

Intensive Behavioral Intervention in a Community-Based Program for  
Children with Autism: A Retrospective Effectiveness Study

by

Lisa Marie Wright

A Thesis submitted to the Faculty of Graduate Studies of  
The University of Manitoba  
in partial fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

University of Manitoba

Winnipeg

Copyright © 2012 by Lisa Marie Wright

### Acknowledgements

I would like to express my gratitude to the children and staff participating in the St. Amant ABA Preschool Program: without them, this research would not have been possible. Furthermore, I would like to express my sincere appreciation to Dr. C.T. Yu (research advisor) and Dr. Garry Martin (academic advisor) for their guidance and encouragement during this project; as well as Drs. Angela Cornick, Dennis Hrycaiko, Adrienne Leslie-Toogood, and Adrienne Perry (external examiner) for their support and recommendations. I would also like to thank Dr. Leslie Thorne and Dr. Ian Clara for their invaluable feedback and suggestions, as well as Carly Thiessen for her assistance with the ABA database. Finally, I am forever grateful for the support and encouragement given by my immediate and extended family. Most especially, I would like to thank my husband, Paul, for his enduring love and support, and our son, Sam, who inspires me every day to succeed.

## Abstract

Early intensive behavioral intervention (EIBI), when delivered through small model programs, has been unequivocally established as an effective treatment for many children with autism. The next major research need is to evaluate the effectiveness of EIBI when delivered to large, community-based samples, since this is how the majority of children access treatment. The present study involved a retrospective analysis of archival data from children who have participated in the EIBI program at St. Amant, in Winnipeg, Manitoba. Similar to findings reported by Perry and colleagues (2008), results showed that children ( $n = 100$ ) achieved statistically and clinically significant reduction in autism symptom severity, as well as statistically and clinically significant improvement in language skills, adaptive behaviour skills, cognitive level, and rate of development after one year of intervention. Furthermore, when it came to achieving both statistically *and* clinically significant gains (including more “typical” or normative levels of functioning), children identified as “higher functioning” at intake outperformed “lower functioning” children on all standardized outcome measures. For the smaller subset of children ( $n = 50$ ) with data available at Intake, Year 1, and Year 2, statistically and clinically significant reductions in autism symptom severity were observed after two years of intervention, as well as statistically and clinically significant improvement in language skills, adaptive behaviour skills, and rate of development. Moderate gains in IQ were also observed, but these gains did not reach statistical significance, perhaps owing to sample size restriction for this measure ( $n = 11$ ). Across all outcome measures, children identified as “higher functioning” at intake were far more likely to achieve “typical” levels of functioning after two years of intervention. As for predictors of outcome at

Year 1, language and cognitive functioning at intake offered the strongest predictive value after one year of intervention. Finally, 13 children (13%) had achieved Best Outcome at Year 1; these children achieved notable gains on all outcome measures, and tended to be less severe diagnostically and much higher functioning at intake. Implications of these findings and directions for future research are discussed.

## Table of Contents

Acknowledgements.....	i
Abstract.....	ii
Table of Contents.....	iv
List of Tables.....	vi
List of Figures.....	vii
Introduction.....	1
Lovaas' 1987 Study and Follow-up.....	3
EIBI Evaluations with Less Intense Treatment than the Lovaas Study.....	7
EIBI Evaluations with Treatment Intensity Comparable to the Lovaas Study.....	9
EIBI in the "Real World": Efficacy vs. Effectiveness.....	18
Statement of the Problem.....	27
The Present Study.....	29
Hypotheses.....	30
Method.....	31
Participants and Settings.....	31
St.Amant Applied Behavior Analysis (ABA) Preschool Program.....	34
Measures and Frequency.....	36
Statistical Procedures.....	41
Results.....	44
Part A: Analysis of Outcomes for <i>Year 1 Sample</i> ( $n = 100$ ).....	45
Part B: Analysis of Outcomes for <i>Year 2 Sample</i> ( $n = 50$ ).....	49
Part C: Correlates and Predictors of Outcomes at Year 1.....	56
Hierarchical Models: Simultaneous Predictors of Outcomes.....	59

Best Outcome Children .....	60
Discussion.....	63
References.....	77
Appendix.....	84

## List of Tables

Table 1: Age, gender, and diagnosis for <i>Year 1 Sample</i> and the two subgroups .....	86
Table 2: Intake scores for <i>Year 1 Sample, Group A</i> (Higher functioning), and <i>Group B</i> (Lower functioning) .....	87
Table 3: Age, gender, and diagnosis for <i>Year 1 Sample</i> vs. <i>Year 2 Sample</i> .....	88
Table 4: Intake scores for <i>Year 1 Sample</i> vs. <i>Year 2 Sample</i> .....	89
Table 5: <i>Year 1 Sample</i> : Outcome measures at Intake (T0) and Year 1 (T1).....	90
Table 6: <i>Year 1 Sample</i> : Rate of Development at Intake and During Year 1 .....	91
Table 7: <i>Year 2 Sample</i> : Outcomes for children with scores at Intake, Year 1 and Year 2 .....	92
Table 8: <i>Year 2 Sample</i> : Rate of Development at Intake, During Year 1, During Year 2, and During 24 Months .....	93
Table 9: <i>Year 1 Sample</i> : Correlations of outcome variables with predictors .....	94
Table 10: <i>Year 1 Sample</i> simple regression: Variance accounted for by T0 predictor variables on dependent variables at T1 .....	95
Table 11: <i>Year 1 Sample</i> : Results for <i>Best Outcome Children</i> at Intake and Year 1 .....	96

## List of Figures

Figure 1: Mean Preschool Language Scale-4 Total Language Standard Scores at Intake and Year 1 for higher (Group A) and lower (Group B) functioning children .....	97
Figure 2: Mean Scales of Independent Behavior Broad Independence Standard Scores at Intake and Year 1 for higher (Group A) and lower (Group B) functioning children .....	98
Figure 3: Mean Pervasive Developmental Disorders Behavior Inventory Autism Composite Standard Scores at Intake and Year 1 for higher (Group A) and lower (Group B) functioning children .....	99
Figure 4: Mean WPPSI-III Full Scale IQ Standard Scores at Intake and Year 1 for higher (Group A) and lower (Group B) functioning children .....	100
Figure 5: Mean Preschool Language Scale-4 Total Language Standard Scores at Intake Year 1, and Year 2 for higher (Group A) and lower (Group B) functioning children .....	101
Figure 6: Mean Scales of Independent Behavior Broad Independence Standard Scores at Intake, Year 1, and Year 2 for higher (Group A) and lower (Group B) functioning children .....	102
Figure 7: Mean Pervasive Developmental Disorders Behavior Inventory Autism Composite Standard Scores at Intake Year 1, and Year 2 for higher (Group A) and lower (Group B) functioning children.....	103
Figure 8: Mean WPPSI-III Full Scale IQ Standard Scores at Intake, Year 1, and Year 2 for higher (Group A) and lower (Group B) functioning children .....	104



Intensive Behavioral Intervention in a Community-Based Program for Children with  
Autism: A Retrospective Effectiveness Study

Autism is an early-onset, sustained developmental condition marked by severe social dysfunction, communication difficulties, and a restricted repertoire of interests and activities (American Psychiatric Association, 2000). To meet the diagnostic criteria for autistic disorder (AD), the *Diagnostic and Statistical Manual of Mental Disorders, 4th edition – Text Revision* (DSM-IV-TR) requires a triad of deficits that include: (a) severe and ongoing impairments in social interaction; (b) marked developmental delays in communication; and (3) the presence of restricted, repetitive, and stereotyped behaviors. Under DSM-IV-TR, AD is one of five pervasive developmental disorders (PDDs), which include autistic disorder (AD), Asperger’s disorder, pervasive developmental disorder not otherwise specified (PDD-NOS), Rett’s disorder, and childhood disintegrative disorder. The first three listed are commonly referred to as the *autism spectrum disorders* (ASDs), due to their high degree of overlapping symptoms. The prevalence of ASDs has been estimated to be as high as 60 per 10,000 (1 in 166) individuals and approximately 70% of individuals with ASD also have a diagnosis of mental retardation (Fombonne, 2003).

For years, the prognosis for children diagnosed with autism or ASD was very poor, as professionals implemented a wide variety of psychological interventions that were largely unsuccessful, and in some cases, even harmful (Matson & Mishawaka, 2006; Perry & Condillac, 2003). However, at a time when autism and the more broadly defined ASDs were generally considered untreatable, Ivar Lovaas (1987) published a landmark outcome study that caused unprecedented excitement and optimism regarding the effectiveness of behavioral treatment for these children. Using an operant treatment

model that relied heavily on discrete-trials teaching, almost half (9 out of 19) of the children with autism who received early, intensive behavioral treatment (40 hours of 1:1 treatment per week, starting prior to age four, for a period of two or more years) achieved significant gains in intellectual functioning and completed first grade successfully in regular classrooms, without special instruction. More importantly, a follow-up study by McEachin, Smith, and Lovaas (1993) revealed that these same children had maintained their intellectual and academic gains several years later, and continued to function better than the original, minimally treated comparison children on measures of personality and adaptive behaviour.

Results from the Lovaas (1987) study and follow-up by McEachin et al. (1993) created much excitement in the field of autism intervention, providing new hope for these children and their families, and fueling a number of replications. As a result, there is now a substantial body of research showing that many children can make significant gains, even achieve some degree of “normalcy”, with early intensive behavioral intervention (EIBI; see reviews by Eikeseth, 2009; Eldevik et al., 2009; Howlin, Magiati, & Charman, 2009; Makrygianni & Reed, 2010; Perry & Condillac, 2003; Virués-Ortega, 2010). It seems abundantly clear at this juncture that the internal validity of EIBI, when delivered through small model programs, has been unequivocally established, relative to: (a) lower intensity behavioral treatment, (b) more “eclectic” community programs, and (c) high quality special education (see Perry et al., 2008). The next major step is to assess the effectiveness of EIBI delivered in less-than-ideal, “real world” conditions, since this is how most children actually receive treatment. Until recently, there were very few studies examining the effectiveness of EIBI in community settings, and these studies were

generally noted for their methodological shortcomings (e.g., Bibby, Eikeseth, Martin, Mudford, & Reeves, 2002; Sheinkopf & Seigel, 1998). More recently, however, Perry and colleagues conducted a large-scale study examining the effectiveness of publicly-funded EIBI in Ontario, Canada, and their results were very promising, suggesting that this treatment *can* be delivered effectively in the community, even when it involves “. . . less than ideal clients, less than ideal staffing and supervision procedures, and less than ideal treatment integrity” (p. 624). Given the extent to which EIBI is now being publicly funded in virtually all Canadian provinces, additional evaluations of EIBI delivered by service programs is needed to better understand the likelihood of meaningful, measurable progress for these children.

