

Quality indicators for single–case research on social skill interventions for children with Autistic Spectrum Disorder

Shin-Yi Wang & Rauno Parrila
University of Alberta

In this paper, we describe a quality checklist that parents, teachers, clinicians, and policy-makers with basic research skills can use to systematically evaluate the methodological quality of single-case studies on social skill training of children with autistic spectrum disorder (ASD). We provide a rationale for included quality indicators, and two examples of how the checklist can be used to assess the quality of individual papers and the quality of a body of research.

Introduction

Social interaction problems are recognized as one of the core deficits for children with autistic spectrum disorder (ASD; White, Keonig, & Scahill, 2007). Recently, the diagnostic criteria and conceptualization of autistic disorder have been broadened from autism to ASD (Fombonne, 2005), resulting in an increase of individuals with the central deficit in social interaction, such as many with pervasive developmental disorders not otherwise specified (PDDNOS), or Asperger’s disorder. Children with high function autism (HFA), PDDNOS, or Asperger’s disorder show fewer cognitive and language deficits, but social interaction issues can be a major barrier for them that impacts negatively on their adjustment in school and community. When children with ASD are placed in inclusive settings, they tend to be isolated or experience difficulties in establishing friendships with peers. Even for the children with PDDNOS, Asperger’s Syndrome, or HFA, the ones with more preserved cognitive skills, interacting appropriately with others can be a difficult task (Rao, Beidel, & Murray, 2008). Traditional intervention models that stress academic or basic living skills fail to meet their needs and, as a result, interventions targeting social skills have started to gain in popularity.

Several models have been developed for the social skill training of children with ASD, including behavior modification, peer-mediated training, social story, video modeling, self-management, pivotal response training, joint attention training, and buddy system (e.g., Bass & Mulick, 2007; Matson, Matson, & Rivet, 2007; Scattone, 2007). Although the strategy of modeling target behaviours and providing reinforcement tends to be the most commonly used, there are increasing varieties of the intervention approaches in the newer studies. While some approaches, such as the peer-mediated approach (Matson et al., 2007), have generally proved to be effective for children with ASD, others have produced less consistent findings across different studies.

As more studies on social skill interventions are completed, there is a growing need to assess and integrate the evidence of the efficacy or effectiveness of the interventions they provide. Parents want to know how to choose an effective model for their children with ASD, clinicians would like to adopt the most effective model for their evidence-based practice, and policy-makers are interested in funding programs with proven effectiveness. Therefore, how to examine the quality of the intervention research systematically has become a critical issue for those who are interested in social skill interventions for children with ASD.

This paper describes development of a quality checklist that parents, teachers, clinicians, and policy-makers with basic research skills can use to systematically evaluate the methodological quality of single-case studies (i.e., studies that use each participant as his/her own control, and that aim to demonstrate experimental control; see e.g., Horner, Carr, Halle, McGee, Odom, & Wollery, 2005) on social skill training of children with ASD. We focus on single-case studies because a recent review by Matson et al. (2007) indicated that more than 90% of the intervention studies employ this design. Below, we will describe in more detail the development of the quality checklist and provide an explanation of the items included. We will also provide two examples of how the checklist can be used, first to examine the overall quality of individual studies, and then to examine the quality of a small body of research. The complete checklist is provided in Appendix A.

Quality Indicators

The first step involved identifying quality indicators that could be used to assess methodological quality of single-case studies. The initial list of the quality indicators was adapted from Horner et al., (2005). Horner et al. list multiple criteria that can be used to examine different dimensions of single-case research, including information given on participants, settings, dependent and independent variables, baseline data collection, experimental control/internal validity, external validity, and social validity. As Horner et al.'s indicators were not specific to social skills interventions for children with ASD, their criteria were compared to those used in three recent papers that focused more closely on this specific topic (Lord et al., 2005; Smith et al., 2007; Reichow, Volkmar, & Cicchetti, 2008). Several quality indicators were added to the checklist, such as use of standardized instruments for diagnosis, information of the peers and interventionists, the criteria for the percentage of the sessions used to examine the inter-rater agreement, utilization of multiple baseline or reversal design, the amount of data points in the baseline and intervention phases, and the use of blind agents for establishing social validity. Table 1 lists the initial quality indicators.

