

# Determining treatment outcome in early intervention programs for autism spectrum disorders: A critical analysis of measurement issues in learning based interventions

Johnny L. Matson\*

*Department of Psychology, Louisiana State University,  
Baton Rouge, LA 70803, United States*

Received 2 June 2005; received in revised form 12 July 2005; accepted 28 July 2005

---

## Abstract

One of the areas receiving the greatest attention from researchers studying autism spectrum disorders in recent years involves psychologically based early intervention programs. Various claims of cure, marked improvement in social and communication skills, and improved I.Q. are among the conclusions that have been drawn by various researchers. However, little has been done to analyze the dependent variables used in these studies and their impact on the conclusions reached regarding treatment effectiveness. Obviously, this set of measures is crucial since these methods define which behaviors “improved” and to what extent. The present review analyzes the current status, strengths, and weaknesses of these measurements.

© 2006 Elsevier Ltd. All rights reserved.

*Keywords:* Early intervention; Autism spectrum disorder

---

## 1. Introduction

One of the most discussed issues in the child treatment literature is the success of early intervention programs with autism spectrum disorder (ASD) children (Matson & Minshawi, 2006; Moore & Goodson, 2003). These interventions, which address the range of symptoms evinced by ASD children, are behavioral/learning-based procedures. This pragmatic factor defines the parameters for the present review.

---

\* Tel.: +1 225 578 4104/4125; fax: +1 225 388 4125.

*E-mail address:* [Johnmatson@aol.com](mailto:Johnmatson@aol.com).

Various claims have been made about the specific effects of these programs, but a general consensus appears to be growing in the research literature. The consensus is that early intervention is valuable and, within limits, the more intense the intervention, the greater the gains despite variability in outcomes within groups of children treated (Symes, Remington, & Brown, 2006). Despite the considerable research focus on these treatment studies, however, little has been done to analyze the various dependent variables that have been used and the effect these assessment methods have on how treatment effects are perceived. When guidelines regarding assessment for children with autism spectrum disorders (ASD) are published, they tend to focus almost exclusively on differential diagnosis (AACAP, 1999).

To date, behaviorally-based interventions have been found to be by far the most effective means of intervention for children with ASD (Schreibman, 2000). Many more studies are available on behaviorally-based interventions than other topics and thus, the bulk of the review will focus on the dependent measures used in these studies. More recently pharmacological interventions have begun to appear and some guidelines for assessment have been described. Therefore, while assessment is gleaned primarily for behavioral treatments these data also have implications for pharmacology outcome studies. Other methods of intervention, some of which are quite unusual, have been proposed. Some authors have suggested that treatments have become so unusual that they refer to autism in particular as a fad magnet for off the wall interventions (Metz, Mulick, & Butter, 2005). For this review, assessment guidelines for mainstream behavioral and pharmacological treatment will be the theme for the paper.

## **2. Outcome measures for behavioral interventions**

It would not be possible to cover the vast literature on treatment in one paper. Therefore, selective studies were reviewed to develop the general points about the type of dependent variables employed. Two methodologies are evident. The first method follows a strict behavior analytic approach. These studies employ single case designs with one to five children being evaluated. Multiple baseline and/or reversal designs are used as the primary means of methodological control. A second, much smaller, set of studies uses more conventional control group pre-test–post-test designs with the application of statistical tests of significance aimed at differentiating change between groups across time and within groups.

## **3. Behavior analysis outcome measures**

Typical of these applied behavior analysis single case design studies in general are the operational definition of one to three problems/challenging behaviors such as self-injury, stereotypies, and aggression. Studies of this nature typically note in the description of participants that the children evince an ASD, typically autism. However, the purpose of the study typically, as noted, is on dealing with a few specific behaviors, usually self-injury (Mohr & Sharpley, 1985), aggression, noncompliance, stereotypies (Matson, Benavidez, Compton, Paclawskyj, & Baglio, 1996; Matson & Minshawi, 2006), or spontaneous language (Matson, Sevin, Fridley, & Love, 1990; Matson, Sevin, Box, Francis, & Sevin, 1993). This research was of considerable value in establishing the initial intervention strategies that later were expanded to cover a much greater range of targets for intervention. The goal of the studies were as much about developing and refining intervention strategies that could be used for a broad array of ASD symptoms as they were about intervening on the ASD child's core symptoms. Thus, the move from research of this type to broad program research is viewed as evolutionary. It should also be

stressed that as a general point one can develop a treatment strategy powerful enough to positively effect aggression, self-injury, noncompliance and stereotypies, then it is also likely to be effective with ASD core symptoms.