The sections that follow review, in some detail, the original findings reported by Lovaas (1987) and McEachin and colleagues (1993), and then critically examine the reported results from subsequent EIBI outcome studies, both in terms of (a) small program models, and (b) the (limited) available data from community-based programs. Following this review, results from the present study, which involved a retrospective analysis of archival data from children participating in the Applied Behavior Analysis Preschool Program (also referred to as the EIBI program) at St. Amant are presented in detail.

### **Lovaas’ 1987 Study and Follow-Up**

Building on prior research from the Young Autism Project at UCLA (Lovaas, Koegel, Simmons, & Long, 1973), a prospective study by Lovaas in 1987 was the first to report substantial improvement in functioning in a significant proportion of children receiving intensive, long-term, behaviour analytic intervention. The study compared

outcomes on two key variables, IQ and school placement, for three groups of autistic children: an experimental treatment group ( $n = 19$ ), a “minimal” treatment group ( $n = 19$ ), and a community treatment group ( $n = 21$ ). There were no significant differences across groups before treatment on eight measures (which included mental age, abnormal speech, self-stimulation, etc.), all children were less than 4 years old at intake, and all received treatment for at least two years.

The experimental group ( $n = 19$ ) received intensive treatment, which consisted of 40 hours per week of one-to-one, home-based, applied behavioural intervention. The control participants ( $n = 40$ ) consisted of Control Group 1 ( $n = 19$ ), receiving minimal treatment (10 hours per week or less) from UCLA staff, and Control Group 2 ( $n = 21$ ), receiving treatment through other community agencies not associated with the UCLA project. Lovaas (1987) noted the benefits of having two distinct control groups, in that data from Control Group 1 could be used to assess “spontaneous improvement” in very young autistic children, while data from Control Group 2 was useful in guarding against the possible confound of “favourable” or “unfavourable” referrals to the experimental group. It was also noted that participants were always assigned to the experimental group if there were enough staff members available to render treatment at the time of intake, thus resulting in a quasi-experimental design.

After two to three years of intervention, the children in the experimental group substantially outperformed those in both control groups. Children in the experimental group had gained an average of 31 IQ points over children in Control Group 1 (post-treatment mean IQ of 83 vs. 52), and had made other significant advances in educational achievement. Furthermore, nine of the 19 children (47%) receiving intensive treatment

had cognitive and language scores in the Average or High Average range on standardized tests of intelligence, and had completed Grade 1 successfully without special instruction. In contrast, the children who received either 10 hours per week of behaviour analytic treatment (Control Group 1) or treatment through typical community services (Control Group 2) made smaller gains in cognitive functioning (post-treatment mean IQs of 52 and 58, respectively), and only one child out of 40 (2.5%) was assessed to have reached Average intellectual functioning and normal first-grade placement at follow-up. Finally, Control Group 1 and Control Group 2 did not differ significantly at intake or at follow-up on standardized measures of IQ or educational placement. This last finding was notable, lending further support to the idea that gains made by the experimental group were, in fact, due to the intervention (and not likely the result of “spontaneous recovery”, placebo effects, or potential biases in the referral process).

McEachin et al. (1993) conducted a follow-up on the experimental group and Control Group 1 from the Lovaas (1987) study, with the goal of evaluating durability of treatment gains over time. Participants from Control Group 2 in the Lovaas study were referred from an outside agency and were not available for the follow-up investigation. McEachin and colleagues reported three main findings. First, participants in the original experimental group had maintained their level of intellectual functioning since their intake assessment at age 7, with their mean IQ continuing to be approximately 30 points higher than the control participants. Second, participants in the experimental group demonstrated significantly higher functioning than control participants on measures of adaptive functioning and personality. Finally, in a more in-depth analysis of the nine “best outcome” participants from Lovaas (1987), findings revealed that eight of the nine

children were comparable to average children on measures of intellectual functioning and adaptive behaviour, while one child had been removed from regular education classes and placed in a classroom for children with special needs. Thus, according to McEachin and colleagues, these eight children (42% of the experimental group) “. . . may be judged to have made major and enduring gains and may be described as ‘normal functioning’ ” (p. 368), whereas none of the control group participants had achieved such favourable gains.

Overall, these findings suggested that early, intensive behavioural treatment, when designed and implemented early and consistently for several years, could result in significant improvements in functioning for many children with autism. Moreover, treatment gains could be maintained over a prolonged period. These findings sparked a series of systematic replications to extend the generality of the observed treatment effects and to address some methodological limitations in the 1987 Lovaas study (e.g., potential sampling bias, small sample size).

The next two sections review several “partial replications” of the 1987 Lovaas study. The first section reviews studies with treatment applied at a lesser intensity than the Lovaas study; the second section reviews four studies with treatments delivered at a level of intensity comparable to the Lovaas study. The focus is on outcome studies with preschool children with ASDs participating in 1:1 behavioral intervention. As such, studies involving minimal behavioral treatment (Eldevik, Eikeseth, Jahr, & Smith, 2006), behavioural treatment in classroom settings (Fenske, Zaluski, Krantz, & McClannahan, 1985; Harris & Handlemann, 2000) or older children (Eikeseth, Smith, Jahr, & Eldevik, 2002; 2007) are not reviewed here.

### **EIBI Evaluations with Less Intense Treatment than the Lovaas Study**

There have been several published “partial replications”, all without the use of “aversives” that were used in the 1987 Lovaas study (e.g., Anderson, Avery, Di Pietro, Edwards, & Christian, 1987; Birnbrauer & Leach, 1993; Sheinkopf & Siegel, 1998; Remington et al., 2007; Smith, Groen & Wynn, 2000). In three of these replications (Anderson et al., 1987; Birnbrauer & Leach, 1993; Sheinkopf & Seigel, 1998), the participants received less intensive treatment (18-25 hours per week), and from less experienced personnel. These studies showed significant average increases in nonverbal IQ (range of 22 to 29 points), but gains in other areas, particularly adaptive functioning and school placement, were less substantial than those reported by Lovaas (1987).

Smith et al. (2000) were the first to publish a randomized clinical trial of EIBI. This study followed 28 children diagnosed with AD or PDD-NOS receiving either EIBI or parent training over a period of two to three years. The EIBI group ( $n = 15$ ; 7 with AD, 8 with PDD-NOS) averaged 25 hours of treatment per week over the first year, with treatment intensity fading over the next one to two years. In contrast, the comparison group ( $n = 13$ ; 7 with AD, 6 with PDD-NOS) received 10-15 hours of special education classes per week, as well as five hours per week of parent training for three to nine months. At follow-up, results showed that the EIBI group had significantly higher scores on measures of intellectual, language, visual-spatial, and academic functioning, but not on measures of adaptive functioning or behavior problems. They also found that the least impaired children (in this study, those with PDD-NOS) made greater gains than those diagnosed with AD. Furthermore, the gains achieved were less substantial than in the original Lovaas study. For example, the IQ difference between groups at follow-up was

16 points (compared to the 31 points reported by Lovaas), and only 4 of the 15 children (27%) had achieved an IQ above 85 and were attending regular classes.

More recently, Remington et al. (2007) conducted a 2-year controlled comparison study, where they examined EIBI vs. standard treatment for autism within the United Kingdom education system. The EIBI group ( $n = 21$ ) received home-based, 1:1 teaching from trained tutors and parents for 18 to 34 hours per week ( $M = 25.6$  hrs;  $SD = 4.8$  hrs), and the comparison group ( $n = 23$ ) received “treatment as usual” for autism provided by the local education authorities, none of which were delivered 1:1 or intensively in nature. Assessment of outcomes occurred at intake, Year 1, and Year 2 for both groups. While the two groups did not differ on any assessment measure at intake, findings revealed that there were “robust” statistically significant differences in favor of EIBI after two years on measures of intelligence, language skills, positive social behavior, adaptive functioning, and a statistical measure identifying “best outcome” children. Results showed significant gains in language for the EIBI group at Year 1 and again at Year 2, whereas changes in IQ and positive social behavior (e.g., initiating and responding to joint attention) were established by Year 1 and then maintained through to Year 2. Furthermore, after 2 years of treatment, 5 of the 21 children in the EIBI group (23.8%) were identified as “best outcome”, defined by the authors as a reliable and clinically significant change in IQ score (see Kazdin, 2005). Note that three of the 23 comparison group children (13%) also met this criteria, but another 3 children in this group experienced a statistically significant regression (e.g., loss of more than 24 IQ points), whereas none of the EIBI children had significantly regressed.



Overall, results from these studies showed that a number of children diagnosed with autism or ASD can achieve notable gains in intellectual, academic and adaptive functioning with EIBI, although in most of these studies, fewer children reached the average levels of functioning reported by Lovaas and colleagues (with the exception of Remington et al., 2007). Arguably, however, the experimental treatment in these studies differed from the original UCLA model (most notably, children in these studies had consistently received fewer treatment hours per week).