Table 1
The Initial Checklist of Quality Indicators for Single-Case Studies of Social Skills Training for Children with ASD

Primary Quality indicators	Quality Indicators
<i>Participants:</i>	Measuring procedure generated a quantifiable index
Gender and age of ASD participant is provided	The data on each baseline phase present a stable pattern/trend
Ethnicity information of ASD participant is provided	DVs measured at least 3 times on each intervention phase
Recruiting procedure of ASD participant is explained	The data on each intervention phase present a stable pattern/trend
IQ, academic performance, or adaptive skills data provided	The inter-rater agreement was collected on at least 20% of sessions

(Table 1 Cont'd)

Selection criteria of ASD participant are explained	The inter-rater agreement is over 80% or Kappa over .60 between raters
ASD diagnosis made by professionals specialized in autism	The raters were blind to research
The study used a standardized instrument for diagnosis	The raters were different from the interventionist
Detailed information on training & qualifications of interventionists provided	Research Design: The study used multiple baseline or reversal design
Detailed information on the recruiting procedure of peers provided (if applicable)	Secondary Quality indicators
Detailed information of selection criteria of peers (if applicable) provided	External Validity:
Settings/materials used for social skill training:	The researcher reported data on maintenance effect
Information of the settings and materials sufficient for replication	The data on generalization of effects are collected across different contexts
Potential confounding factors caused by the settings/materials controlled	Social Validity:
Independent Variables:	Data on direct gains (other than DVs) caused by intervention reported
IVs were described in sufficient detail for replication	Data on secondary gains caused by intervention reported
Standardized procedure used for implementation (i.e., manual)	Data on consumer satisfaction reported
IV implemented at least three times at three different time points	Qualitative data reported for social importance of change in DVs
Researchers controlled the contamination between subjects	The implementation of IV cost- and time-effective
The researchers assessed the fidelity of implementation	IV implementation needs minimal adjustment to natural settings
Dependent Variables:	The research examined SV over extended (3 month later) period
DVs were operationally defined	The agents used to establish SV blind to research
DV is clearly linked to target behaviors	The agents used to establish SV adopted from typical contexts

The quality indicators are divided into two parts: primary and secondary quality indicators. The primary quality indicators focus on the internal validity of the research. The more the studies meet the criteria for the primary quality indicators, the better the studies manage the confounding factors and can demonstrate the causal relationship between the intervention and the observed outcomes. We designate internal validity indicators as primary a study without internal validity cannot produce useful information. The secondary indicators are concerned with the external and social validity of the studies. A study earns credit in external validity when it presents evidence for generalizability of the results across various time frames, settings, or participants. A study demonstrates better social validity if more people recognize the importance of the intervention or give credit to the outcome of the intervention. A more detailed description of the initial quality indicators together with a rationale for their inclusion follows.

Primary Quality Indicators

The primary quality indicators are used mainly to examine whether (a) the study includes sufficient information about the participants, settings/material, and the independent and dependent variables, (b) the researchers manipulated and measured variables faithfully, (c) there were sufficient data points and stable pattern in the data, and (d) the study adopted one of the designs that can demonstrate a functional relationship between target behavior and intervention.

Participants. The researcher should provide detailed and precise information regarding the participants. This information is necessary for others to be able to replicate the research, or to apply the results to different groups of children. Detailed information should be provided about the children with ASD, interventionists, and the peers and parents, if applicable.

First, the information about the children with ASD should include gender, age, ethnicity, recruiting procedure, selection criteria, information on relevant ability such as IQ, academic ability, or adaptive skills, and confirmative information of ASD diagnosis. Age and gender is

the basic information in order to facilitate the selection of participants for replication (Horner et al., 2005). Furthermore, age has been found to be related to the differential gains from the intervention, with younger children tending to gain more from the intervention than older children (Baker-Ericzen, Stahmer, & Burns, 2007; Corsello, 2005).

Ethnicity information can demonstrate the demographic characteristic of the population and can be related particularly to the effectiveness of parent education interventions (Baker-Ericzen et al., 2007). Selection criteria and recruiting procedure provide explicit standards regarding what kinds of characteristics the participants exhibited and how they were selected. In addition, the information on relevant abilities such as IQ, language abilities, or the index of social interaction for ASD participants should be provided in detail. Because ASD represents a heterogeneous group and the abilities of children within subgroups of ASD can be diverse (Fombonne, 2005; NRC, 2001; White et al., 2007), children with different levels of abilities can respond to the same intervention differently (Shea, 2004; Sherer & Schreibman, 2005). The levels of language ability for different subgroups of ASD can range from no speech to fluent but idiosyncratic communication (NRC, 2001). The social interaction of children with ASD can be categorized as aloof, passive, or active but odd (Wing & Gould, 1979). The intelligent levels of the children with ASD can range from severe mental retardation to superior levels. NRC (2001) indicates that there likely is no single intervention approach that benefits all different types of children with ASD equally. Thus, detailed information about the abilities of participants with ASD is necessary for both replication and assessment of generalizability.

