies employed a treatment manual (Koenig et al. in press; Yang et al. 2003; Tse et al. 2007). Some studies used a comparison group and others used pre- and post-treatment measures to assess outcome (Solomon et al. 2004; Provencal 2003; Barnhill et al. 2002). Additionally, different informants were used to provide outcome data across studies, including teachers, parents, direct observers or the self-report of the affected child. Overall, outcomes varied widely and presented a very mixed picture as to the efficacy of SST (White et al. 2007).

Critically important is that inconsistent outcomes were noted within studies as well as between studies. For example, Koenig et al. (in press) showed improvement on a global measure of social functioning based on parent report, but less impressive improvement on a parent reported standardized questionnaire targeting pro-social behavior and social initiative. Webb et al. (2004) described significant improvement in four of five specific skills taught in a structured intervention, but no concomitant improvement in social skills broadly measured. In a study of 18 boys with PDDs, significant improvement in face recognition skills but no improvement in perspective-taking skills was

observed (Solomon et al. 2004). Ozonoff and Miller (1995) noted significant improvement in participants' ability to use perspective-taking skills, but no concurrent improvement in social skills on a valid, reliable questionnaire. While not flawless, each of these studies was designed and executed well enough, so that inconsistent results can not be attributed wholly to random or measurement error.

A second issue in the evaluation of this body of research is that the targets for intervention ranged from very specific behaviors to global improvement in social functioning. Barry et al. (2003) focused specifically on greeting and conversation, while Bauminger (2002) implemented a broader curriculum, emphasizing social cognition, emotional understanding and social interaction skills. Tse et al. (2007) focused on improving eye contact and listening skills, and behaving courteously, while Golan and Baron-Cohen (2006) targeted the recognition of emotion in others. These studies and others (MacKay et al. 2007; Sansosti and Powell-Smith 2006) have been included in prior reviews of the efficacy of SST. That said, the targets for intervention vary widely across studies, limiting the conclusions that can be derived from meta-analyses of these studies for the purpose of deriving the mean effect of treatment.[1]

Understanding Variability in Outcome

In the psychological sciences, inconsistent results in outcome measurement following intervention research are consistently noted (Achenbach 2006; De Los Reyes and Kazdin 2005). A possible source of variation is that *multidimensional, psychological constructs present unique challenges for intervention, intervention research and the evaluation of outcomes* (De Los Reyes and Kazdin 2006). In group SST for children with PDDs, the focus of treatment is enhancing social reciprocity, an extraordinarily complex, multi-dimensional construct. This target for intervention requires comprehensive evaluation at a level rarely observed in the empirical literature. Further, complicating the interpretation of clinical outcomes are the substantial differences between characteristics of PDD and other childhood conditions currently treated through established, evidence-based treatments. Specifically, current models of intervention research for childhood medical and psychiatric disorders are well established, based on methods for improving or eradicating symptoms of illness, rather than on the remediation of developmental deficits. In the assessment of an intervention's efficacy for treating, for example, a psychiatric disorder in a previously healthy child, the characteristics of illness can be clearly described so that improvement is readily recognized. For example, a sign of depression, such as irritability or oppositional behavior, can be contrasted with a period of time in the child's life when this behavior was not present or with normative data from same-age peers (Jacobson and Truax 1991). Thus, improvement in the clinical picture can be measured fairly unambiguously. This is not to suggest that intervention research for these kinds of difficulties is not complex or without challenges. Indeed, the sophistication of this body of research has resulted in the well-articulated methods for establishing what constitutes evidence-based treatment (Chambless and Ollendick 2001) in many circumstances. However, the evaluation of intervention for the remediation of difficulties associated with the PDDs and their complex sequelae is by no means straightforward. The dynamic nature of these disorders and the multiple and diverse outcomes possible with intervention must be considered. The complexity of conceptualizing change and identifying beneficial outcome is challenging for most psychological constructs (De Los Reyes and Kazdin 2006) and extraordinarily challenging for intervention to promote social reciprocity in children with PDDs.