### **EIBI Evaluations with Treatment Intensity Comparable to the Lovaas Study**

Four recent studies have evaluated EIBI where intensity was comparable to the 1987 Lovaas study. The first compared a clinic-directed and a parent-directed EIBI program. Sallows and Graupner (2005) randomly assigned 24 children diagnosed with autism (not ASD) to one of two treatment groups: (a) a clinic-directed group ( $n = 13$ ), designed to replicate Lovaas model and following the parameters of the EIBI developed at UCLA, and (b) a parent-directed group ( $n = 10$ ), also based on the Lovaas model, but involving slightly less intensive treatment hours and less staff supervision (by personnel that were comparably trained). Both groups were matched on pre-treatment IQ and other key variables. A primary purpose of the study was to assess whether the Lovaas model could be successfully implemented in a community clinic setting with a comparable population of autistic children. Other key research questions included whether it was possible to accurately predict outcomes from specific pre-treatment variables, and whether children achieving post-test scores in the average range continue to show significant residual symptoms of autism. Finally, the study evaluated the effectiveness of a less costly parent-directed model.

After four years of treatment, performance measures on cognitive, adaptive, social, and academic functioning were similar for both the clinic- and parent-directed treatment groups (an unexpected finding). Furthermore, when combining the results for all participants, findings revealed that 11 of the 23 children (48%) demonstrated “rapid learning” early in treatment, had achieved IQ scores  $\geq 85$  (middle of the Low Average range), and by age 7, were doing well in regular classroom settings (without additional support), where they demonstrated average academic abilities, spoke fluently, and were included regularly in play activities with peers. In terms of predictors of outcome, results suggested that the best predictors were pre-treatment imitation, language, and social responsiveness. Although the 11 “best outcome” children (rapid learners) showed much larger improvements than the other 12 children (moderate learners), all children showed marked improvements in most skill areas over the four years of treatment. Notably, the average Full Scale IQ for all 23 children increased from 51 to 76 (25 points), approaching the 31-point increase in IQ reported by Lovaas (1987) and McEachin and colleagues (1993).

As for residual symptoms of autism among the rapid learners, some symptoms did remain, although generally not clinically significant. Approximately one-third of the children were identified as having “mild delays in social skills” as rated by teachers on the Vineland Socialization subscale (Sparrow, Balla, & Cicchetti, 1984) and the Child Behavior Checklist (Achenbach, 1991a, 1991b), and three of the 11 children were assigned aides because they “. . . needed reminders to stay on task” (Sallows & Graupner, 2005, p. 433). These residual symptoms are consistent with the results reported by McEachin and colleagues (1993), who found that, despite some clinically significant

scores in one or two areas, children were still succeeding in regular classes, six years after treatment was completed. With regards to the comparative effectiveness of a less costly parent-directed model, a surprising finding was that parent-directed children did as well as the clinic-directed group. The parent-directed group received six hours per month of supervision, and consultation every two months by the clinic supervisor or the senior author, whereas the clinic-directed group received six to 10 hours a week of supervision in the home from a senior therapist and weekly consultations with the clinic supervisor or senior author. However, all children received intensive treatment based on the Lovaas model for at least the first two years, and the parent-directed group received much more supervision than other outcome studies involving “parent managed” or “workshop” supervision (Bibby et al., 2002; Sheinkopf & Siegel, 1998). These results suggested that parents who are highly motivated and well-educated in the treatment methods *can* potentially co-ordinate and deliver EIBI effectively to their children at a much-reduced treatment cost.

Overall, the findings from this study were consistent with the results reported by Lovaas and colleagues (Lovaas, 1987; McEachin et al., 1993), even with respect to the “best outcome” children. Treatment intensity in this study was closer than in any previous replication, averaging 38 hours per week over two years for the clinic-directed group (and over 31 hours for the parent-directed group), and findings were the most consistent with the original Lovaas study. Limitations noted by the authors included (a) the possibility of assessment bias (most of the pre- and post-testing of the moderate learners was conducted by the second author), (b) use of different IQ measures at intake vs. follow-up, making pre-post test comparisons more difficult (although statistical analysis showed that an

effect on IQ attributable to the use of different testing instruments was unlikely), and (c) small sample sizes ( $n = 13$  and  $n = 11$ ), limiting the power of statistical tests to detect differences. Sallows and Graupner (2005) acknowledged that the use of “. . . many tests on such a small sample increased the likelihood of spurious findings, thereby limiting the implications of results for the larger population of children with autism” (p. 434).

The second evaluation of EIBI that involved treatment parameters similar to the original Lovaas study compared three treatments for preschool-age children with ASDs (Howard, Sparkman, Cohen, Green & Stanislaw, 2005). This study was unique in that outcomes for an experimental group ( $n = 29$ ) receiving traditional EIBI (1:1 adult-child ratio, 35+ hours per week) were compared to two other treatment groups: one group ( $n = 16$ ) receiving intensive “eclectic” intervention (combination of methods, 1:1 or 1:2 adult-child ratio, 30 hours per week) in early public intervention school programs, and a second group ( $n = 16$ ) receiving less-intensive, early public intervention within the school system (combination of methods, 1:6 adult-child ratio, 15 hours per week). The authors noted that although the available published studies “. . . offer compelling evidence that many children with autism who received early intensive behaviour analytic treatment make substantial gains” (p. 5), there is little research evidence available regarding the efficacy of “eclectic” interventions (publicly funded early intervention programs using a combination of methods) that are widely available to young children with autism. The three groups were assessed as comparable on key variables at intake, and standardized tests of cognitive, language, and adaptive functioning were administered by independent examiners to all children at intake and again after approximately 14 months of treatment.

At follow-up, findings revealed that children who received EIBI outperformed comparable children in the other two eclectic treatment groups in every skill domain assessed, and these differences were statistically significant for all domains (with the exception of motor skills). In contrast, there were no statistically significant differences in mean standard scores between the two eclectic treatment groups. This was a notable finding, as it showed that “eclectic” treatment (in the case of the first comparison group, a detailed program involving a combination of the non-behavioural analytic models, sensory integration therapy, and some applied behaviour analysis techniques) was not very effective, even when provided early and intensively “. . . in classrooms specifically designed for children with autism by staff with considerable training and experience with the population” (p. 14). Furthermore, in terms of learning rates, findings revealed that children in the EIBI group showed substantially higher rates of skill acquisition than the children in the other two treatment groups. In fact, during the 14 months of treatment, children in the EIBI group “. . . acquired skills in most domains at a rate that matched or exceeded the normal rate of one year of development per year of age” (p. 13), whereas learning rates remained well below normal for the vast majority of children in the comparison groups (and in many cases, reflected losses rather than gains).

Overall, these findings supported previous research suggesting that children who undergo EIBI can make substantial gains in areas of cognitive, language, and adaptive functioning. Although gains in several areas were smaller than have been reported in other studies (e.g., Lovaas, 1987; Sallows & Graupner, 2005), these other studies measured post-treatment gains after two or more years of EIBI, rather than just 14 months. Limitations noted by the authors included parent-determined group assignment

(although groups were similar on key dependent measures before treatment), lack of treatment integrity measures, and the reality that examiners, although independent of the investigators and all three treatment programs, were not blind to group assignment at follow-up assessment, introducing the possibility of measurement bias.

The third EIBI evaluation that attempted to “fully replicate” the parameters of the 1987 Lovaas study was a 3-year prospective study conducted in a community setting (Cohen, Amerine-Dickens, & Smith, 2006). The authors highlighted the importance of replicating previous successes with EIBI in the community, as this is ultimately where most children with ASD will receive treatment. The study compared two treatment groups: (a) an experimental group ( $n = 21$ ) that received EIBI (35 to 40 hours per week) from a community agency replicating the Lovaas treatment model, and (b) an age- and IQ-matched comparison group ( $n = 21$ ) obtained from special education classes at various local public schools. The design was quasi-experimental, as group assignment was based on parental preference. Standardized measures of cognitive, language, and adaptive functioning were administered by independent examiners to all children at intake, and repeated annually for three years, for a total of four repeated measures per child.

Findings revealed that, after the three years of treatment, the group receiving EIBI had obtained significantly higher IQ and adaptive behaviour scores than the comparison group. With regards to IQ (the primary outcome measure), the experimental group had gained an average of 25 points, compared to a mean gain of 14 points for the comparison group. Similar results were obtained on measures of adaptive behaviour. Although the experimental group showed greater gains in language comprehension, the differences between the two groups on expressive language and nonverbal cognitive skill were not

statistically meaningful. As for school placement, 6 of the 21 children in the experimental group were fully integrated without support at Year 3, and 11 others were integrated with varying levels of support. In contrast, one child from the comparison group was removed from special education and placed primarily in a regular classroom setting.

Overall, results from this study were encouraging, suggesting that EIBI based on the Lovaas model can be successfully executed in a community-based setting. However, Cohen and colleagues (2006) acknowledged that “. . . the difference between the EIBI group and the comparison group on outcome measures was smaller than that in other studies, as the comparison group also made gains” (p. 152). Several limitations of their study were noted, which may have resulted in smaller observed differences between the two groups. For example, although scores on pre-treatment measures did not differ significantly across groups, the EIBI group did have more children diagnosed with AD and fewer diagnosed with PDD-NOS. Given that PDD-NOS is often a milder diagnosis with a more favourable prognosis (Smith, Groen, & Wynn, 2000; see Perry, Condillac, Freeman, Dunn-Geier, & Belair, 2005), this difference may have benefited outcomes for the comparison group. This being said, the EIBI group may have had an advantage when it came to family variables: the EIBI group had better educated parents and more dual-parent families, perhaps making it easier for families to participate in treatment and to generalize skills beyond treatment sessions. Given these pre-existing group differences, and given the relatively small sample size in this study, the authors acknowledged that random assignment would have allowed for stronger conclusions regarding the relative efficacy of EIBI. Another key limitation cited by the authors was that the comparison group received such diverse interventions that it was not possible to monitor treatment

fidelity. As such, children in this comparison group may have been receiving more of the key behavioural components of EIBI than comparison groups in other studies, resulting in greater treatment gains. Yet another concern was that the criteria for Autism Spectrum Disorder (ASD) were not rigorously assessed throughout the study. The authors noted that inclusion of an autism screening measure at pre-treatment would have increased confidence in the initial diagnosis, and administration of this measure at yearly follow-up assessments would have confirmed whether or not participants continued to meet criteria for ASD.