#### **4. Program treatment outcome**

As the field has advanced and matured with respect to this massive literature on testing a few specific targets, more ambitious efforts geared toward multiple behavior domains have emerged. Occurring with this shift in focus has been the use of more and more expansive dependent measures, which assess a broad range of success criteria on which individual and groups of autistic and PDD-NOS children can be measured. These two diagnostic groups among all the ASD groups have received most of the research attention, primarily for pragmatic reasons. These two ASDs are by far the most common and readily identified in young children.

One of the first of these studies was by [Lovaas, Koegel, Simmons, and Long \(1973\)](#). The measurements included what they refer to as multiple-response recordings and measures of intelligence: the Stanford-Binet Intelligence Test and the Vineland Social Maturity Scales. This study is important for a number of reasons, not the least of which because the paper helped establish a series of assessments methods, which in turn helped establish what constitutes how a successful “program” is evaluated (e.g., self-stimulation/now called stereotypy, echolalic speech, appropriate speech, social nonverbal behavior, appropriate play, and intelligence). Very lengthy definitions, upwards of a page, were provided for multiple-response behaviors. Using appropriate play as an example, considerable effort was exerted not only to give specific examples of the behaviors (i.e., stacking tiles or blocks, putting crayons in boxes, handling and examining various toys) but also the context in which the behaviors were displayed.

[Sallows and Graupner \(2005\)](#) suggest in their treatment paper that the interventions developed at UCLA in the 1960s and 1970s are perhaps the best known and documented treatment methods. This statement is probably an overgeneralization. Clearly, there is a vast literature on individual procedures to remediate specific behaviors developed by an array of behavioral researchers ([Matson & Minshawi, 2006](#)). It is argued that a more precise, but certainly not a trivial distinction, is that the UCLA program is perhaps the most studied package program and, along with TEACCH ([Schopler, Brehm, Kinsbourne, & Reichler, 1971](#); [Schopler, Mesibov, & Hearsey, 1995](#)), is the best known. Because the focus has largely shifted from the development of specific treatment strategies to the application of packages aimed at treating the entire syndrome, perhaps an unintentional but even greater impact has been to define what success or effective is for autistic children, based on the outcome measures that have been selected. To date, little has been done to establish empirically what the most critical variables are in claiming improvement or cure. This literature needs much greater attention since it will shift the debate as to what optimal change is and concomitantly which tasks and skills should receive the greatest attention during intervention.

A number of general assessment domains appear to be relevant when surveying the control group treatment program literature ([Matson & Minshawi, 2006](#)). Again, while these points are largely based on behavioral studies, they also have considerable applicability for pharmacotherapy trials. Eight general topic areas have been identified as essential for an adequate discourse regarding cure or substantial improvement in ASD symptoms and the treatments used for this set of disorders. A brief review of these assessment areas follows.

#### 4.1. Group assignment

Particularly for children, where development is occurring in the lifespan, more rapidly than at other points often in spurts, carefully matched control groups are essential. To the extent this is possible with young children, they should be matched on age in months, intellectual disability (ID), and overall severity of ASD in addition to challenging behaviors. Outcome studies have shown that these variables can be prognostic in establishing long-term outcome, regardless of the intervention (Charman et al., 2005; Howlin, Goode, Hutton, & Rutter, 2004; Stoelb et al., 2004). Without such careful matching, at least some change is likely due to developmental or behavioral artifact versus true improvement from the intervention. These variables overlap with methodology to some extent and the reader is referred to Kasari (2002) who makes a number of cogent points on the topic. Despite the need for adherence to such methodological factors, such precautions have rarely occurred to date in the intervention literature. To be sure, the small number of children who qualify for treatment and the heterogeneity of symptoms and symptom severity pose great challenges in this regard. However, it is at the same time a bit disappointing that greater care has not been taken to ensure such matching, or to at least acknowledge in the research papers that such limitations exist. To some extent these assessment problems compromise results and what can be concluded about the interventions employed.