A Range of Possible Changes in Outcome

Using two extremely well-designed, rigorous studies of cognitive-behavioral treatment for anxiety as exemplars, De Los Reyes and Kazdin (2006) illustrated that measurement of the same construct with the same instrument between intervention studies can result in different outcomes (significant versus non-significant); measures of the same construct within and between studies using different instruments may also result in widely different outcomes. Based on this work and a review of other studies of the evidence base of particular treatments, the authors conclude that varying outcomes are likely to occur when the target of psychosocial intervention is a complex and multidimensional construct. Further, "a range of possible changes" may occur as a result of intervention (De Los Reyes and Kazdin 2006). With regard to SST in children with PDDs the essential point is that the complex, multidimensional construct of social reciprocity is a dynamic entity, buffeted by a myriad of different factors at different times and in different contexts. This construct cannot be treated as an unmoving, stable and static target for intervention. Thus, when considering social reciprocity as a target for intervention for children with PDDs, the complexity of the construct should be acknowledged, a priori, as well as the likelihood that change post-intervention will be *variable*, not uniform. As mentioned previously, social reciprocity includes, but is not limited to, the ability to read

[1] Meta-analyses of *methods* might yield useful information regarding variability in outcome as it is linked to specific methods for measuring intervention effects.

facial expressions and interpret nonverbal communication, appropriate use of gaze and gesture, verbal fluency and appropriate prosody, adept pragmatic speech, interpersonal and intra-personal insight, self-awareness, and the capacity for behavioral regulation. Moreover, the change associated with intervention will be variable, involving multiple influences. Delay and deviance in social reciprocity skills, central characteristics of the PDDs, take varied forms over time, reflecting the dynamic nature of development. Predictive models and outcome measurement must incorporate estimates of growth as well as stable and time-specific effects (Curran and Bollan 2001). Accordingly, outcome ought not to be measured assuming one is targeting a unidimensional entity when the target is conceptualized as multi-dimensional.

These complexities have major implications for the design and evaluation of group delivered interventions developed to remediate problems associated with impaired social reciprocity. For example, the common practice of choosing a primary outcome measure for a particular study necessarily reduces the target of intervention to a single dimension, in the sense that outcome is limited to one source or one method of measurement. While the demand that intervention research select a primary outcome measure was made to force researchers to specify a priori an indicator representing the extent to which an evaluated intervention changes a target behavior, a laudable objective, there are limitations to this methodological strategy. The approach can mislead investigators and consumers to the false assumption that a single indicator can capture change in the targeted construct in its entirety. Specifically, investigators cannot capitalize on the utility of inconsistent information for informing whether and how interventions work (De Los Reyes and Kazdin 2006).

An additional point with regard to study design is that estimates of effect size for the calculation of sample size will depend on the particular aspect of the construct under study, as well as the direction and magnitude of change predicted as a result of intervention; in other words, the possible *range of changes* that may occur. Estimates of possible effects and the magnitude of these effects must be considered a priori. By examining these issues systematically, one can begin to tease out not only whether an intervention is effective, but when, where, and for whom. Indeed, one ought not to expect uniform estimates of effect size across multiple outcome measures because disparate skills comprise the broad construct of social reciprocity. Further, conventional assessments of the magnitude of effects (e.g. 0.20 = small effect) may not be useful (Lipsey and Wilson 1993). When considering the complexity and subtlety of some deficits in social reciprocity skills, a small treatment effect could make a real difference to the individual. For example, unusual prosody sets children with PDDs apart from the peers during social interaction (Paul et al. 2005). Incremental improvement in prosodic skill might conceivably make the difference between peer acceptance and peer rejection.