Similar to the ability levels, the accuracy of ASD diagnosis can interfere greatly with the efficacy of the intervention. Smith et al. (2007) suggested that researchers should ensure faithfulness of ASD diagnosis by using standardized diagnostic tools. Hence, use of diagnostic tools with standardized procedure or description such as CARS (Childhood Autism Rating Scale), ADOS (Autism Diagnostic Observation Schedule), ADI-R (Autism Diagnostic Interview-Revised), DSM-IV, or ICD-10 (International Statistical Classification of Diseases and Related Health

Problems-10th version) is included as one of the quality indicators. In addition, diagnosis from psychologists, psychiatrists, or pediatricians is included as a separate quality indicator as it can support the accuracy of the diagnosis in addition to the use of standardized diagnostic tools.

The research on social skill training usually involves interventionists to implement the training and the information regarding the background and training experience of the interventionists should be provided. Replicating research with insufficiently trained interventionists can result in insignificant outcomes and undermine the efficacy of the intervention model. Furthermore, if the peers or parents participated as mediators in the research (i.e., peer-mediated or parent-mediated model), the researchers should present information such as their recruiting procedure and selection criteria to facilitate future replication and meta-analyses.

Settings and materials used for social skill training. The information on settings and materials is important as different settings and materials may motivate children differently and interfere with their social interaction dramatically even without intervention. Therefore, the researcher should provide sufficient information regarding how the setting was arranged or what type of materials – such as games or tools – were available and used. Structuring the setting and materials in a consistent way can demonstrate the functional relationship between the outcomes and intervention more clearly.

Independent Variables (IV). In social skill training, independent variables (IVs) are the specific procedures or strategies used for intervention, and the implementation of IVs should lead to change of the social behavior. Clear and detailed descriptions of independent variables (IV) are necessary for replication and generalization studies. Using standardized manuals for implementing IVs generally ensures there is detailed information to repeat the procedure, and the manual also can be used for creating a checklist to examine if the intervention is being implemented faithfully.

Furthermore, a study can earn credit if it tries to control possible

confounding factors (i.e., contamination effects between children), manipulate IVs at least three different times (Reichow et al., 2008), and assess the fidelity of IV implementation. These quality indicators examine whether the researchers have provided sufficient evidence to support the linkage between the observed behaviors and the IV.

Dependent variables (DV). Dependent variables (DV) are the measurements of the target behaviors that the researchers aim to change (either increase or decrease) with the implementation of IVs. All possible target behaviors need to be defined operationally so that they can be measured with minimal error, and the measured behaviors have to be clearly connected to socially desired outcomes that they are chosen to represent. Finally, the measurement procedure has to be clearly described to allow replication.

In order to demonstrate the effect of intervention, data on DVs should be collected a minimum of three times during each baseline and intervention phase (Horner et al., 2005; Reichow et al., 2008). Further, data should display a stable pattern or trend at each phase. Without stable pattern or trend, the study cannot provide sufficient evidence for the differences between the phases. Lack of stable pattern or trend may also indicate presence of confounding factors. As a result, verifying a fundamental link between IV and DV by contrasting the patterns at different phases becomes difficult.

In addition, because most measurements of DVs in social skill intervention studies involve raters, the researchers should test the reliability of the measurement by comparing the rating outcomes across different raters for a minimum of 20% of the sessions (Reichow et al., 2008), and the inter-rater agreement should be at least 0.6 or Kappa coefficient over .60 (Horner et al., 2005; Reichow et al., 2008). Moreover, if the study includes raters that are different from interventionists and raters are blind to the research, validity of the ratings is further increased.

Research Designs. An additional quality indicator was added to indicate whether the study used a design that clearly can support a functional

relationship between targeted social behaviors and the intervention. Multiple-baseline and reversal designs were chosen as preferred designs because both decrease threats to internal validity and provide a more powerful statement for the efficacy of the intervention (NRC, 2001; Richards, Taylor, Ramasamy, & Richards, 1999; Smith et al., 2007). In the multiple-baseline design, the researchers implement the intervention to different participants at different settings, or to different behaviors at different time frames. If the change in dependent variables corresponds to the implementation of the intervention at different time points across different participants, settings, or behaviors, more convincing evidence to support the effect of the intervention is generated. With the use of reversal design, the study can rule out the effects of maturation and history that generally confound the interpretation of the intervention effect in a simple A-B design. The reversal design also provides opportunities to examine the generalization effect of the intervention. The alternating design is not appropriate for examining intervention outcomes of social skill training because the effect of one intervention can interfere with the possible effect of the other intervention. In addition, the changing criterion design is aimed to increase or decrease developed skills, and changing criterion design may not be appropriate in social skill training because social skill training usually involves developing new skills.

Secondary Quality Indicators

The secondary quality indicators are used to examine external and social validity of the research. External validity is mainly concerned with the generalizability of the results to different settings and participants, whereas social validity is mainly concerned with recognized social importance of the intervention outcomes.