Consistently, the biggest problem at this point in group design/assessment is the total lack of a control group. In her 2002 review, Kasari noted that only five programmatic group studies reported a control group, and of these only one reported random assignment. Unfortunately, these issues continue to persist in published programmatic studies (Sallows & Graupner, 2005). Again, it is emphasized that studies of this sort are extremely labor intensive. Also, some might argue that it is unethical or at the very least impractical to assign children to a no treatment control list. Alternative solutions exist, such as using single case designs for each child, delaying intervention and then introducing intervention for controls after some period of intervention (e.g. 6 months–1 year) for the experimental group, or providing a standard preschool or school program as the control group.

#### 4.2. I.Q. and adaptive skill measures

Dating to the earliest attempts at programmatic intervention for autism, I.Q. and concomitant adaptive skill measures have been common assessments of treatment outcome. Often, marked improvements in these measures occur and the authors report increased I.Q. (Lovaas, 1987). On some level this assertion is no doubt accurate. However, multiple factors require additional redress in this regard. First, many outcome studies vary the I.Q. test from pre- to post-test. Thus, for example, a Bailey Scale for Infant Development might be used at pre-test and a WPPSI or Binet might be used at post-test making direct comparisons of pre-test–post-test data problematic. Additionally, since considerable variability in language is evident in these children, tests of intelligence for nonverbal children may be mixed with other I.Q. measures at pre-test. Finally, adaptive measures such as the Vineland Social Maturity are simply not designed for use with very young ASD children, many of whom have marked language and intellectual disabilities. Thus, an artificial ceiling effect emerges since test items simply cannot discriminate differences in skill levels. Clearly, trying to compare different tests within pre-test or from pre-test to post-test, and using measures that are insufficient to adequately evaluate skill levels of ASD children is highly problematic.

Second, the claim of increased I.Q. (e.g., Weiss, 1999) might be better framed as increase in I.Q. test scores. This point is not trivial since the difference is between reporting an outcome for a

dependent variable versus suggesting the cause of the test score change. A probable explanation for increased scores, at least in part, is likely due to treatment effects on compliance and attention. Additionally, given the very young age of many children studied, i.e., 24–60 months, even a 6-month treatment study could result in marked developmental gains in compliance and attention simply due to developmental maturation. Control groups carefully matched on I.Q. (to the extent that is possible) and the other pertinent variables noted previously, should be of considerable value in addressing this issue.

Third, predictive validity of I.Q. tests at very young ages are at best unstable. Therefore, it may be difficult to lay claim to improved I.Q. because of unreliable pre-test classification on this variable, or due to marked lack of uniform developmental gains, due to individual differences in children.

#### *4.3. Measures of autism and other ASDs*

Intuitively, change in core symptoms of ASD over time would appear to be one of the most important, if not the most important measures of change for programs claiming to improve or even cure ASD. This approach is strongly endorsed and has been used in many of the program evaluation studies of autism/PDD-NOS published. Even with this approach, problems have surfaced. One major issue is when intervention begins. No matter what the measure, instability in scores over time becomes greater as the pre-test scores are obtained for younger and younger children. Thus, despite assertions that treatment programs for ASD classification start as early as 6 months (Maestro et al., 2002), 3 years may be the earliest age at present for accurately diagnosing ASD, and then only for the most severe cases (Charman et al., 2005). In one study that addressed this issue, atypical autism and PDD were frequently misdiagnosed as language delay for children 20 months of age (Cox et al., 1999). To date the majority of outcome studies do not use one of the primary measures of autism as an outcome measure (Rogers, 1998). This omission is of concern if one is aiming at improvement in core symptoms of autism. Making claims of improvement in autism versus I.Q. scores or adaptive skills would seem to suggest this avenue. It is suggested that the ADI-R or CARS, which have been used in outcome research, be considered for any group outcome study (Charman et al., 2005; Weiss, 1999).