Finally, the most recent study examining EIBI based on the Lovaas treatment model examined outcomes for children with autism (not ASD) receiving intensive UCLA ABA treatment in the United Kingdom (Hayward, Eikeseth, Gale, & Morgan, 2009). Two different service models were compared: (a) an intensive clinic-based treatment model ( $n = 23$ ), with all treatment personnel provided, and (b) an intensive parent-managed treatment model ( $n = 21$ ), where high levels of supervision were provided but tutors were recruited and monitored by parents. Average intensity of treatment for the clinic-based group was 37.4 hours per week ( $SD = 3.47$ ) versus 34.2 hours per week for the parent-managed group ( $SD = 5.29$ ). All children were between the ages of 24 and 42 months at intake (mean ages of 35.7 and 34.4 months, respectively), and both groups were comparable on all outcome measures at entry. Standardized tests of intelligence, visual-spatial skills, language, and adaptive skills administered by independent examiners to all children at intake and again after 12 months of treatment. In addition, this study included a non-concurrent multiple baseline design for the first 13 children entering treatment, to assess whether progress was associated with EIBI versus confounders (e.g.,



maturation effects, reaction to being assessed, or possible impact of prior interventions). The children were randomly assigned to baseline periods of 6, 12 or 18 weeks, during which time they received no EIBI intervention, and they were assessed every three weeks on receptive and expressive language and vocal/non-vocal imitation (see Early Language Measure, or ELM, in Smith et al., 2000).

At Year 1 follow-up, results showed that children in both treatment groups had achieved significant gains in the areas of cognitive functioning (including a measure of visual-spatial IQ), language skills, and adaptive functioning. Furthermore, consistent with findings reported by Sallows and Graupner (2005), there were no significant differences between the clinic-based and parent-managed groups on any of the outcome measures. When combining the results for all children, mean IQ had increased by 16 points, with half of the children achieving gains  $\geq 15$  points, and the vast majority (89%) achieving some level of improvement in cognitive functioning. As for results from the multiple baseline design, findings confirmed that children made little if any progress before treatment started, but once EIBI was implemented "... marked improvement was demonstrated on items in all four skill domains" (p. 621), suggesting gains could be attributed to EIBI. Finally, the most consistent predictor of outcomes was visual spatial IQ, "...as it predicted follow-up IQ, visual spatial IQ, language comprehension, expressive language and adaptive behaviour, as well as changes in IQ and adaptive behaviour" (p.630). Although intake IQ was correlated with outcome on these variables, it was not predictive of *change scores* from intake to Year 1. Although some previous studies had implicated intake age as an important predictor of outcomes (e.g., Fenske et

al., 1985; Harris & Handlemann, 2000), in this study, age at entry was not correlated with treatment outcomes or gains made during treatment.

Overall, these findings are consistent with results from earlier studies examining EIBI based on the Lovaas model (Cohen et. al, 2006; Howard et al., 2005; Sallows & Graupner, 2005), showing that children made substantial gains on all measures of interest after one year of intervention. Furthermore, similar to the findings reported by Sallows and Graupner, results suggested that the clinic-based and parent-managed models were equally effective, which has implications for increased access to treatment (particularly for rural families seeking outreach services). Strengths of this study included independent and blind assessment of the children, and inclusion of treatment integrity measures to ensure that EIBI was reliably administered. Furthermore, a unique contribution was that, with the inclusion of single-case experimental control, children's progress could be more confidently attributed to EIBI versus other possible confounding variables. Limitations noted by the authors included smaller sample size, which may have affected the power to detect important predictors of outcome (beyond visual-spatial IQ), and the lack of an alternative-treatment or no-treatment comparison group.

### **EIBI in the “Real World”: Efficacy vs. Effectiveness**

Barlow (1996) provides a clear distinction between efficacy and effectiveness research. *Efficacy* refers to “the results of a systematic evaluation of the intervention in a controlled clinical research context. Considerations relevant to the internal validity of these conclusions are usually highlighted” (p. 1051). In contrast, *effectiveness* refers to “the applicability and feasibility of the intervention in the local setting where the treatment is delivered”, in order to “determine the generalizability of an intervention

with established efficacy” (p. 1055). In short, efficacy research focuses on internal validity and replicability, whereas effectiveness research emphasizes external validity and generalizability (see discussion in Nathan, Stuart, & Dolan, 2000).

Clearly, there is much compelling evidence demonstrating the *efficacy* of EIBI, when delivered through small model programs, in treating children diagnosed with an ASD. This has been unequivocally established relative to: (a) lower intensity behavioral treatment, (b) more “eclectic” community programs, and (c) high quality special education. But can it work in the “real world”, where treatment program adherence may be more varied, where children have more severe diagnoses or more co-morbid conditions, and where families are not chosen based on resources (e.g., financial, education) that might correlate with better outcomes (see Perry et al., 2008)? To what degree can EIBI for children with ASDs be implemented successfully under these less-than-ideal conditions, without the support and resources of a university research centre to train treatment staff, provide appropriate supervision, and ultimately help to defray the immense cost of such an intensive treatment model (see Sallows & Graupner, 2005)? This speaks to the *effectiveness* of an intervention: that is, whether an intervention is robust enough to work in “real life” treatment settings. At present, there are only a handful of available studies examining the success of EIBI under such conditions. This section reviews the limited research examining outcomes for children participating in community-based EIBI. Three of the studies reviewed are based on smaller samples (Bibby, et al., 2002; Boyd & Corley, 2001; Sheinkopf & Seigel, 1998), and their findings have suggested less favorable outcomes than reported from the small-model efficacy studies. The fourth study, however, represents the first published outcome study for ASD

children participating in a large, publicly-funded EIBI program ( $n = 332$ ; Perry et al.), and their results suggest that EIBI *can* be effectively delivered in a community setting. The methodology and findings of this latter study are reviewed in some detail.

**Community-based EIBI in smaller samples.** Sheinkopf and Siegel (1998) conducted a controlled EIBI effectiveness study, where they compared outcomes for 11 preschool children participating in a home-based, parent-managed treatment program to a matched control group attending regular preschool. The children were matched on key variables such as age (both mental and chronological), intake IQ, diagnosis, sex, and time elapsed. After a mean duration of 16 months with an average of 26 hours per week of 1:1 treatment, the children receiving EIBI had better outcomes than the control group (who attended, on average, 11 hours per week of regular school). More specifically, there were notable differences in post-treatment IQ between the two groups (approximately 25 points), and modest improvements in symptom severity. However, the majority of the children still met diagnostic criteria for an ASD at follow-up.

Boyd and Corley (2001) presented findings from an outcome survey examining the effects of a community-based EIBI program. Data was available for twenty-two children, most under four years of age at entry, who had received a mean of 23 months of EIBI (range: 9 – 36 months). Results from parent surveys and retrospective file review revealed that all children had improved to some degree, and seven of the children (32%) were judged to be of “normal intellect” following intervention (although standardized assessment data was not available). However, none of the children in this study met the Lovaas (1987) criteria of “recovered” or “best outcome”, given that no child was able to participate in a regular classroom setting without 1:1 support at follow-up. The authors

noted that these children were older at intake than the Lovaas experimental group, and may also have been lower functioning at the outset of treatment.

Bibby and colleagues (2002) examined outcomes for 66 children participating in a parent-managed, self-funded EIBI program in the United Kingdom. Findings revealed modest but significant gains in adaptive functioning and speech and behavior pathology, but no meaningful group change in cognitive functioning (IQ scores). Furthermore, only 6% of the children were comparable to Lovaas' (1987) "best outcome" children (i.e., average IQ, integration into regular classrooms without support) upon exit from the program. However, compared to the Lovaas study, many children were older at intake (57%); the program was home-based and parent-managed, introducing a higher degree of variability in the quantity and quality of treatment; and the supervisors and consultants, most of whom would not have met UCLA training standards, only met with children approximately four times a year.

**EIBI in a large, community based sample.** To examine the feasibility of success in a large, community-based program, Perry and colleagues (2008) examined outcomes for children with ASDs in a large, publicly-funded EIBI program in Ontario, Canada. Two broad research questions were addressed. First, did the children show statistically significant and clinically significant changes on developmental and diagnostic measures from the time of entry to the time they were discharged? Secondly, what was the range of progress seen upon exit from the program and how many children resembled the "best outcomes" reported previously in the literature?

For this study, assessment data were available for 332 children participating in nine EIBI programs across different regions of Ontario, Canada. Most children had a

DSM-IV diagnosis of AD (58%), with the remainder diagnosed with either PDD-NOS (14%) or a less specific diagnosis of PDD/ASD (28%). The majority of the participants were male (5:1 ratio), mean age at entry into the program was 4.5 years (range: 20-86 months), and mean duration of treatment was approximately 18 months (range: 4-47 months). For each child, two data points were available on all assessment measures: one within three months of program entry, and another upon exit. Although the core components of EIBI treatment were similar for all programs, this sample was relatively heterogeneous when it came to treatment settings and the specifics of the curriculum.

Children were classified, a priori, into one of three Initial Level of Functioning groups, based on their Vineland Adaptive Behavior Composite (ABC) Standard Scores (adjusted to exclude motor skills). This resulted in three mutually exclusive subgroups: Group A (26%), a (relatively) higher functioning group (ABC scores above 59); Group B (42%), an intermediate functioning group (ABC scores 50-59); and Group C (32%), the lowest functioning group (ABC scores under 50). Statistical analyses showed significant differences across the three Initial Level of Functioning groups on all key measures at intake (i.e., severity of autism, adaptive functioning, cognitive level, and developmental rate), thus supporting the use of these subgroups when examining outcomes within the overall sample.

In terms of core measures, *autism severity* was assessed using the Childhood Autism Rating Scale (CARS; Schopler, Reichler, & Renner, 1988), which gives a Total Score with a trichotomous classification: not autism, mild/moderate autism, and severe autism. *Adaptive functioning* was assessed using the VABS survey form (Sparrow et al., 1984), a semi-structured interview conducted with parents by a trained interviewer.