External Validity. External validity is regarded as high when the target behavior is maintained over longer periods of time and we have a reason to believe that the positive effect of intervention can be generalized to different individuals in different contexts; after all, the ultimate goal of the intervention research is to find interventions that benefit more participants with similar difficulties and maintain the gains across

different settings and time. The researchers can assess maintenance effects of the intervention by measuring the DVs again some time after the intervention has been discontinued. In order to distinguish maintenance effect from generalization effect, the data on maintenance effect should be obtained with the presence of the same experimental setting, participants, and materials as were used during the intervention. The generalized effects of the intervention can be assessed through measuring DVs while having ASD participants interact with different persons, or with the same persons in different settings, or with the same person but with different activities or toys. Hence, the external validity of the study is increased if the researchers included maintenance or follow-up sessions over an extended period, and if they verify the effect of the intervention in different contexts.

Social Validity. Social validity is increased in a single-case study if the study examines social importance of the intervention outcomes to the children with ASD and to other people around the children, such as school staff, teachers, friends, siblings or parents, or society. Social validity can be established directly or indirectly. For instance, children, parents, or teachers may report how well the intervention had improved children's social interactions other than the DVs that the researchers measured. In other cases, the intervention may indirectly benefit children's self-esteem or child-parent relationship that were not the primary focus of the intervention. Therefore, social validity indicators include measurements of the direct and secondary gains of the intervention, consumer satisfaction reports, and qualitative reports of the progress. The direct gains are related to the improvement of social behavior other than the DVs, whereas the secondary gains are defined as the improvement of non-social behavior or psychological status such as child-parent relationship, self-value, self-confidence, happiness, disruptive behaviors or social alliance. Moreover, if the implementation of IV is conducted in a context close to natural settings, there will be better chances for adaptation to real world settings. As a result, the social validity of the study is strengthened if minimal adjustment of IV implementation is required for real world settings.

In addition, using raters that are blind to the research to evaluate social

validity is counted as one of the quality indicators because blindness of raters will decrease the possible confirmation bias. Additional credit will be given if the researchers examine social validity three months or longer after the intervention has stopped. Furthermore, another quality indicator, based on a suggestion by Horner et al. (2005), is placed in the checklist to inspect whether the IV implementation is cost-effective and time-effective.

Examples of how the quality indicator checklist can be used

After the initial checklist of quality indicators was developed, we used the checklist to examine research papers that reported single-case research with focus on the social skill training of children with ASD. The quality indicators in the initial list used to examine the cost-effectiveness and time-effectiveness of the implementation were excluded after probing the first two papers. The reason for exclusion was the lack of agreement as to the operational definitions for cost-effectiveness or time-effectiveness. The revised quality checklist had 39 quality indicators remaining (see Appendix A).

Target Papers

Thirty recent (published between 2000 and 2007) papers were located either through Academic Search Premier, Web of Science, and TOC Premier databases using keywords “autism,” “social skill,” “intervention” and “training,” or from the reference lists of recent review articles on social skills interventions for children with ASD (Bass et al., 2007; Matson et al., 2007; Rao et al., 2008; Scattone, 2007). Five papers were excluded because they did not meet most criteria of single-case studies. Ten of the remaining 25 papers that adopted one or more models of behavior modification, peer-mediated training, social story, pivotal response training, joint attention training, or buddy system were randomly selected for this review. The 10 papers are summarized in Table 2 and numbered by superscript in the reference list. The total number of children with ASD was 28 including two females and 26 males. The ages of children with ASD ranged from three to nine years old. The number of participants with ASD within these studies ranged

Developmental Disabilities Bulletin, 2008, Vol.36, No. 1 & 2

from one to five. The intervention models or strategies used in these studies included behavior analysis, pivotal response training, peer-mediated approach, social story, role play/modeling/prompt/prime/reinforcement, and social script; 70 % of them used more than one intervention model or strategy. Two studies did not report the duration or frequency of the intervention because they used varied cut-off criteria for different phrases of intervention and the duration or frequency of the intervention for each participant with ASD varied. Two studies did not do so because the duration or intensity of the intervention could not be accumulated due to the adoption of all classmates or the context as independent variables. Four of the ten studies indicated significant improvement across all target behaviors, while the remaining studies indicated that some target behaviors improved or that some of the ASD participants showed improvement in social behaviors. In terms of settings, four of the studies were conducted in a public school, two were conducted in laboratory settings, two in a private school or private education center, one in a community clinic, and one in children's community or their homes.