Attending to and interpreting inconsistent findings in the examination of group intervention to promote social skills in children with PDDs involves considering the unique nature of this problem. The complexity of this construct is exemplified by the many factors that influence it, including (but not limited to) social perception, social communication skills, the presence of co-morbid conditions such as obsessive-compulsive disorder, the degree of cognitive impairment and learning disability, the degree of social motivation, and the level of competency with regard to adaptive living skills (Volkmar et al. 2005). The outcome of an intervention might change any one, or any combination of these factors (in a positive or negative direction), although the magnitude of the effect will vary based on the particular factors that are the focus of intervention, as well as the measures used to gauge improvement (and their varying sources; Fig. 1). For example, an increase in social motivation (considered a positive outcome) might cause a concomitant increase in anxiety (a negative outcome). Improved ability to recognize emotion in others may lead to greater empathy and social motivation, but increase the frequency of inappropriate social communication. Moreover, the hypothesis or underlying assumption that substantive change in social behavior within a limited range (for example, social perception) will then result in global changes in social reciprocity, may simply be untenable.

An additional and critical point is that there may be limited utility in expecting uniform improvement following intervention in a complex psychological construct such as social reciprocity (De Los Reyes and Kazdin 2006). Traditionally, intervention researchers base their evaluation of a study on the assumption that if an intervention is truly effective, all indices used to measure outcome will move in an expected, positive direction and inconsistency between outcome measures can be attributed to error variance or lack of efficacy of the treatment (Smith et al. 2007). The expectation of uniform improvement across measures and studies in response to intervention is consistent with the way clinical trials are evaluated, particularly in psychopharmacology research (Lipsey and Wilson 1993; Wang and Bakhai 2006). Nonetheless, in the evaluation of group SST, the premise of uniform improvement on all measures needs examination given the construct targeted for intervention.

With regard to psychosocial intervention in general, reports from different informants regarding a child's emotional or behavioral difficulties are most often *not* highly correlated, with correlations (r values) ranging from

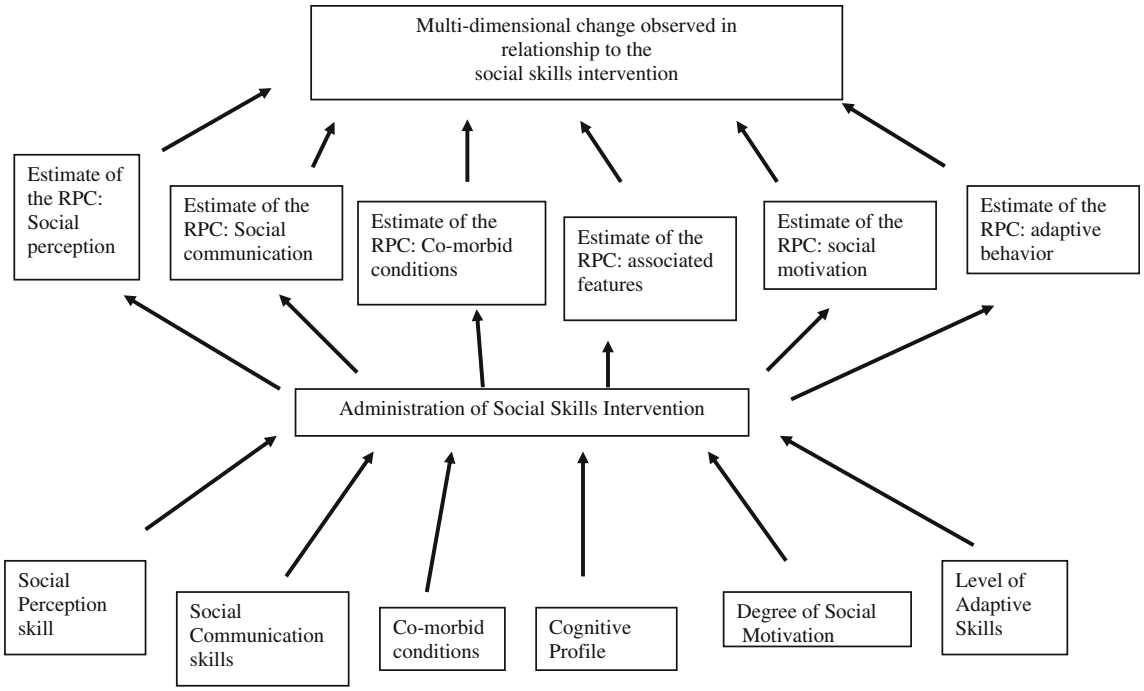

Fig. 1 Schematic of range of possible changes (*RPC*) in the construct of social reciprocity as a result of intervention