*Developmental rate* was calculated using the VABS ABC Age equivalent scores. Initial Development Rate was obtained by dividing the Age Equivalent (AE) score by the child's chronological age, whereas Development Rate During EIBI was calculated by subtracting Intake AE from Exit AE, and then dividing the difference by the duration of treatment. Perry and colleagues noted that this measure of change is based on a number of questionable assumptions: namely, that development prior to treatment had been linear, that development during treatment was linear, and that two data points (entry/exit) were sufficient to obtain a slope.

For *cognitive level*, data from a number of different measures was used, most commonly: the Mullen Scales of Early Learning (Mullen, 1995), the Bayley Scales of Infant Development (Bayley, 1993), the Wechsler Preschool and Primary Scales of Intelligence (WPPSI-III, or in some cases, the WPPSI-R; Wechsler, 2002), and the Stanford-Binet Intelligence Scale, Fourth Edition (Thorndike, Hagen, & Sattler, 1986). For the WPPSI and the Stanford-Binet, a Full Scale IQ (FSIQ) was calculated, as well as a Mental Age (MA). However, for the Mullen and Bayley, a MA score was obtained and then used to calculate a Ratio IQ, or IQ estimate ( $MA/CA \times 100$ ). As such, scores from different tests were combined and referred to as "cognitive level", and different measures were sometimes used at intake vs. exit (more than a third of children with cognitive scores at both time points). The authors acknowledged that this was regrettable from a methodological perspective (see Matson, 2006), but noted that these types of concessions are often unavoidable in large, community-based samples.

Findings revealed that the children in the study showed statistically significant and clinically significant changes on developmental and diagnostic measures from the

time they entered the program to the time they were discharged. There were statistically and clinically significant improvements in all core areas assessed, both within individuals and across the three Initial Level of Functioning groups. First, with regards to autism severity, results indicated that the total sample, as well as all three subgroups, showed statistically significant improvement on the CARS (i.e., mean decreases in Total Score). Furthermore, clinically significant improvement in symptom severity was evidenced by changes in CARS categories (not autism, mild/moderate autism, and severe autism): at discharge, 50% of the total sample had improved categories, and 75% of the severe children had moved to a milder category (15% were now in the non-autism range, and 59% had improved to the mild/moderate range).

With regards to adaptive behavior, findings revealed that the children had gained significant skills during treatment in all areas of adaptive behavior. Statistically significant increases in Age Equivalent scores were noted on all subscales across the three subgroups. Standard scores revealed a similar overall outcome, but different patterns were noted: the “higher functioning” group (Group A) showing substantial increases in functioning on all subscales (with the exception of Daily Living Skills), while the lowest functioning group (Group C) showed no change on Communication or Socialization, and decreases in Daily Living Skills and ABC scores. As for rate of development, notable changes in developmental trajectory were observed, with all subgroups (approximately) doubling their Initial Developmental Rate. Specifically, Group A more than doubled (2.5 times) their initial rate, with over half (53%) showing rates of development equal to or above a typical rate after treatment. Group B showed a similar proportional change (approximately double the initial rate), with 30% of the



children functioning equal to or above a typical developmental rate upon exit. Finally, Group C also evidenced a similar proportional change (2.5 times their initial rate) during treatment, with 15% of the children in this group functioning equal to or above a typical rate upon exit.

As for cognitive level, findings revealed gains in cognitive functioning that were both statistically and clinically significant. Overall, there was a statistically significant increase in IQ estimate of approximately 12 points, a more modest gain relative to the average increase of 20 points observed in the previous efficacy studies. However, Group A (which was higher functioning and more comparable to samples in studies in small model programs) did show a similar increase of approximately 20 points. In contrast, Group B showed a significant, but less substantial increase of approximately 10 points, and Group C did not show a change in IQ, although a significant increase in MA (19 vs. 25 months) was observed. In terms of *clinically significant* gains in IQ (defined by the authors as an increase of at least 15 standard points), results showed that 39% of the total sample made gains of 15 standard points or more, and 17% of the total sample evidenced gains of 30 standard points or more. Only three children (2%) scored lower at exit, and the remaining children (59%) had scores within 15 points of their initial IQ. The authors also defined clinically significant gains in MA as an increase greater than 12 months over the course of treatment, and findings revealed that almost two-thirds of children (61%) met this criterion, with approximately one-third of the total sample (32%) showing gains of 24 months or more. Only one child showed a decline in MA, while the remaining children (38%) were similar to their initial levels upon exit.

In terms of the range of progress seen upon exit from the program, the results were (as expected) quite heterogeneous. Based on the available data (e.g., CARS Total scores and corresponding subgroups, VABS AE and standard scores, cognitive standard scores, and developmental rates), seven mutually exclusive categories were generated (a priori) and labeled as follows: (a) Average functioning (10.8%), (b) Substantial improvement (14.5%); (c) Clinically significant improvement (30.4%); (d) Less autistic (10.5%); (e) Minimal improvement (8.4%); (f) No change (18.6%); and (g) Worse (6.8%; see Perry et. al for specific criteria). Overall, most of the children (75%) showed some gains during the program, and approximately 11% reached Average functioning, similar to “best outcome” children in the literature. With regards to the three subgroups, more than half of the children in Group A achieved very favorable outcomes (Categories a and b above), and another quarter were assessed to be clinically significantly improved (Category c). In contrast, Groups B and C showed a wider range of outcomes: for Group B, there were some children in every category upon exit (including five assessed to be Average functioning), and similar heterogeneity was observed in Group C (although none reached Average functioning).

In a recent study, Perry et al. (2011) examined possible predictors of outcomes for the same children in their 2008 study, and found that while all four predictors of interest (age at entry, IQ, adaptive functioning, and autism severity) were important to some extent, IQ at intake was (by far) the strongest predictor of outcomes, accounting for significant unique variance (5-12%) in every regression they conducted. They also found that children who were most successful in the program (those who achieved *Average Functioning*,  $n = 32$ ) were significantly younger than the remaining children at entry

(mean of 41.71 months vs. 54.75 months), had higher rates of cognitive and adaptive functioning and milder autism ratings of autism severity at entry, and tended to remain in treatment for longer periods (mean of 25.92 months vs. 17.70 months). In contrast, the children with poor outcomes (those who fell in the *No change* or *Worse* categories;  $n = 75$ ) were not appreciably different than the remaining children in terms of intake scores or diagnostic characteristics at entry. Finally, it should be noted that Freeman and Perry (2010) reported on outcomes and predictors for the subset of children participating in the Toronto Preschool Autism Service (the largest of the nine publically-funded regional programs included in the 2008 provincial study). Findings for this smaller subset of children ( $n = 89$ ) were comparable to those reported in Perry et al. (2008) and Perry et al. (2011).

### **Statement of the Problem**

To date, randomized control studies (Remington et al., 2007; Sallows & Graupner, 2005; Smith et al., 2000) and matched-control group studies (Howard et al., 2005; Cohen et al., 2006) have provided compelling evidence of the efficacy of EIBI for preschool-aged children with ASDs when delivered and supervised by university-based researchers. Researchers have emphasized the need for additional EIBI outcome research, particularly in “real world” community settings, where treatment is most often delivered (Cohen et al., 2006; Perry et al., 2008; 2011).

In Canada, EIBI programs are now government funded to varying degrees in virtually all provinces, and evaluation of program effectiveness is one way to provide much-needed fiscal accountability. Such evaluations will also provide insights on the robustness of EIBI in “less-controlled” treatment environments, and help to identify

program variables that may be critical to EIBI success. Potentially critical variables could then be further researched, which will ultimately inform and improve future delivery of EIBI programs.

The St. Amant Applied Behavior Analysis Program for Children with ASDs has offered a home-based EIBI program based on the Lovaas model for a number of years. The program has collected a rich set of data for a large number of children. First, the overall sample size is quite large ( $n = 100$  for the *Year 1 Sample*), smaller only to the Perry et al. provincial study (2008). Second, the St. Amant data set contains standardized measures of intellectual, language, adaptive functioning, and ASD symptom severity. Third, measures were taken for each child annually over a three-year period. This will provide a better understanding of the developmental trajectories of these children relative to previous studies with fewer data points and varying treatment durations. Fourth, the program is delivered in relatively homogeneous treatment settings. Fifth, the larger data set allowed for examination of possible predictors of treatment outcomes at *Year 1*, which is often limited in studies with smaller samples.

A commonly cited criticism of EIBI research to date is that many studies have lacked a comparison group, or if a comparison group is included, participants are not randomly assigned, resulting in a quasi-experimental design (Matson, 2006; Kasari, 2002). In fact, Perry and colleagues (2008) recognized their lack of a comparison group as a key limitation to their research. Clearly, comparison groups are needed if we are to judge whether one treatment is superior to another. However, if an important goal is to gain additional information about whether a treatment is sufficiently effective to accelerate the developmental rates of children with ASDs, then arguably, a well-

conducted, retrospective study that does not include comparison group data would still contribute valuable information about treatment outcomes for these children. At this stage, we still have much to learn about the durability or “robustness” of EIBI when delivered within a community service model. Studies that have both strengths and limitations should be included for others to critique, provided that the limitations are acknowledged and compensated for whenever possible, and provided that the information reported contributes to our available knowledge base and appropriately directs future research endeavors.

### **The Present Study**

The purpose of the current study was to conduct a retrospective analysis of archival data from children who have participated in the early intervention preschool program at St. Amant (referred to as the Applied Behavior Analysis Preschool Program), in Winnipeg, Manitoba. In particular, we were interested in outcomes for children with assessment data at yearly time intervals (Intake, Year 1, and Year 2). Although some children participated in the program for three years, there was not enough data at the time of analysis to examine outcomes for children at Year 3. Furthermore, we were interested in whether children designated as “higher functioning” would make greater gains than those designated as “lower functioning” in five main areas of interest: language skills, adaptive functioning, autism symptom severity, cognitive functioning, and rate of development. Finally, we were interested in whether certain intake variables (e.g., age at entry, cognitive level, language skills) might predict outcomes for children, and whether certain intake variables were associated with “best outcome” results at Year 1.