Table 2
Summary of the Reviewed Papers

Study ID	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Publication year	2005	2007	2003	2006	2000	2003	2008	2007	2008	2005
Settings	Priv	Lab	Clin	P.S.	P.S.	Lab	P.S.	P.S.	Priv	home/ community
Number of children with ASD	1	4	4	3	2	5	2	2	3	2
Number of Male	1	4	3	3	2	5	2	2	3	1
Number of Female	0	0	1	0	0	0	0	0	0	1
Design	M-B	AB	AB	M-B	reversal	M-B	M-B	M-B	M-B	M-B
Models/ strategies										

(Table 2 Cont'd)

Behavior Analysis	☆					☆		☆		
pivotal response training						☆	☆			☆
peer mediated		☆		☆				☆	☆	
social story					☆					
social script					☆					
role play/modeling/prompt/prime/reinforcement	☆	☆	☆	☆	☆	☆	☆		☆	
outcome of intervention	★	★	★	★	★★	★	★	★★	★★	★★
generalization of intervention	★	NA	★	★	★★	★	★	★	★★	★★
Score for Primary QI	.28	.48	.62	.53	.67	.74	.64	.72	.53	.66
Score for Secondary QI	.50	.00	.40	.38	.25	.40	.30	.40	.30	.40
TOTAL SCORE	.33	.38	.56	.50	.57	.65	.55	.64	.47	.59

NOTE: Study ID identifies the study in question in the reference list; Priv = private school or educational center; Lab = laboratory; Clin = clinic; P.S. = public school; M-B: multiple-baseline design; AB = one baseline session + one intervention session; Reversal = design includes withdrawal phase; ★ = indicate partial improvement in the outcome of target behaviors; ★★ = indicate improvement in all target behaviors; NA = not applicable.

Scoring of Individual Papers. Thirty-nine quality indicators were used to examine the ten papers. If the paper met the criterion for a specific quality indicator, it was given one point for that item. If partial criterion was met, 0.5 point was given. If the item was not applicable (for example, the quality indicator for detailed information about peers is not applicable to the studies that don't involve peers for intervention), it was counted as "not applicable." Thus, the maximum score varied across studies and was less than 39 for the studies for whom not all quality indicators could be applied. To provide a common metric across studies, we calculated the proportion of applicable quality indicators met. For

example, if the study received 25 points across 38 applicable items, its total score was $25/38 = .66$. The total score results are presented on the last line of Table 2 and can be construed as representing an assessment of the overall quality of the paper. Note, however, that the total scores are somewhat simplistic estimates of the total quality of the studies as all quality indicators were given equal weighting. Similar ratio scores were also calculated separately for the primary quality indicators and the secondary quality indicators.

The total quality scores of the reviewed papers ranged from .33 to .65 with the mean of .52 ($SD = 0.10$) indicating that, on average, these studies met about half of the applicable quality indicators. The scores of the 10 papers ranged from .28 to .74 over the twenty-nine primary indicators with the mean of .59 ($SD = 0.14$). The scores of the 10 papers ranged from 0 to .5 over the ten secondary indicators with the mean of .33 ($SD = 0.14$).

Compared with other papers, the paper with the highest total score met most primary quality indicators, with the exception of the indicators of using standardized procedure for implementation, providing detailed ethnicity information of ASD participants, providing detailed information regarding the training or qualification of interventionist, demonstrating stable patterns in baseline and intervention phases, and having the interventionist different from experimenters or blind to the research. However, the study demonstrated the functional relationship by adopting multiple-baseline design.

Examination of the Results by Quality Indicators. Table 3 presents the results across different quality indicators and can be used to examine the overall quality of this small body of research and to identify specific problems that may be replicated across multiple studies. On the positive side, Table 3 shows that all of the ten papers provided information of their participants' gender and age, manipulated IV at least three different times, provided an operational definition of DV, linked DV as measured clearly to the target behaviors, generated quantifiable index for DV, and repeated measurement at least 3 times at each intervention phase. Seven to nine papers also provided detailed information of recruiting procedures for peers, described the IV in detail, measured DV at least 3

times at baseline phase, reached 80 % interrater agreement or 0.6 kappa index, adopted either multiple baseline or reversal design, and used interventions that require minimal adjustments for implementation in natural settings. Six papers provided detailed information about the selection criteria for the ASD children, used standardized instruments for diagnosis, and collected data on maintenance effects, and three out of five papers that used peers included information on their selection criteria.

Table 3

The percentage of papers meeting the criteria of each primary and secondary quality indicator

Quality indicators	%			
	Yes	No	Part	NA
Primary Quality indicators				
<i>Participants:</i>				
Gender and age of ASD participant is provided	100	0	0	0
Ethnicity information of ASD participant is provided	40	60	0	0
Recruiting procedure of ASD participant is explained	30	70	0	0
Selection criteria of ASD participant are explained	60	40	0	0
Information of relevant abilities (IQ, academic performance, or adaptive skills) provided	40	60	0	0
ASD diagnosis made by professionals specialized in autism	20	80	0	0
The study used a standardized instrument for diagnosis	60	40	0	0
Detailed information on training & qualifications of interventionists provided	30	70	0	0
Detailed information on the recruiting procedure of peers provided	70	10	0	20
Detailed information of selection criteria of peers provided	30	50	0	20
<i>Settings/materials used for social skill training:</i>				
Information of the settings and materials sufficient for replication	40	60	0	0
Potential confounding factors caused by the settings/materials controlled	40	60	0	0
<i>Independent Variables:</i>				
IVs described in sufficient detail for replication	70	30	0	0