#### *4.4. Psychopathology and challenging behavior*

Researchers describe a variety of forms of psychopathology, which can covary with ASD. These disorders include depression (Ghaziuddin, Ghaziuddin, & Greden, 2002; Long, Wood, & Holmes, 2000), generalized anxiety disorder (Woodard, Groden, Goodwin, Shanower, & Bianco, 2005), attention deficit hyperactivity disorder (Geurts, Verte, Oosterlaan, Roeyers, & Sergeant, 2004), and phobias (Luscre & Center, 1996). Despite a sporadic but growing literature, these issues have simply not been addressed. Future research should be conducted with assessments of comorbid psychopathology for selecting and matching children across groups. Additionally, outcome of studies should assess not only core symptoms of ASD but also core symptoms of any other form of psychopathology requiring treatment. Simply because a researcher or clinician is more interested or knowledgeable about ASD versus another problem, such as the psychopathologies noted above, does not minimize the potential importance of these covarying disorders.

A major issue, however, is the lack of adequate assessment measures to screen for psychopathology in this population. However, until such instrumentation, with norms on ASD

children are available, standard childhood instruments should be used for group classification and as pre-test/post-test measures. Thus, for example, the Child Depression Inventory (Kovas, 1980) could serve as an outcome measure for depression or the Fear Survey Schedule for Children could be used for children evincing anxiety disorders (Ollendick, 1983). These measures, while not normed on this particular childhood population are well-respected and established measures that should be useful until more specialized measures are available.

Challenging behaviors have not fared much better in programmatic research. This situation is vexing since the behaviors are common in ASD children and very debilitating. To the extent they have been addressed in programmatic studies the state of affairs could be described best as hit and miss.

#### *4.5. Operational target behaviors and maintaining variables*

Operationally defined target behaviors are a hallmark of behaviourally-based treatment programs (Matson et al., 1996). These assessment methods are particularly salient and commonly used for challenging behavior, which frequently occur concomitantly with ASD (Schreibman, 1988). Similarly, functional assessment technology has become a major approach to determining maintaining variables and establishing more valid and effective treatments for this problem area (Matson & Minshawi, 2006). The most commonly reported methods in the research literature are experimental functional analyses (Iwata, Dorsey, Slifer, Bauman, & Richmond, 1982) and the Questions About Behavior Change (Matson, Bamburg, Cherry, & Paclawskyj, 1999). Remarkably, given that these methods have been in place for some time and have been codified in legal decisions, functional assessment has not been routinely used in programmatic outcome studies of ASD.

Various treatment studies in the ASD literature emphasize the fact that they do not use aversives (Eikeseth, Smith, Jahr, & Eldevik, 2002; Sallows & Graupner, 2005). The reasons for such omissions are not reported. However, it would be very concerning if children with challenging behaviors, which are much more likely to require aversive procedures such as brief time-out, overcorrection or similar procedures compared to other typical intervention targets for ASD, were excluded from such programs. Similarly, a movement to use antipsychotic drugs is also beginning to appear. Using drugs which have the potential for serious and permanent neurological long-term side effects, without first resorting to proven behavioral methods would be unfortunate. Thus, studies which routinely look at not only I.Q., academic, social and language issues, but challenging behaviors employing the assessment methods just described appear warranted.

#### *4.6. Measurement of fidelity*

Given that the bulk of treatment is provided by parents or bachelors level staff versus professionals with advanced degrees and extensive experience in applied behavior analysis, accuracy of training procedure applications is particularly salient. Most treatment research, both for ASD and in psychology and psychiatry in general, do not use such methodology. Matson and Senatore (1981) provide a model for such assessment. In their study, sessions were taped and raters blind to treatment conditions (behavior therapy or traditional psychotherapy raters) were given checklists with salient characteristics of each intervention and asked to rate the tenants evident in training sessions. Data were used to provide feedback to therapists to enhance therapy accuracy and ensure against therapist drift over time away from tenants of a specified therapy.

This latter point is particularly relevant to ASD program research where interventions may be in place for a year or more since length of training increases the possibility of therapist drift from accuracy of treatment implementation.

Sallows and Graupner (2005) provide a somewhat different but excellent description of treatment fidelity measurement for an ASD early intervention program. Therapists were required to pass written tests on training procedures from *The Me Book* (Lovaas et al., 1981). Videotaped therapy was assessed prior to initiation of treatment and weekly supervision was provided to each therapist by the senior author. Some combination of the best elements from this study and the fidelity measures described by Matson and Senatore (1981) should be included in any programmatic study on ASD children.