## Hypotheses

In terms of hypotheses, four main findings were expected. The first hypothesis was that children participating in EIBI at St. Amant would achieve statistically and/or clinically significant reduction in autism symptom severity, and statistically and/or clinically significant improvement in cognitive level, language skills, and adaptive functioning. More specifically, it was predicted that children identified as “higher functioning” (Group A) at intake would make greater gains on outcome measures than those identified as “lower functioning” (Group B) at intake (based on global adaptive functioning standard scores at entry; see Perry et al., 2008).

The second hypothesis was that EIBI treatment offered through St. Amant would significantly accelerate the developmental rates of most children in the program. More specifically, it was expected that rate of development would increase significantly during treatment as compared to rate of development prior to treatment, and that during EIBI, the greatest gains would be made during the first year (see Howard et al., 2005, for a discussion of “learning rate”).

The third hypothesis was that skill level at entry would be positively correlated with skill acquisition in the later stages of treatment (Sallows & Graupner, 2005; Howard et al., 2005). Furthermore, it was expected that certain key variables at entry (e.g., cognitive level, language skills, age) might predict which children would benefit most during intervention.

The final hypothesis was that, in keeping with findings reported by Perry and colleagues (2008), approximately 10-15% of children in this study would resemble the “best outcome” children described in the literature after one year of treatment (i.e.,

cognitive and/or language skills approaching the Average range, as well as significant reduction in autism severity and significant increase in adaptive skills). Furthermore, it was expected that higher scores at intake (e.g., higher IQ and language skills) and possibly lower age at entry would be associated with reaching this level of outcome.

## Method

### Participants

The available archival data set consisted of 217 children who were served, in some capacity, by the St. Amant ABA Preschool Program between 2003 and 2011. Note, however, that approximately one-third of these children ( $n = 71$ ) were excluded from analysis because data were missing either at Intake or at Year 1. Furthermore, an additional 21% of children ( $n = 46$ ) with Intake and Year 1 data were not included for analysis because they did not have an adaptive functioning standard score at intake, and therefore could not be designated as “higher” or “lower” functioning at entry. As such, one hundred children ( $n = 100$ ) met criteria for analysis at Year 1 (i.e., intake standard score on the SIB-R and Intake/Year 1 data on at least one other outcome measure), representing just under half (46%) of all children served by the program during this time period. Although many children did not meet criteria for analysis, it should be noted that the children who did meet criteria for analysis (hereafter referred to as *Year 1 Sample*) remained representative of all children served in terms of age at entry, ratio of males vs. females, and diagnosis; furthermore, intake scores on all outcome measures for *Year 1 Sample* were virtually identical to those for all children served (see Tables A and B in Appendix). Since inclusion at the outset was not based on entry status, progress during the first year, or any other known factors, and given these comparable values, the

assumption was made (going forward) that *Year 1 Sample* is representative of all children served by the program during this time period.

**Year 1 Sample.** Age, gender, and diagnosis for the *Year 1 Sample* ( $n = 100$ ) are presented in Table 1 (first column). This sample had a mean age of 46.75 months (range: 26-63 months) at intake, with 85 (85%) boys and 15 (15%) girls. Diagnoses included autism/AD (43%), autism spectrum disorder (54%), and PDD-NOS (3%). All children had received an independent diagnosis of AD, ASD, or PDD-NOS by a psychiatrist, psychologist, or family physician prior to starting the program. Information regarding possible co-morbid conditions was not available for analysis.

In order to examine potentially different outcomes for children based on initial levels of adaptive functioning, children were classified as either “higher functioning” (Group A) or “lower functioning” (Group B), defined *a priori* based on their SIB-R Broad Independence Standard Scores (SIB-R BI SS) at intake. Specifically, children were designated as relatively “higher functioning” if their SIB-R BI SS at intake was 60 or above, and designated as relatively “lower functioning” if their SIB-R BI SS was 59 or lower. Adaptive functioning scores have been used in previous research as a grouping variable, with a similar cut-off point for identifying “higher functioning” children. For example, Perry et al. (2008) used a similar strategy for classifying their sample into one of three *Initial Level of Functioning Groups*, defined *a priori* based on their Vineland Adaptive Behavior Composite (ABC) Standard Scores (excluding motor skills) at intake. Note that the SIB-R BI SS scale has excellent concurrent validity with the Vineland ABC scale (Sparrow et al., 1984), with identical distributions ( $M = 100$ ,  $SD = 15$ ) and similar principal domains. Furthermore, unlike most standardized measures, the SIB-R has an



extended lower range of standard scores (i.e., a child can obtain a standard score as low as 1), thus allowing for more accurate differentiation among children who are very low functioning (see discussion in Wells, Condillac, Perry, & Factor, 2009).

Age, gender, and diagnosis for both subgroups are summarized in Table 1, and intake status on all outcome measures for both subgroups is summarized in Table 2.

These comparisons revealed that, while Group A and Group B were almost “matched” in terms of age and gender ratios, there was a higher number of children with the more severe diagnosis of Autistic Disorder in the lower functioning group (56% in *Group B* vs. 22% in *Group A*, Table 1). Furthermore, results from independent *t*-tests indicated that mean comparisons between *Group A* and *Group B* for all outcome measures were statistically significant ( $p < .01$  for PDDBI Autism Composite Standard Score and  $p < .001$  for the remaining measures, Table 2).

**Year 2 Sample.** Fifty children ( $n = 50$ ) from the *Year 1 Sample* had data at all three time points (Intake, Year 1, and Year 2), allowing for analysis over a 24-month treatment period (hereafter referred to as *Year 2 Sample*). Age, gender, and diagnoses for *Year 2 Sample* are summarized in Table 3 (second column), and intake scores for all outcome measures for *Year 2 Sample* are summarized in Table 4 (second column). While age remained comparable with *Year 1 Sample*, the gender ratio was now considerably lower (3.5:1, Table 3), and there was a slightly higher percentage of children with the more severe diagnosis of Autistic Disorder (AD; 52% vs. 43%). Furthermore, intake status on all outcome measures was lower in *Year 2 Sample* relative to *Year 1 Sample*. Overall, results suggests that the children with three data points are generally more

delayed (i.e., have greater skill deficits) at intake relative to children with only Intake and Year 1 scores.

### **St. Amant Applied Behavior Analysis (ABA) Preschool Program**

The St. Amant ABA Preschool Program for children with ASDs provides 31 hours per week of one-to-one instruction for up to three years. Parents are trained and supervised to implement a minimum of 5 additional hours per week. Since intensity and consistency are key characteristics of the effectiveness of EIBI for children with ASDs, family commitment is a key component for success. For example, families are required to limit their time away from the program to three planned weeks per calendar year (two weeks in the summer and one week during the December holiday season).

The ABA Preschool Program uses the *Assessment of Basic Language and Learning Skills* (ABLLS; Partington & Sundberg, 1998) as their skill development curriculum, and individualizes it for each child based on his or her strengths and weaknesses. Intervention is designed to progress gradually and systematically from relatively simple tasks, such as responding to basic requests made by an adult, to more complex skills, such as dressing, responding to requests that have multiple components, and appropriate social interaction. Intervention is based on applied behavior analytic procedures derived from operant learning principles (see Martin & Pear, 2011).

**Personnel.** A Director and a Clinical Coordinator head up the program, both of whom have a doctorate in Applied Behavior Analysis. The Director is a registered psychologist in Manitoba and the Clinical Coordinator is certified by the Behavior Analyst Certification Board. They oversee program operations and supervise Consultants, Senior Tutors, and Tutors.

Consultants are responsible for assessing skills, programming, training, analyzing outcomes, and supervising team members. Consultants are behaviour analysts with a Masters degree or Doctoral degree, either complete or in progress, in Applied Behaviour Analysis. They are supervised closely by the Director and the Coordinator.

Senior tutors assist the Consultants in their tasks and train other team members. They also conduct a shift with each child once every two weeks. Senior Tutors have completed two courses on basic principles and applications of behaviour analysis, and receive ongoing on-the-job training and supervision, with a minimum of 6 hours per week of direct contact with Consultants.

Tutors conduct 31 weekly hours of one-to-one behavioural intervention with each child, implementing the programs designed by the Consultant. Tutors participate in an initial training session of seven hours, in addition to pairing up with more experienced tutors for one month prior to starting independent sessions. They also receive ongoing on-the-job training and supervision, with a minimum of six hours per month of direct contact with Consultants and 12 hours per month of direct contact with Senior Tutors.

**Program activities.** Consultants are assigned to each child for approximately 4.5 hours per week. Senior Tutors are assigned to each child for approximately 9 hours per week. These hours include direct contact with children and parents, clinical and administrative meetings, and professional development.

Clinic meetings are led by the Consultant every two weeks, for 3 hours, for each child, and all members of the child's intervention team can attend. For children living in rural areas, the clinic meeting occurs monthly; however, the same number of hours is allotted to children living in or outside the city. According to the child's needs and the

clinical activities that are prioritized, the Consultant sometimes conducts shorter clinic meetings or skips a clinic meeting. During that time he or she will work on other activities to benefit the child's program, such as conducting assessments, writing progress reports, and participating in school and review meetings. During clinic meetings, team members review the child's successes, programming needs, and difficulties that have arisen in the previous weeks. Some meeting time is spent taking turns working with the child to directly observe skills being taught and to assess programming needs. Team performance is also reviewed at each meeting.

### **Measures and Frequency**

In addition to the data collected by tutors through direct observation during teaching sessions, the St. Amant ABA Program administers a battery of norm-referenced or criterion-referenced measures, many of which are standardized. All measures are administered at intake and then yearly during the program. For the present study, five main areas of interest were examined: (1) language skills, (2) adaptive functioning, (3) autism symptom severity, (4) cognitive functioning, and (5) rate of development. In order to assess outcomes in the first four areas, the composite (or total) standard score was examined from each of the following standardized instruments: (1) *Preschool Language Scale, 4th edition* (PLS-4; Zimmerman, Steiner, & Pond, 2002), (2) *Scales of Independent Behavior-Revised* (SIB-R; Bruinicks, Woodcock, Weatherman, & Hill, 1996), (3) *Pervasive Developmental Disorders Behavior Inventory* (PDDBI; Cohen & Sudhalter, 2005), and (4) *Wechsler Preschool and Primary Scale of Intelligence, 3rd edition* (WPPSI-III; Wechsler, 2002). In addition, *rate of development* pre-intervention and *rate of development* during each year of treatment was calculated (based on SIB-R

age equivalence scores relative to chronological age). Each measure is described in more detail below.