(Table 3 Cont'd)

Standardized procedure used for implementation (i.e., manual)	30	70	0	0
Researchers controlled the contamination between subjects	40	60	0	0
IV implemented at least three times at three different time points	100	0	0	0
The researchers assessed the fidelity of implementation	40	60	0	0
<i>Dependent Variables:</i>				
DVs were operationally defined	100	0	0	0
DV is clearly linked to target behaviors	100	0	0	0
Measuring procedure generated a quantifiable index	100	0	0	0
DVs measured at least 3 times on each baseline phase	80	20	0	0
The data on each baseline phase presents a stable pattern/trend	30	40	30	0
DVs measured at least 3 times on each intervention phase	100	0	0	0
The data on each intervention phase present a stable pattern/trend	10	80	10	0
The inter-rater agreement is over 80% or Kappa over .60 between raters	90	0	10	0
The inter-rater agreement was collected on at least 20% of session	90	10	0	0
The raters were blind to research	0	100	0	0
The raters were different from the interventionist	40	60	0	0
<i>Research Designs:</i> using multiple baseline or reversal design	70	30	0	0
Secondary Quality indicators				
<i>External validity:</i>				
The researcher reported data on maintenance effect	60	40	0	0
The data on generalization of effects are collected across different contexts	50	50	0	0
<i>Social validity:</i>				
Data on direct gains (other than DVs) caused by intervention reported	40	60	0	0
Data on secondary gains caused by intervention reported	20	80	0	0
Data on consumer satisfaction reported	10	90	0	0

(Table 3 Cont'd)

Qualitative data reported for social importance of change in DVs	20	80	0	0
IV implementation needs minimal adjustment to natural settings	70	30	0	0
The research examined SV over extended (3 month later) period	0	100	0	0
The agents used to establish SV blind to research	10	60	0	30
The agents used to establish SV adopted from typical contexts	40	30	0	30

Note: Total numbers of reviewed papers=10; Yes = Meet the criterion; No = Do not meet the criterion; Part = Meet the criterion partially; NA = The criterion not applied to the paper

The criteria that half or more than half of the papers did not meet included providing information on ethnicity and relevant abilities of the ASD participants, as well as whether they were diagnosed with ASD by professionals specialized in autism. Half of the papers reported data on generalization of the effects to different contexts. Only three to four papers provided detailed information on the training and qualifications of the interventionists, or the settings and materials that were used. Four papers controlled for materials and settings, contamination between subjects, fidelity of implementation, and rater-bias. While most papers collected sufficient amounts of data, only three showed a stable pattern/trend in baseline phase, and only one paper showed a stable pattern/trend on each intervention phase.

Finally, most papers fared poorly in terms of the social validity quality indicators, indicating that this is an area where there is ample room for improvement.

Discussion

Interacting appropriately with others is a significant challenge to many children with ASD and intervention studies targeting social skills have increased both in popularity and in variety. Several models have been developed and tested for the social skill training of children with ASD, and both the outcomes and the quality of the studies evaluating the models vary widely. With more studies published there is a growing

need for tools that help not only researchers but also parents, teachers, clinicians, and policy-makers to assess the accumulating evidence for different models. One important part of this assessment is the examination of the quality of research used to support different intervention programs; only high-quality studies can provide a basis for evidence-based practice, and choosing an intervention program or programs to implement and fund requires examination of both the effectiveness of those programs as well as the quality of the studies establishing the effectiveness. How to examine the quality of the intervention research systematically has become a critical issue for those who are interested in social skill interventions for children with ASD. Hence, this paper aimed to develop a checklist of quality indicators that can be used by a variety of people with basic research skills to systematically review the quality single-case studies of social skill intervention for children with ASD.

The developed checklist includes several quality indicators for examining internal, external and social validities of the single-case research papers. Parents, teachers, clinicians, and policy-makers with basic research skills can go through and check the criteria of the checklist one by one while reading each research paper. They can give credits to the study for its internal validity by examining whether there is detailed information of participants, interventionist, IVs, and DVs, sufficient and reliable data-points across phases, control over confounding factors, and a research design that can demonstrate a functional relationship between the intervention and the outcome. In particular, providing sufficient information on different aspects of the study is important because it allows replications of the studies that are necessary for establishing the efficacy of any intervention. External validity is established if the study applies the intervention to different interactive people, settings, or materials. Furthermore, the study can earn credits on social validity when it reports on how the participants and other people recognized the contribution of the intervention. However, the indicators of internal validity are more important when judging the overall quality of the study than the indicators of external validity or social validity. Simply put, if the study lacks internal validity, there are no valid results that can be generalized or proven socially important. To acknowledge this, we

clustered the indicators related to internal validity under the heading of primary quality indicators.