0.20 to 0.60 (Achenbach 2006; De Los Reyes and Kazdin 2005). In this regard, variability is due to the perspective of the informant, his or her perception of the child's difficulties, and the context within which the informant observes the child (Kraemer et al. 2003). Further, the difference between cross-informant reliability versus inter-rater reliability is pertinent. Cross informant correlation reflects ratings of behavior based on varying perspectives, contexts, and roles vis-à-vis the child. In contrast, inter-rater reliability represents correlation based on informant ratings of concurrent observed behavior (Achenbach 2006). The assumption that baseline ratings or outcomes from different sources should yield consistent results may indicate some confusion of these two constructs.

In children with PDDs, differences in behavior and performance depending on context are the rule, not the exception. Lack of generalization is a major issue for interventionists. As such, seeking to reconcile disparate points of view among informants regarding a child's social behavior in an attempt to get a consistent picture seems to be somewhat counterproductive. Rather than working to eradicate, reduce or partition the factors contributing to variability in outcome, it may be more productive to both seek to understand these differences and consider that that they represent different realities, all of which may have some validity in a particular context (Kraemer et al. 2003). Stated another way, variability in outcomes may represent different expressions of intervention effect (or lack thereof), depending on the context in which intervention effects were observed, or the contexts in which social reciprocity domains are expressed, or both.

The inevitability of variable outcomes directly impacts the choice of outcome measures. Again, choosing a primary outcome measure, a standard practice in clinical trials, is commonly employed based on a plan to use simple analysis of variance models for examining change due to intervention, rather than selecting data analysis strategies that match the complexity of the measurement problem (Gueorguieva and Krystal 2004; Jacobson and Truax 1991). The basic problem is that "it is difficult to argue that adequate examination of such complex constructs can be captured with a single indicator" (De Los Reyes and Kazdin 2006, p. 556). Specificity regarding the estimated effect of the intervention on different aspects of the impairment requires the use of multiple indicators of change. A number of data analytic strategies are available to model change due to intervention and development. Differential structural equation models (dSEM) that incorporate estimates of the impact of one change in an individual's state on other changes within that individual and within the environment may be most appropriate for modeling the kind of intra-individual, dynamic variability associated with social growth and development as the result of intervention (Boker 2001). In contrast to growth curve analysis, with dSEM, measures are included as predictors as well as outcomes. Second order latent growth models incorporate multiple indicators of change over time, separating measurement error associated with a

specific indicator from time specific variance (Sayer and Cumsille 2001). Further, the assumption that the same indicator measures the same construct at different time points in development can be tested (Sayer and Cumsille 2001). For example, a questionnaire assessing social initiation behavior may not tap the same construct of this behavior at age 7 vs. age 17. Appropriate social initiation behavior is quite different at these two developmental time points, and the influence of latent factors or error may not be consistent. These issues complicate the analysis of longitudinal data with straightforward statistical models that treat data from two time points as representing the same construct.

Analysis of Group Data

Yet another complexity in measuring change in social reciprocity is that the analysis of outcome data obtained from intervention delivered in a group format requires that the impact of participation in a *particular* group be considered as a random variable. This adds another layer of complexity to the analysis of treatment effects. Baldwin et al. (2005) identified significant difficulties with the analysis of data related to group intervention for psychological/psychiatric problems in 101 research studies purported to support the efficacy of particular group-delivered interventions. After re-analysis of data, incorporating estimated intra-group correlations, less than 50% of the studies analyzed showed statistically significant effects, yielding 20 treatments for psychological disorders that no longer could be considered evidence based according to current American Psychological Association guidelines (Baldwin et al. 2005). Intra-group correlation must be considered in the design of SST studies for children with PDDs delivered in a group format, given that this factor has major implications for estimating appropriate sample size. Considering this issue, a number of the group studies included in recent reviews were severely underpowered.