#### 4.7. Measures for group assignment versus measures of treatment outcome

Another factor to consider is separating measures for group assignment versus measures used to evaluate treatment outcome. A measure or measures may not be sufficient to make accurate classification. Researchers have argued that more accurate diagnoses, and thus better group assignment result when multiple methods such as observation and clinical interviews are used in combination with scaling methods. This broader set of available information in diagnostic decision-making is defined in this context as clinical judgment (Charman & Baird, 2002).

In addition to the argument for creating a more robust means of group classification, there is also the issue of confounding of effects. Using the same dependent measure to classify groups and then using the same measure to determine change adds a level of bias toward positive outcome, or at least the possibility of a confound. This is the case since the investigator is measuring identical items used for selection and treatment effects versus independent methods used to measure the same construct. A more stringent method would be to use another test for autism in the group classification phase, or exclude checklists entirely, relying on observation, interview, DSM and WHO criteria with multiple raters (two or three) who would all need to independently arrive at the same classification. Checklists such as the CARS or ADI-R could then be used as pre-test post-test outcome measures. Again, the areas of program research in ASD are still relatively new, but these safeguards seem to be reasonably easy to introduce and would enhance the design and thus robustness of treatment outcome results.

#### 4.8. Side effects

Side effects are defined as unintended effects. No intervention is likely to be side effect free, yet this topic has largely been ignored in the intervention literature on children with ASD. Operant punishment procedures have been criticized for producing adverse side effects yet the minimal data available on such side effects do not support such a claim (Matson & Taras, 1989). The point here is that clinical folklore and scientific data are likely to differ, at least in some instances. For ASD research these claims should be empirically tested, regardless of the intervention used, then refuted or accepted. With young children evincing ASD, who are asked to comply to structured tasks over extensive periods of time on a daily basis, it is hard to imagine that no unintended side effects will emerge (e.g., tantrums, noncompliance, yelling, etc.). Medication is also an issue, particularly with the trend toward using antipsychotic medications in children as young as 4. Geddes, Freemantle, and Harrison (2000) have reported serious long-term side effects of the typical and atypical antipsychotics. If they are to be used, systematic scaling methods such as the MEDS (Matson et al., 1998) which measures tardive dyskinesia, akathisia



and other long-term neurological effects would seem to be prudent, particularly for young children where long-term outcome is so critical. Additionally, these data would need to be considered when evaluating overall program effectiveness. Thus, if brief time-out or an atypical antipsychotic resolved challenging behavior but produced serious side effects the intervention might still be rejected for young ASD children.

## **5. How do researchers define effective**

The assertion made here is that the outcome measures that researchers choose to include in their studies define the “universe” of potential effects. Using this definition it is clear that I.Q. and measures of language are the clear favorites to date in this regard. Interestingly, behaviorally-based programs rarely include measures of the core symptoms of autism such as the ADI-R or CARS, despite the fact that autism or PDD-NOS children, versus Rett’s, Childhood Disintegrative Disorder, or Asperger’s are almost universally the only children studied. Additionally, while the primary rationale for introducing pharmacotherapy into this field is to treat challenging behaviors such as self-injury or aggression, these behaviors are rarely outcome measures for the programmatic studies to date. This finding is a bit disappointing since children evincing such behaviors are far more likely to advance at lower rates in learning and a vast treatment literature with applied behavior analysis is available for successful application to these problem behaviors (Matson & Minshawi, 2006).

A second parameter is the criteria used to determine improvement. This goal is largely determined by the experimental design used. Thus, single case research relies on visual inspection while group designs use more conventional levels of statistical significance (e.g., .05 or .01). Both methods are certainly appropriate but more applications of social validity criteria also seem to be warranted. That is, how do the children in experimental groups compare to normally developing same-aged children at the conclusion of the treatment phase. This issue has been addressed sporadically in two ways. McEachin, Smith, and Lovaas (1993), for example, have looked at measures such as the Personality Inventory for Children to determine if experimentals scored in the “typical child” range on test norms at post-test. More efforts of this sort are encouraged in future studies with a broader range of normed scales across a broader range of domains. For example, from pre-test to post-test if experiments could demonstrate that “X” percent of the experimental group went from severe or moderately autistic, to not scoring as autistic with only a few PDD symptoms on a standardized measure of ASD, this would be a substantial and powerful argument for the intervention.