In the second half of the paper, we provided examples of how the quality indicator checklist can be used to assess both the quality of individual studies as well as the quality of a body of research on a specific topic. Taken together, Tables 2 and 3 indicate significant flaws in both internal and external validity indicators, and that no single study is clearly above the criticism. For example, none of the papers in this review provided all of the information needed for a replication study, and only one included most required information (except the information regarding the training background or qualifications of the interventionist). In addition, many of those studies could improve their quality if they have had examined the fidelity of the implementation or provided operational definitions and measurable indexes for both the IV and the DV. Although the ten papers selected for the review may not fully represent the field, the results of this review highlight the need to examine the quality of studies carefully before accepting their results. Furthermore, researchers interested in studying social skill intervention programs for children with ASD would benefit from using this quality checklist to examine how well they have designed and reported their studies.

Some limitations with the quality indicators should be noted. First, the quality indicator checklist may require additional modification when it is used to examine a large body of papers. For example, the now excluded criteria of cost- and time-effectiveness could be added back if proper operational definitions become available. Those criteria can be important when we try to examine if the intervention model can be implemented in natural settings. In addition, using the total scores to rank the studies should be done with care because each indicator is now given equal weight. The primary indicators should be weighted more in the final decision since those items are central to the quality of the research.

References

- Baker-Ericzen, M.J., Stahmer, A.C., & Burns, A. (2007). Child demographics associated with outcomes in a community-based
Developmental Disabilities Bulletin, 2008, Vol.36, No. 1 & 2

- pivotal response training program. *Journal of Positive Behavior Interventions*, 9(1), 52-60.
- ⁽³⁾ Barry, T. D., Klinger, L. 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(6), 685-701.
- Bass, J. D., & Mulick J.A. (2007). Social play skill enhancement of children with autism using peers and siblings as therapists. *Psychology in the School*, 44(7), 727-734.
- ⁽²⁾ Chung, K.M., Reavis, S. Mosconi, M., Drewry, J., Matthews, J., & Tasse, M. J. (2007). Peer-mediated social skills training program for young children with high-functioning autism. *Research in Developmental Disabilities*, 28, 423-436.
- Corsello, C.M. (2005). Early intervention in autism. *Infant and Young Children*, 18(2), 74-88.
- ⁽⁴⁾ Delano, M., & Snell, M. E. (2006). The effects of social stories on the social engagement of children with autism. *Journal of Positive Behavior Interventions*, 8(1), 29-42.
- Fombonne, E. (2005). The changing epidemiology of autism. *Journal of Applied Research in Intellectual Disabilities*, 18(4), 281-294.
- ⁽⁷⁾ Harper, C. B., Symon J. B. G., & Frea, W. D. (2008). Recess is time-in: Using peers to improve social skills of children with autism. *Journal of Autism and Developmental Disorders*, 38(5), 815-826.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71(2), 165-179.
- ⁽¹⁰⁾ Koegel, R. L., Werner, G. A., Vismara, L. A., & Koegel, L. K. (2005). The effectiveness of contextually supported play date interactions between children with autism and typically developing peers. *Research and Practice for Persons with Severe Disabilities*, 30(2), 93-102.
- ⁽⁵⁾ Laushey, K. M., & Heflin, L. J. (2000). Enhancing social skills of kindergarten children with autism through the training of multiple peers as tutors. *Journal of Autism and Developmental Disorders*, 30(3), 183-193.
- ⁽⁸⁾ Lee, S., Odom, S. L., & Loftin, R. (2007). Social engagement with peers and stereotypic behavior of children with autism. *Journal of Positive Behavior Interventions*, 9(2), 67-79.