**Preschool Language Scale–Fourth Edition (PLS-4).** The PLS-4 is a norm-referenced, individually administered test used to identify children who have a language disorder or delay (Zimmerman et al., 2002). This measure consists of two subscales: Auditory Comprehension and Expressive Communication. The Auditory Comprehension subscale is used to evaluate how much language a child understands, including skills that are important precursors for language development such as attention to speakers and appropriate object play, as well as comprehension of basic vocabulary, concepts, grammatical markers, complex sentences, and the ability to make comparisons and inferences. The Expressive Communication subscale is used to determine how well a child communicates with others including vocal development, social communication, vocabulary, preposition use, grammatical markers, sentence structures as well as pre-literacy skills such as phonological awareness, ability to tell a short story, and use of language to define words.

Scores from the Auditory Communication and the Expressive Communication subscales are combined to yield a Total Language Score. For all three composites, a standard score of 69 or less indicates significant deficits in comparison to the normative sample of same-aged peers. The PLS-4 has been shown to have high test-retest reliability, ranging between .82 and .95 for the subscale scores, and between .90 and .97 for the Total Language Score. Internal consistency reliability coefficients were also very high (Cronbach's alpha = .86, .91, and .93 for the Auditory, Expressive, and Total Language Scores, respectively), and the test shows strong overall inter-rater reliability (99%).

Additionally, the PLS-4 has been demonstrated to have strong test-criterion (predictive) validity and is highly correlated with other measures of receptive and expressive language ability.

**Scales of Independent Behavior--Revised (SIB-R).** The SIB-R is a norm-referenced, structured interview that provides a measure of a child's functional independence and adaptive behaviours (Bruininks, Woodcock, Weatherman, & Hill, 1996). A Broad Independence Standard Score (BI SS) is calculated based on four main areas of functioning: motor, social interaction and communication, personal living, and community living skills. Standard scores for BI and the adaptive behavior clusters have a mean of 100 and standard deviation of 15, with scores of 69 and below indicating significant deficits. There are three different forms available (Full Scale, Short Form, Early Development Form), all of which provide a BI SS and assess frequency and severity of problem behaviours (the latter was not examined in the present study). Note that the SIB-R is not a measure of a child's potential; rather, it provides an overview of the child's current strengths and challenges from the point of view of the person(s) being interviewed. Thus, results reflect the respondent's judgment regarding the level of independence on specific adaptive skills and frequency/severity of several challenging behaviours. The SIB-R has been demonstrated to have test-retest reliabilities in the .70s and .80s, and inter-rater reliabilities are reported to range from the .60s to the .90s (Sattler, 2002). Furthermore, the SIB-R has been shown to have excellent concurrent validity with the VABS (Sparrow et al., 1984), a measure which is generally considered the "gold standard" for adaptive behaviour assessment in preschool children with developmental disabilities (Msall, 2005).

**Pervasive Developmental Disability Behavior Inventory (PDDBI).** The PDDBI (Cohen & Sudhalter, 2005) measures autism symptom severity and is a norm-referenced rating scale developed for clinical, educational, and research applications. It is standardized with a population diagnosed with PDD, and is intended as a tool for assisting in the diagnosis and intervention of this population. The PDDBI–Parent Rating Form is a comprehensive questionnaire where parents rate a child’s repertoire across 10 domains, divided into two sections: Approach/Withdrawal Problems, and Receptive-Expressive Social Communication Abilities. Scores from these two sections are combined to yield an Autism Composite score. Results are expressed in standard T-scores, with a mean of 50 and a standard deviation of 10. For the autism composite, scores between 40 and 60 are considered typical of most children with PDD in the normative sample, scores below 40 indicate fewer challenges in comparison to the normative sample, and scores above 60 indicate greater challenges in comparison to the normative sample. The PDDBI–Parent Rating Form has been demonstrated to have strong internal consistency (range of coefficient alpha = .80 to .98 across all domains and composite scores), strong long-term test-retest reliability (over a 12-month interval, coefficients ranged from .38 to .91, with only two of the 15 coefficients below .50), strong construct validity, and good criterion-related validity when compared with components of other comparable scales, such as the CARS and the VABS (Cohen, Schmidt-Lackner, Romanczyk, & Sudhalter, 2003).

**Wechsler Preschool and Primary Scale of Intelligence–3rd Ed. (WPPSI-III).**

The WPPSI-III is a standardized, individually administered assessment of general intellectual ability of young children, designed to assess both verbal and non-verbal

abilities (Wechsler, 2002). Verbal abilities assessed include verbal reasoning and comprehension, acquired knowledge, and attention to verbal stimuli; non-verbal abilities assessed include fluid reasoning, spatial processing, attentiveness to detail, and visual-motor integration. The Full Scale score is derived from these verbal and nonverbal subscales, and is presented as an index of the child's general cognitive functioning at the time of the assessment. The WPPSI-III has very strong psychometric properties, including high internal consistency/reliability at both the composite scale level (range: .89 - .96) and the individual subtest level (range: .83 - .95), high inter-rater reliability (overall agreement rate: .98 - .99), and high correlation with other standardized tests of intelligence for preschoolers (Gordon, 2004).

**Rate of Development.** In addition to the above measures, rate of development pre-intervention was compared with rate of development during intervention, in order to isolate rate of growth occurring between Intake and Year 1 (and between Year 1 and Year 2; see Perry et al., 2008). *Rate of Development at Intake* was calculated by dividing the age equivalent (AE) score at intake by the child's age at entry. For example, a 24-month-old child who began the program with an AE score of 12 months would have an initial development ratio of 0.50. *Rate of Development During Year 1* was calculated by taking the difference between the Intake and Year 1 AE scores divided by the time interval between them. For example, if a 24-month old child with an AE score of 12 months at intake reached an AE score of 27 months after one year of EIBI, the rate of development during the first year would be  $(27-12)/12 = 1.25$  (see Howard et al., 2005). Examining rate of development during specific intervention periods allows for evaluation of both the magnitude of improvement in AE and the proportional rate of change that has



occurred (keeping in mind the key assumption that development prior to treatment had been linear). As such, the child in this example gained more than a year (15 months) in AE scores between Intake and Year 1, and more than doubled their rate of development after one year of intervention (2.5-fold increase from 0.50 to 1.25).

### **Statistical Procedures**

In order to assess outcomes for children at yearly time intervals and to examine possible predictors, we conducted Repeated Measures Analysis of Variance (RM ANOVA) models, as well as bivariate correlations and linear regression models. RM ANOVA was used to test the hypotheses that (1) children identified as relatively “higher functioning” at intake would make greater gains on outcome measures than children identified as “lower functioning” at intake, and that (2) for both groups, the greatest gains would occur during the first year of treatment. Bivariate correlations were generated to test the hypothesis that skill level at entry would be significantly related to skill acquisition at Year 1, and linear regression models were used to assess whether certain key variables at entry (such as FSIQ) might strongly predict outcomes after 1 year. A brief description of each type of analysis is provided below, including an examination of the principal assumptions that must be considered for each model.

**Repeated Measures ANOVA.** As with any ANOVA, the purpose is to test the equality of means between two or more groups. However, *repeated measures* ANOVA is used when all members of a random sample are measured under a number of different conditions. As the participants are exposed to each condition in turn, the measurement of the dependent variable is repeated, in this case, across different time points.

When a dependent (outcome) variable is measured repeatedly for all sample members across time points, this set of repeated conditions is called a *within-subjects factor*. When a dependent (outcome) variable is measured for independent groups of sample members, the set of independent groups is called a *between-subjects factor*. For the current study, the between-subjects factor consisted of the following groups: Group A (higher functioning) vs. Group B (lower functioning). Finally, when an analysis has both within-subjects factors (time) and between-subjects factors (group), it is referred to as *repeated measures ANOVA with between-subjects factors*.

When conducting RM ANOVA, there are two main assumptions made about the sample being analysed: (1) multivariate normality (i.e., the dependent variables are assumed to follow a multivariate normal distribution), and (2) multivariate homogeneity of variance (i.e., the assumption that the variance between groups is equal, or that the variance of the dependent variables is constant across cells formed by the between-subjects effects).

With regards to multivariate normality, ANOVA is robust in the face of most violations of this assumption if group sample size is not small (e.g.,  $n \geq 20$ ). When the sample size is smaller than  $n = 20$  per group, it is important to look at indicators such as skewness and kurtosis when assessing normality. Generally speaking, although values ranging from -1 to +1 are ideal, skewness values of -3 to +3 are acceptable and kurtosis values of -8 to +8 are acceptable (Kline, 2010). In our samples, looking at values of skewness and kurtosis for each outcome variable at all three time points, all values for skewness fall between +2.17 and -0.40, and all measures of kurtosis fall between +4.53 and -1.10 (with the vast majority of values falling between +0.9 and -1.10). As such, from

the perspective of multivariate normality, these numbers are acceptable to proceed with RM ANOVA.

With regards to the homogeneity of variance assumption (i.e., variance between groups is equal), this assumption can easily be tested with Levene's Test. If the test is statistically significant ( $p < .05$ ), this suggests that the variances between groups may not be equal, and the  $F$ -statistic with "equal variances not assumed" should be reported. In our sample, the assumption of homogeneity of variance was not violated in our analyses in most cases; when it was, the "corrected" statistic and accompanying  $p$ -value were reported.