A second method is to actually test normal developing children matched with ASD children treated on the specific target behaviors assessed at post-test. Children are matched on age, sex, I.Q. and related variables to make the most valid comparison of ASD and related symptoms possible. Success would be defined as training the ASD children to “normal performance” as judged by the scores obtained on normal developing peers based on scores from operationally defined targets (Matson, Kazdin, & Esveldt-Dawson, 1980). Combining both this method and the procedures of McEachin et al. (1993) would seem to be prudent.

## **6. Trends**

Relative to assessment of outcomes, a few observations are in order. First, group designs are becoming more common as a means of assessment versus single case research designs. Second, standardized I.Q. and adaptive measures initially introduced over 25 years ago still appear to be a



primary measure of outcome (Lovaas et al., 1973). Third, operational target behaviors as outcome measures are being used less frequently. Fourth, in the place of operation target behaviors are measures of general behavior problems such as the Aberrant Behavior Checklist (Achenbach, 1991), and subcomponents of scales such as the Vineland Social Maturity Scale (Sparrow, Balla, & Cicchetti, 1984). Fifth, despite initial resistance of many behavior analysts, the people who have developed most of the effective interventions for this group, the construct of Autism Spectrum Disorders appears to have been established in the thinking of researchers, therefore, measures of autism such as the ADI-R (Lord, Rutter, & LeCouteur, 1994), should be common in programmatic studies as outcome measures. Sixth, measures of treatment fidelity are now being used, infrequently. When employed paper and pencil tests completed by treatment staff and analysis of videotaped treatment sessions are typical (Sallows & Graupner, 2005). Seventh, mixed results exist with respect to employing the same measurements for group assignment as are used for determining treatment outcome.

## **7. Concluding remarks**

Despite various methodological flaws from an assessment standpoint in the existing early intervention literature with ASD, there have been a sufficient number of replications to declare the methods “promising”. This finding should be no surprise since the techniques employed are based on a massive empirical literature on applied behavior analysis (Matson & Minshawi, 2006). Despite these promising gains, one would hope and expect to see more defined and refined assessment methods as the treatment literature on the topic evolves.

There is ample evidence that I.Q. test scores improve whether it is a true increase in I.Q. or better attention and compliance by the child. Authors should consider stating their findings in this manner while insuring that the same I.Q. measures are used within and across groups at pre-test and from pre-test to post-test. Second, while the Vineland Social Maturity Scale is a fine test, more specialized adaptive behavior tests, particular for very young children and/or those with the greatest level of developmental delays, will need to be addressed.

Using language measures is excellent since along with social skills they are hallmarks of ASD (Matson & Minshawi, 2006). Again, measures more specific to young ASD children are needed, particularly with respect to social skills. Social skills tests such as the MESSIER, which are specifically designed for intellectual disabilities, are rarely used but should be considered. Similarly, and perhaps remarkably, core tests of autism such as the CARS and ADI-R are rarely used as pre-test post-test assessments. It is suggested that a compelling argument for improving or curing ASD requires that these measures be included before all others.

ASD children also frequently evince comorbid psychopathology and/or challenging behaviors. It is unclear in treatment studies to date if these children are being ruled out for treatment or whether these disorders and behaviors are simply not being addressed. In either case, they warrant considerably more attention, given past success of behavioral programs for these problems, and given their prevalence in this group of children.

Accurately carrying out treatments is being measured in some studies but needs to be done more uniformly across studies. Some points on how to carry out these procedures have been offered. Methodologically it is also crucial that control groups be used and that participant matching and random assignment be followed. Addressing these issues should put to rest most of the criticisms leveled against this area of research.

Finally, authors need to consider more systematically and carefully constructed ways of establishing clinical success. Social validity measures appear to be a very important methodology

in this regard. It is also recommended that brief consumer satisfaction questionnaires might be constructed for parents and teachers. These people could be interviewed about potential improvements they have noted in children. Data such as attending regular classes or other related real world developments should also be included but they must be tempered with the realization that these are secondary/administrative decisions that may or may not reflect clinical gains.

These programmatic treatment studies do show great promise. The number of published papers is increasing rapidly with no indication that they will slow in the near future. The methodological refinements one would hope for in assessing treatment outcome have not kept pace, however. Given the huge investment in time and resources researchers, parents, teachers, the children themselves and other concerned individuals are putting into this training, a more extensive emphasis on assessment parameters such as those outlined in this paper appears warranted. This point is particularly salient since many of the recommendations made here are not extensive in time or cost relative to overall investment in programmatic treatment studies.