Concluding Comments

Overall, the complexity of designing research to test the effects of group-delivered intervention to remediate social impairments in children with PDDs presents major challenges to the research community. The construct of social reciprocity is far too complex, and the impairment in social functioning in children with PDDs too variable, as to be amenable to intervention that is focused on teaching a discrete set of skills in a necessarily time limited fashion. The construct of social reciprocity, the associated aspects of this construct and the dynamic nature of development and change require reconsideration and refinement of the currently accepted guidelines for evaluating the efficacy of particular interventions. A more painstaking but necessary step in this process is to consider the construct of impaired social reciprocity in all its complexity, chose a specific aspect of the construct for intervention, and then specify the skills within that construct one aims to address. For example, one might identify social perception as a focus, and identify components of social perception that would be targets for intervention (Fig. 2). Using multiple methods to measure change in target skills of social perception and multiple global measures of social perception is preferable to measuring social reciprocity broadly construed. Further, the careful choice of multiple informants, attending to the need for varying perspectives and contexts through which observations of target behaviors are made, as well as anticipating the degree of collinearity among informants a priori strengthens the research design (Kraemer et al. 2003).

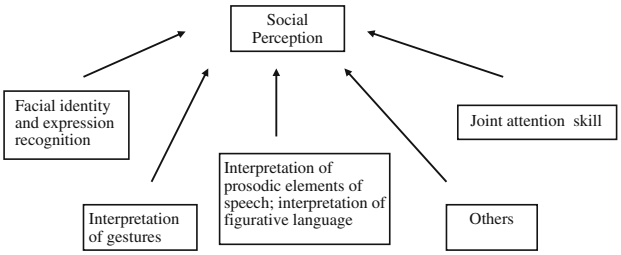

Fig. 2 Components of social perception as targets for intervention

Conclusions

As the focus on developing psychosocial intervention to address social and communication deficits in children with PDDs has increased over the last 10–15 years, concerned stakeholders have called for rigor in the testing of treatments. The need to establish treatments as evidence-based for this population is particularly acute, given the variety of novelty and alternative treatments available, some of which can be dangerous to the child (Offit 2008). For example, in 2005, a 5-year-old child with autism died following chelation therapy to remove "heavy metals" from his bloodstream, purportedly a cause of his autism (Wadman 2008).

Using established standards for determining the strength of the evidence for treatment is an important step, and the NIMH working group on psychosocial intervention charted a step-wise course for those in the field to follow. The guidelines are clear and reasonable, and bring some order to a field with a myriad of purported treatments. The need for well-designed studies to establish the usefulness of particular strategies for the vast number of individual skills that are required to develop social reciprocity skill is clear. However, *social reciprocity is more than the sum of its parts*.

Since it is not clear what essential elements, brought together within a SST program will result in significant change, and because measures designed to assess subtle changes in many aspects of social behavior do not yet exist, efforts to move forward along the pathway established by the NIMH working group are likely to yield disappointing results.

At this juncture, it is critical to acknowledge the tension between the requirement for rigor in the testing of treatments and the use of the current paradigm for establishing the evidence base of intervention with the kind of broad-based group intervention employed to remediate impairment in social reciprocity in children with PDDs (Rogers and Vismara 2008). A more fruitful approach may be to consider the strong evidence that observational studies, carefully done, can provide about the efficacy of treatments. Indeed, the maxim that observational studies always yield less accurate estimates of treatment effect compared to randomized controlled trials seems to be less and less defensible, as recent evidence indicates (Concato 2004; Concato et al. 2000).