- ⁽⁹⁾ Liber, D. B., Frea, W. D., & Symon, J. B. G. (2008) Using time-delay to improve social play skills with peers for children with autism. *Journal of Autism and Developmental Disorders*, 38(2), 312-323.
- Lord, C., Wagner, A., Rogers, S., Szatmari, P., Aman, M., Charman, T., Dawson, G., Durand, V. M., Grossman, L., Guthrie, D., Harris, S., Kasari, C., Marcus, L., Murphy, S., Odem, S., Pickles, A., Scahill, L., Shaw, E., Siegel, B., Sigman, M., Stone, W., Smith, T., & Yoder, P. (2005). Challenges in evaluating psychosocial interventions for autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, 35(6), 695-708.
- Matson, J. L., Matson, M. L., & Rivet, T. T. (2007). **Social-skills treatments for children with autism spectrum disorders an overview.** *Behavior Modification*, 31(5), 682-707.
- National Research Council (NRC). (2001). *Educating children with autism*, National Academy Press, Washington, DC.
- Rao, P. A., Beidel, D. C., & Murray, M. J. (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(2), 353-361.
- Reichow, B., Volkmar, F.R., & Cicchetti, D.V. (2008). Development of the evaluative method for evaluating and determining evidence-based practices in autism. *Journal of Autism and Developmental Disorders*, 38(7), 1311-1319.
- Richards, S. B., Taylor, R. L., Ramasamy, R., & Richards, R. Y. (1999). Single subject research application in educational and clinical settings. (pp.134-147 & 145-164). San Diego, London : Singular Publishing Groups, Inc.
- ⁽¹⁾ Sawyer, L. M., Luiselli, J. M., Riciardi, J. N., & Gower, J. L. (2005). Teaching a child with autism to share among peers in an integrated preschool classroom: Acquisition, maintenance, and social validity. *Education and Treatment of Children*, 28(1), 1-10.
- Scattone, D. (2007). Social skills intervention for children with autism. *Psychology in the School*, 44(7), 717-726.
- Shea, V. (2004). A perspective on the research literature related to early intensive behavioral intervention (Lovaas) for young children with autism. *Autism*, 8(4), 349-367.
- Sherer, M. R., & Schreibman, L. (2005). Individual behavioral profiles and *Developmental Disabilities Bulletin*, 2008, Vol.36, No. 1 & 2

predictors of treatment effectiveness for children with autism. *Journal of Consulting and Clinical Psychology*, 73, 525-538.

Smith, T., Scahill, L., Dawson, G., Guthrie, D., Lord, C., Odom, S., Rogers, S., & Wagner, A. (2007). Designing research studies on psychosocial interventions in autism. *Journal of Autism and Developmental Disorders*, 37, 354-366.

⁽⁶⁾ Whalen, C., & Schreibman, L. (2003). Joint attention training for children with autism using behavior modification procedures. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 44(3), 456-468.

White, S. W., Keonig, K., & Scahill L. (2007). Social skills development in children with autism spectrum disorder: A review of the intervention research. *Journal of Autism and Developmental Disorders*, 37(10), 1858-1868

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders*, 9(1), 11-29.

Author Note

Shin-Yi Wang and Rauno Parrila, Department of Educational Psychology, University of Alberta. Please address all correspondence to Shin-Yi Wang at shinyi@ualberta.ca

Appendix A

The Quality Indicator Checklist for Single-Case Research in ASD

© Shin-yi Wang & Rauno Parrila.

Can be reproduced for personal use without permission.

Quality indicators	%			
Primary Quality indicators	Yes	No	Part	NA

Participants:

Gender and age of ASD participant(s) is provided

Ethnicity information of ASD participant(s) is provided

Recruiting procedure of ASD participant(s) is explained

Selection criteria of ASD participant(s) are explained

Information on relevant abilities (IQ, academic performance, or adaptive skills) provided

ASD diagnosis made by professionals specialized in ASD

The study used a standardized instrument for diagnosis

Detailed information on training & qualifications of interventionists provided

Detailed information on the recruiting procedure of peers provided

Detailed information of selection criteria of peers provided

Settings/materials used for social skill training:

Information on the settings and materials sufficient for replication

Potential confounding factors caused by the settings/materials controlled

Independent Variables:

IVs described in sufficient detail for replication

Standardized procedure used for implementation (i.e., manual)

Researchers controlled the contamination between subjects

IV implemented at least three times at three different time points

The researchers assessed the fidelity of implementation

Dependent Variables:

DVs were operationally defined

DVs clearly linked to target behaviors

Measuring procedure generated a quantifiable index

(Appendix A Cont'd)

DVs measured at least 3 times on each baseline phase

The data on each baseline phase present a stable pattern/trend

DVs measured at least 3 times on each intervention phase

The data on each intervention phase present a stable pattern/trend

The inter-rater agreement over 80% or Kappa over .60 between raters

The inter-rater agreement collected on at least 20% of sessions

The raters were blind to research

The raters were different from the interventionist

Research Designs: using multiple baseline or reversal design

Secondary Quality indicators

External validity:

The researcher reported data on maintenance effect

The data on generalization of effects collected across different contexts

Social validity:

Data on direct gains (other than DVs) reported

Data on secondary gains caused by intervention reported

Data on consumer satisfaction reported

Qualitative data reported for social importance of change in DVs

IV implementation needs minimal adjustment to natural settings

The research examined SV over extended (3 month later) period

The agents used to establish SV blind to research

(Appendix A Cont'd)

The agents used to establish SV adopted from typical contexts

Note: Yes = the study meets the criterion; No = the study does not meet the criterion; Part = the study meets the criterion of this quality indicator partially; NA = the quality indicator is not applicable to the study.