## References

- American Academy of Child and Adolescent Psychiatry. (1999). Practice parameters for the assessment and treatment of children, adolescents, and adults with autism and other pervasive developmental disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 325–545.
- Achenbach, T. M. (1991). *Manual for the teacher's report form and 1991 profile*. Burlington: University of Vermont Department of Psychiatry.
- Charman, T., & Baird, G. (2002). Practitioner review: Diagnosis of autism spectrum disorder in 2- and 3-year-old children. *Journal of Child Psychology and Psychiatry*, 43, 289–305.
- Charman, T., Taylor, E., Drew, A., Cockerill, H., Brown, J., & Baird, G. (2005). Outcome at 7 years of children diagnosed with autism at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry*, 46, 500–513.
- Cox, A., Klein, K., Charman, T., Baird, G., Baron-Cohen, S., Swettenham, J., et al. (1999). Autism spectrum disorders at 20 and 42 months of age: Stability of clinical and ADI-R diagnoses. *Journal of Child Psychology and Psychiatry*, 40, 719–732.
- Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive behavioral treatment at school for 4- to 7-year-old children with autism. *Behavior Modification*, 26, 49–68.
- Geddes, J., Freemantle, N., & Harrison, P. (2000). Atypical antipsychotics in the treatment of schizophrenia: Systematic overview and meta-regression analysis. *British Medical Journal*, 312, 1371–1376.
- Geurts, H. M., Verte, S., Oosterlaan, J., Roeyers, H., & Sergeant, J. A. (2004). How specific are executive functioning deficits in attention deficit hyperactivity disorder and autism? *Journal of Child Psychology and Psychiatry*, 45, 836–854.
- Ghaziuddin, M., Ghaziuddin, W., & Greden, J. (2002). Depression in persons with autism: Implications for research and clinical care. *Journal of Autism and Developmental Disorders*, 32, 299–306.
- Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry*, 45, 212–229.
- Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richmond, G. S. (1982). Toward a functional analysis of self-injury. *Analysis and Intervention in Developmental Disabilities*, 2, 3–20.
- Kasari, C. (2002). Assessing change in early intervention programs for children with autism. *Journal of Autism and Developmental Disorders*, 32, 447–461.
- Kovas, M. (1980). Rating Scales to assess depression in school-aged children. *Acta Paedopsychiatrica*, 46, 305–315.
- Long, K., Wood, H., & Holmes, N. (2000). Presentation, assessment and treatment of depression in a young woman with learning disability and autism. *British Journal of Learning Disabilities*, 28, 102–108.
- Lord, C., Rutter, M., & LeCouteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 23, 659–685.
- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, 55, 3–9.