Social reciprocity is a multi-dimensional, complex construct. The large number of factors that influence competence in social reciprocity, and the way in which these dynamic factors impact one another over the course of development needs further exploration and modeling. If one considers the original distinction between nomothetic and ideographic methods (Grice et al. 2006), it may be most useful, at this time, to employ ideographic methods and procedures to the analysis of individual and group data in the service of identifying context and time based knowledge versus seeking to make larger generalizations based on aggregate data (Grice et al. 2006; Kraemer et al. 2003). In other words, accepting the fact that social reciprocity is context-, time- and culture-dependent will allow for qualification of the results of any given study, and allow for exploration of variability in response to intervention. Effective intervention must be comprehensive and in many cases will need to be customized to address the particular needs of the affected child. This requires a more nuanced approach to testing—one which incorporates the dynamic nature of social reciprocity and change associated not just with intervention, but also associated with growth and development.

Acknowledgments We thank Fred Volkmar for his comments on this manuscript, and the participants in the Organization for Autism Research Convocation, Atlanta, 2008 for their comments as well. We thank Rebecca Pearlson for assistance with references.

References

Achenbach, T. (2006). As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Current Directions in Psychological Science, 15*(2), 94–98. doi:10.1111/j.0963-7214.2006.00414.x.

Baldwin, S., Murray, D., & Shadish, W. (2005). Empirically supported treatments or type I errors? Problems with the analysis of data from group-administered treatments. *Journal of Consulting and Clinical Psychology, 73*, 924–935. doi:10.1037/0022-006X.73.5.924.

Barnhill, G., Tapscott-Cook, K., Tebbenkamp, K., & Smith Myles, B. (2002). The effectiveness of social skills targeting nonverbal communication for adolescents with Asperger syndrome and related pervasive development delays. *Focus on Autism and Other Developmental Disabilities, 17*, 112–118. doi:10.1177/10883576020170020601.

Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. (2000). *Understanding other minds: Perspectives from cognitive neuroscience*. New York: Oxford University Press.

Barry, T. D., Klinger, G., Lee, J. M., Palardy, N., Gilmore, T., & Bodin, S. D. (2003). Examining the effectiveness of an outpatient clinic-based social skills group for high-functioning children with autism. *Journal of Autism and Developmental Disorders, 33*, 685–701. doi:10.1023/B:JADD.0000006004.86556.e0.

Bauminger, N. (2002). The facilitation of social-emotional understanding and social interaction in high-functioning children with autism: Intervention outcomes. *Journal of Autism and Developmental Disorders, 32*, 283–298. doi:10.1023/A:1016378718278.

Beelman, A., Phingsten, U., & Loesel, F. (1994). Effects of training social competence in children: A meta-analysis of recent evaluation studies. *Journal of Clinical Child Psychology, 23*, 260–271. doi:10.1207/s15374424jccp2303_4.

Begeer, S., Koot, H., Rieffe, C., Terwogt, M., & Stegge, H. (2008). Emotional competence in children with autism: Diagnostic criteria and empirical evidence. *Developmental Review, 28*(3), 342–369. doi:10.1016/j.dr.2007.09.001.

Boker, S. (2001). Differential structural equation modeling of intraindividual variability. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 3–28). Washington DC: American Psychological Association.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52*, 685–716. doi:10.1146/annurev.psych.52.1.685.

Concato, J. (2004). Observational versus experimental studies: What's the evidence for a hierarchy? *NeuroRx: The Journal of the American Society for Experimental NeuroTherapeutics, 1*, 341–347.

Concato, J., Shah, N., & Horwitz, R. (2000). Randomized, controlled trials, observational studies, and the hierarchy of research designs. *The New England Journal of Medicine, 342*(25), 1887–1892. doi:10.1056/NEJM200006223422507.

Curran, P., & Bollan, K. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 105–136). Washington DC: American Psychological Association.

De los Reyes, A., & Kazdin, A. (2005). Informant discrepancies" in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*(4), 483–509. doi:10.1037/0033-2909.131.4.483.

De los Reyes, A., & Kazdin, A. (2006). Conceptualizing changes in behavior in intervention research: The range of possible changes model. *Psychological Review, 113*, 554–583. doi:10.1037/0033-295X.113.3.554.