- Lovaas, O. I., Ackerman, A. B., Alexander, D., Firestone, P., Perkins, J., & Young, D. (1981). *Teaching developmentally disabled children: The me book*. Autism, TX: Pro-Ed.
- Lovaas, O. I., Koegel, R., Simmons, J. O., & Long, J. S. (1973). Some generalization and follow-up measures on autistic children in behavior therapy. *Journal of Applied Behavior Analysis*, 6, 131–166.
- Luscre, D. M., & Center, D. B. (1996). Procedures for reducing dental fear in children with autism. *Journal of Autism and Developmental Disorders*, 26, 547–556.
- Maestro, S., Muratori, F., Cavallaro, M. C., Pei, F., Stern, D., Golse, B., et al. (2002). Attention skill during the first 6 months of age in autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41, 1239–1245.
- Matson, J. L., & Taras, M. E. (1989). A 20-year review of punishment procedures and alternative methods to treat problem behaviors in developmentally delayed persons. *Research in Developmental Disabilities*, 10, 85–104.
- Matson, J. L., Bamburg, J. W., Cherry, K. E., & Paclawskyj, T. R. (1999). A validity study on the Question About Behavior Function (QABF) Scale: Predicting treatment success for self-injury, aggression and stereotypies. *Research in Developmental Disabilities*, 20, 163–176.
- Matson, J. L., Benavidez, D. A., Compton, L. S., Paclawskyj, T., & Baglio, C. (1996). Behavioral treatment of autistic persons: A review of research from 1980 to the present. *Research in Developmental Disabilities*, 17, 433–465.
- Matson, J. L., Kazdin, A. E., & Esveltd-Dawson, K. (1980). Training interpersonal skills among mentally retarded and socially dysfunctional children. *Behaviour Research and Therapy*, 18, 419–427.
- Matson, J. L., Mayville, E. A., Bielecki, J., Barnes, U. H., Bamburg, J. W., & Baglio, C. S. (1998). Reliability of the Matson evaluation of drug side-effects (MEDS). *Research in Developmental Disabilities*, 19, 501–506.
- Matson, J. L., & Minshawi, N. F. (2006). *Early intervention for autism spectrum disorders: A critical analysis*. Oxford, England: Elsevier Science Inc.
- Matson, J. L., & Senatore, V. (1981). A comparison of traditional psychotherapy and social skills training for improving interpersonal functioning of mentally retarded adults. *Behavior Therapy*, 12, 369–382.
- Matson, J. L., Sevin, J. A., Box, M. L., Francis, K. L., & Sevin, B. M. (1993). An evaluation and comparison of two methods for increasing self-initiated verbalizations in autistic children. *Journal of Applied Behavior Analysis*, 26, 389–398.
- Matson, J. L., Sevin, J. A., Fridley, D., & Love, S. R. (1990). Increasing spontaneous language in three autistic children. *Journal of Applied Behavior Analysis*, 23, 227–233.
- McEachin, J. J., Smith, T., & Lovaas, O. I. (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal on Mental Retardation*, 97, 359–391.
- Metz, B., Mulick, J. A., & Butter, E. M. (2005). Autism: A late 20<sup>th</sup> century fad magnet. In J. W. Jacobson, R. M. Foxx, & J. A. Mulick (Eds.), *Controversial therapies for developmental disabilities: Fad, fashion, and science in professional practice*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Mohr, C., & Sharpley, C. F. (1985). Elimination of self-injurious behavior in an autistic child by use of overcorrection. *Behavior Change*, 2, 143–147.
- Moore, V., & Goodson, S. (2003). How well does early diagnosis of autism stand the test of time? Follow-up study of children assessed for autism at age 2 and development of an early diagnostic service. *Autism*, 7, 47–63.
- Ollendick, T. H. (1983). Reliability and validity of the revised fear survey schedule for children (FSSC-R). *Behaviour Research and Therapy*, 21, 685–692.
- Rogers, S. J. (1998). Empirically supported comprehensive treatments for young children with autism. *Journal of Clinical Child Psychology*, 27, 168–179.
- Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *American Journal of Mental Retardation*, 110, 417–438.
- Schopler, E., Brehm, S. S., Kinsbourne, M., & Reichler, R. J. (1971). Effect of treatment structure on development in autistic children. *Archives of General Psychiatry*, 24, 415–421.
- Schopler, E., Mesibov, G. B., & Hearshey, K. (1995). Structured teaching in the TEACCH system. In E. Schopler, & G. B. Mesibov (Eds.), *Learning and cognition in autism* (pp. 243–268). New York: Plenum Press.
- Schreibman, L. (1988). *Autism*. Newbury Park, CA: Sage.
- Schreibman, L. (2000). Intensive behavioral/psychoeducational treatments for autism: Research needs and future directions. *Journal of Autism and Developmental Disorders*, 30, 373–378.
- Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service. [Interview Ed.]
- Stoelb, M., Yarnel, R., Miles, J., Takahashi, T. N., Farmer, J. E., & McCathren, R. B. (2004). Predicting responsiveness to treatment of children with autism: A retrospective study of the importance of physical dysmorphism. *Focus on Autism and Other Developmental Disorders*, 19, 55–77.

- Symes, M. D., Remington, B., & Brown, T. (2006). Early intensive intervention for children with autism: Therapists' perspective on achieving procedural fidelity. *Research in Developmental Disabilities, 27*, 30–42.
- Weiss, M. J. (1999). Differential rates of skill acquisition and outcomes of early intensive behavioral intervention for autism. *Behavioral Interventions, 14*, 3–22.
- Woodard, C., Groden, J., Goodwin, M., Shanower, C., & Bianco, J. (2005). The treatment of behavioral sequelae of autism with dextromethorphan: A case report. *Journal of Autism and Developmental Disorders, 35*, 515–518.