Golan, O., & Baron-Cohen, S. (2006). Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive

multimedia. *Development and Psychopathology, 18*, 591–617. doi:10.1017/S0954579406060305.

Grice, J., Jackson, B., & McDaniel, B. (2006). Bridging the nomothetic-ideographic divide: A follow up study. *Journal of Personality, 74*, 1191–1218. doi:10.1111/j.1467-6494.2006.00407.x.

Gueorguieva, R., & Krystal, J. (2004). Move over ANOVA: Progress in analyzing repeated-measures data and its reflection in the Archives of General Psychiatry. *Archives of General Psychiatry, 61*, 310–317. doi:10.1001/archpsyc.61.3.310.

Jacobson, N., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19. doi:10.1037/0022-006X.59.1.12.

Klin, A. (2000). Attributing social meaning to ambiguous visual stimuli in higher-functioning autism and Asperger syndrome: The social attribution task. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*(7), 831–846. doi:10.1111/1469-7610.00671.

Klin, A., Jones, W., Shultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: Lessons from autism. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 358*(1430), 345–360. doi:10.1098/rstb.2002.1202.

Koegel, L., & Koegel, R. (1995). *Current issues in autism; learning and cognition in autism*. New York: Plenum Press.

Koenig, K., White, S., Pachler, M., Lewis, L. M., Pachler, M., Klin, A., & Scahill, L. (in press). Promoting social skill development in children with pervasive developmental disorders: A feasibility and efficacy study. *Journal of Contemporary Psychotherapy*.

Konstantareas, M., & Stewart, K. (2006). Affect regulation and temperament in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 36*(2), 143–154. doi:10.1007/s10803-005-0051-4.

Kraemer, H., Measelle, J., Ablow, J., Essex, M., Boyce, W., & Kupfer, D. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *The American Journal of Psychiatry, 160*, 1566–1577. doi:10.1176/appi.ajp.160.9.1566.

Lipsey, M., & Wilson, D. (1993). The efficacy of psychological, educational and behavioral treatments: Confirmation from meta-analysis. *The American Psychologist, 48*, 1181–1209. doi:10.1037/0003-066X.48.12.1181.

Lord, C., Wagner, A., Rogers, S., Szatmari, P., Aman, M., Charman, T., et al. (2005). Challenges in evaluating psychological interventions for autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 14*, 395–404.

Mackay, T., Knott, F., & Dunlop, A. (2007). Developing social interaction and understanding in individuals with autism spectrum disorder: A group work intervention. *Journal of Intellectual & Developmental Disability, 32*, 279–290. doi:10.1080/13668250701689280.

Matson, J., Matson, M., & Rivit, T. (2007). Social skills treatments for children with autism spectrum disorders: An overview. *Behavior Modification, 31*(5), 682–707. doi:10.1177/0145445507301650.

Morrison, L., Kamps, D., Garcia, J., & Parker, D. (2001). Peer mediation and monitoring strategies to improve initiations and social skills for students with autism. *Journal of Positive Behavior Interventions, 3*(4), 237–250. doi:10.1177/109830070100300405.

Offit, P. (2008). *Autism's false prophets: Bad science, risky medicine, and the search for a cure*. New York: Columbia University Press.

Ozonoff, S., & Miller, J. (1995). Teaching theory of mind: A new approach to social skills training for individuals with autism. *Journal of Autism and Developmental Disorders, 25*, 415–433. doi:10.1007/BF02179376.

Paul, R., Shriberg, L. D., McSweeny, J., Cicchetti, D., Klin, A., & Volkmar, F. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 35*(6), 861–869. doi:10.1007/s10803-005-0031-8.

Provencal, S. (2003). *The efficacy of a social skills training program for adolescents with autism spectrum disorders.* Unpublished doctoral dissertation, University of Utah.

Quinn, M., Kavale, K., Mather, S., Rutherford, R., & Forness, S. (1999). A meta-analysis of social skill interventions for students with emotional or behavioral disorders. *Journal of Emotional and Behavioral Disorders, 7*(1), 54–65. doi:10.1177/106342669900700106.

Rao, P., Beidel, D., & Murray, M. (2008). Social skills interventions for children with Asperger's syndrome or high-functioning autism: A review and recommendations. *Journal of Autism and Developmental Disorders, 38*, 353–361. doi:10.1007/s10803-007-0402-4.

Rogers, S., & Vismara, L. (2008). Evidence-based comprehensive treatments for early autism. *Journal of Clinical Child and Adolescent Psychology, 37*(1), 8–38. doi:10.1080/15374410701817808.

Sansosti, F., & Powell-Smith, K. (2006). Using social stories to improve the social behavior of children with Asperger syndrome. *Journal of Positive Behavior Interventions, 8*(1), 43–57. doi:10.1177/10983007060080010601.

Sayer, A., & Cumsille, P. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 177–200). Washington DC: American Psychological Association.

Schneider, B. (1992). Didactic methods for enhancing children's peer relations: A quantitative review. *Clinical Psychology Review, 12*, 363–382. doi:10.1016/0272-7358(92)90142-U.

Schriebman, L. (2000). Intensive behavioral/psychoeducational treatments for autism: Research needs and future directions. *Journal of Autism and Developmental Disorders, 30*(5), 373–378. doi:10.1023/A:1005535120023.

Schultz, R., Grelotti, D., Klin, A., Kleinman, J., Van Der Gaag, C., Marois, R., et al. (2003). The role of the fusiform face area in social cognition: Implications for the pathobiology of autism. *Philosophical transactions of the royal society B-Biolgical Sciences, 358*(1430), 415–427.

Smith, T., Scahill, L., Dawson, G., Guthrie, D., Lord, C., Odom, S., et al. (2007). Designing research studies on psychosocial interventions in autism. *Journal of Autism and Developmental Disorders, 37*, 354–366. doi:10.1007/s10803-006-0173-3.

Solomon, M., Goodlin-Jones, B., & Anders, T. (2004). A social adjustment enhancement intervention for high functioning autism, Asperger's syndrome, and pervasive developmental disorder NOS. *Journal of Autism and Developmental Disorders, 34*, 649–668. doi:10.1007/s10803-004-5286-y.

Spence, S. (2003). Social skills training with children and young people: Theory, evidence and practice. *Child and Adolescent Mental Health, 8*(2), 84–96. doi:10.1111/1475-3588.00051.

Tager-Flusberg, H., Paul, R., & Lord, C. (2003). Language and communication in autism. In F. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders, vol. 1: Diagnosis, development, neurobiology, and behavior* (3rd ed., pp. 335–364). Hoboken, NJ, US: Wiley.

Tse, J., Strulovitch, J., Tagalakis, V., Meng, L., & Fombonne, E. (2007). Social skills training for adolescents with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders* (serial online).

Volkmar, F., Paul, R., Klin, A., & Cohen, D. (2005). *Handbook of autism and the pervasive developmental disorders*. New York: Wiley.

Wadman, M. (2008). Stalled trial for autism highlights dilemma of alternative treatments. *Nature, 253*, 259. doi:10.1038/454259a.

Wang, D., & Bakhai, A. (2006). *Clinical trials: A practical guide to design, analysis, and reporting*. London: Remedica.

Webb, B., Miller, S., Pierce, T., Strawser, S., & Jones, W. P. (2004). Effects of social skill instruction for high-functioning adolescents with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities, 19*(1), 53–62. doi:10.1177/10883576040190010701.

White, S., Koenig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of Autism and Developmental Disorders, 37*, 1858–1868. doi:10.1007/s10803-006-0320-x.

Yang, N., Schaller, J., Huang, T., Wang, M., & Tsai, S. (2003). Enhancing appropriate social behaviors for children with autism in general education classrooms: An analysis of six cases. *Education and Training in Developmental Disabilities, 38*, 405–441.