Beyond these main assumptions, one final point bears mentioning. Ideally, in a RM ANOVA, the convention is  $n \geq 20$  cases per group, in order to be reasonably confident that the statistical results are interpretable. For the analysis of children with three time points (Intake, Year 1, Year 2), the 2 groups x 3 time points RM ANOVAs were run, in order to assess potential interaction effects (as opposed to conducting paired  $t$ -tests, which does not allow for testing of possible interactions). However, due to smaller sample sizes per group (most notably the "higher functioning" groups), the statistics generated need to be interpreted with caution (Howell, 2007).

**Linear Regression Models.** Linear regression is used to model the value of a dependent scale variable based on its linear relationship to one or more predictors. The case of one explanatory variable is called *simple regression*, whereas more than one explanatory variable is termed *multiple regression*. The linear regression model assumes that there is a linear, or "straight line," relationship between the dependent variable and each predictor. For the purpose of testing hypotheses about the values of the model

parameters, linear regression modeling also assumes the following: (1) independence of the errors (no serial correlation), (2) homoscedasticity (constant variance) of the errors, both versus time and versus the predictions (or versus any independent variable); and (3) normality of the error distribution. If any of these assumptions are violated (i.e., if there is nonlinearity, serial correlation, heteroscedasticity, and/or non-normality), then the predictions yielded by a regression model may be (at best) inefficient or (at worst) seriously biased or misleading (Howell, 2007).

In order to test for violations of these main assumptions, a number of steps were taken. For each outcome measure, the following were generated: (1) *autocorrelation plots* of the residuals, to examine independence of the errors (e.g., no serial correlation), (2) plots of *residuals versus time* and *residuals versus predicted values*, to examine variance and be alert for evidence of residuals that are getting larger (i.e., more spread-out) either as a function of time or as a function of the predicted value, and (3) *normal probability plot* of the residuals, to ensure that the points on this plot should fall close to the diagonal line. Examination of these plots revealed that all outcome measures met the necessary criteria for proceeding with linear regression modeling.

## **Results**

In order to summarize the numerous findings, results will be presented in three sections. Part A will examine to what extent children improved during their first year in the program (Intake vs. Year 1), looking specifically at group mean differences, as well as clinically significant changes on an individual level (see Kazdin, 2005). Part B will repeat this process for the smaller subset of children with three time points (Intake, Year 1, and Year 2). Finally, Part C will examine which intake variables best predict outcomes

for children at Year 1, and will present hierarchical regression models (simultaneous predictors of outcome) for cognitive level and language skills. An examination of the specific factors associated with achieving *Best Outcome* at Year 1 will be presented as well.

### **Part A: Analysis of Outcomes for Year 1 Sample ( $n = 100$ )**

For all outcome measures, comparisons of children's standard scores at Intake (T0) and Year 1 (T1) were conducted using SPSS Repeated Measures Analyses of Variance, with group (higher functioning vs. lower functioning) as the between-subjects factor and time (Intake vs. Year 1) as the within-subject repeated measure. For each measure (presented below), a child was included in the analysis *only* if they had a score at both time points. Main effects for time, as well as interaction effects (group x time) are reported. (Main effect for group was significant for all outcome measures at  $p < .002$ ). In addition, the number of children showing a *clinically significant* change at Year 1 (i.e., improvement  $\geq$  one *SD*) is reported for each outcome measure (Kazdin, 2005).

**Language skills (PLS-4 Total Language Standard Score).** For the 87 children who had a PLS-4 Total Language Standard Score (PLS-4 TL SS) at both Intake and Year 1, results from the 2 x 2 RM ANOVA showed a significant group x time interaction (see Table 5 and Figure 1). While both groups showed statistically significant increases in language skills at Year 1, Group A made more substantial gains in language skills (mean gain of 10.86 standard points) relative to Group B (mean gain of 3.38 standard points). As for *clinically significant* changes in language skills, 12 of the 34 children (35.3%) in Group A achieved an increase  $\geq 15$  standard points (i.e.,  $\geq$  one *SD*), whereas only 6 of the 53 children (11.3%) in Group B had achieved this magnitude of improvement. Overall,

the higher functioning group outperformed the lower functioning children at Year 1 (both in terms of pre- vs. post group means and percentage of children achieving clinically significant gains).

**Adaptive functioning skills (SIB-R BI SS).** For the 93 children who had a SIB-R BI SS at both Intake and Year 1, results from the 2 x 2 RM ANOVA revealed a significant main effect for time, and a non-significant group x time interaction (see Table 5 and Figure 2). In other words, the magnitude of improvement in adaptive functioning was similar for both groups (mean gain of 10.35 standard points for Group A and 13.75 standard points for Group B) across the 1-year period. As for *clinically significant* changes in adaptive functioning skills, 17 of the 37 children (45.9%) in Group A had achieved an increase  $\geq 15$  standard points (i.e.,  $\geq$  one *SD*), and 23 of the 56 children (41.1%) in Group B had achieved this level of change. All told, although both groups showed a similar *magnitude* of improvement, there were clear differences favoring the higher functioning group when it came to achieving more “typical” levels of adaptive functioning (e.g., levels comparable to same-age peers in the normative sample).

**Autism severity (PDDBI Autism Composite Standard Score).** For the 66 children who had a PDDBI Autism Composite Standard Scores (PDDBI AC SS) at both Intake and Year 1, results from the 2 x 2 RM ANOVA revealed a statistically significant main effect for time, with the group x time interaction approaching statistical significance (see Table 5 and Figure 3). While both groups showed statistically significant decrease in autism symptom severity at Year 1, Group A showed a more substantial decrease in symptom severity (mean decrease of 7.57 standard points) relative to Group B (mean decrease of 3.29 standard points). As for *clinically significant* changes in autism

symptom severity, 13 of the 28 children (46.4%) from Group A achieved a decrease  $\geq 10$  standard points (i.e.,  $\geq$  one *SD*), whereas 10 of the 38 children (26.3%) in Group B achieved this level of improvement. Overall, Group A outperformed Group B at Year 1 (both in terms of pre- vs. post group means and percentage of children achieving clinically significant decreases in autism symptom severity).

**Cognitive functioning (WPPSI Full Scale IQ Standard Score).** For the 42 children who had a WPPSI FSIQ SS at both Intake and Year 1, results from the 2 x 2 RM ANOVA revealed a statistically significant main effect for time, and a non-significant group x time interaction (see Table 5 and Figure 4). In other words, the magnitude of improvement in FSIQ SS was statistically significant and similar for both groups (i.e., mean gain of 7.78 standard points for Group A and 6.80 standard points for Group B) across the 1-year period. As for *clinically significant* changes in cognitive level, 9 of the 23 children (39.1%) in Group A had achieved an increase  $\geq 15$  standard points (i.e.,  $\geq$  one *SD*), whereas 4 of the 20 children (20%) in Group B had achieved such gains. As such, although both groups achieved a similar magnitude of improvement in mean FSIQ, there were clear differences favoring the higher functioning group when it came to achieving clinically significant gains and more “typical” levels of cognitive functioning (e.g., more comparable to same-age peers in the normative sample).

**Rate of development.** In addition to the four standardized outcome measures, children’s *Rate of Development at Intake* was compared with *Rate of Development During Year 1*, in an attempt to isolate rate of growth occurring specifically between Intake and Year 1 (see Table 6). Results from the RM ANOVA revealed a significant main effect for time, and a non-significant group x time interaction. At intake, Group A

started out with a mean pre-intervention developmental rate of .74; however, over 12 months of intervention, this had increased to a mean rate of 1.49 (more than doubling their initial rate of development in one year). As for Group B, they started out with a mean pre-intervention developmental rate of 0.39; however, over 12 months of intervention, the mean rate was 0.79 (also more than doubling their initial rate in one year). Although the magnitude of improvement was far greater for the higher functioning group, the proportional rate of change was similar for both groups. For both groups, this represents a significant and substantial alteration in developmental trajectory.

In terms of individual outcomes, 22 of the 37 children (59.5%) in Group A more than doubled their initial rate of development, with 9 of these children (24.3%) more than tripling their intake development rate. As for Group B, 24 of the 56 children (42.9%) more than doubled their intake development rate, with 15 children (26.8%) more than tripling their intake development rate. However, in terms of achieving a typical rate of development during intervention (1.0 being “typical”), 25 of the 37 children (67.6%) from Group A had achieved a development rate greater than 1.00 During Year 1, versus 20 of the 56 children (35.7%) from Group B. As such, while there were many children from both groups who showed marked increases in rate of development, children from the higher functioning group were far more likely to achieve a “typical” rate of development During Year 1 (67.6% vs. 35.7%, respectively).

**Part A: Summary.** As predicted, results showed that (on average) children achieved statistically and clinically significant reduction in autism symptom severity, as well as statistically and clinically significant improvement in language skills, adaptive behaviour skills, cognitive level, and rate of development during the first year of EIBI.



There was a wide range of outcomes at Year 1 across both groups, as evidenced by the tendency for *SD*'s to become larger after intervention (with the exception of PDDBI scores, where *SD*'s stayed relatively constant from Intake to Year 1). In terms of differences in group means, Group A made more substantial gains relative to Group B in the areas of language development and reduction in autism symptom severity, whereas both groups showed a similar magnitude of improvement in the areas of broad adaptive functioning and cognitive level. However, when looking at outcomes at Year 1 on an individual level, there were clear differences favoring Group A, on all measures, when it came to achieving clinically significant improvement and more "typical" levels of functioning (e.g., standard scores that were more comparable to same-age peers in the normative sample). As for rate of development, both groups more than doubled their intake rate of development during the first year of intervention, but the magnitude of improvement was far greater for Group A, with these children being far more likely to achieve a "typical" rate of development during Year 1 (67.6% vs. 35.7%, respectively).

**Part B: Analysis of Outcomes for Year 2 Sample ( $n = 50$ )**

For all measures, comparisons of children's scores at Intake (T0), Year 1 (T1) and Year 2 (T2) were conducted using SPSS Repeated Measures ANOVA, with group (higher functioning vs. lower functioning) as the between-subjects factor and time (Intake, Year 1, Year 2) as the within-subject repeated measure. A significant main effect for group ( $p < .012$ ) was observed for all outcome variables; main effects for time and interaction effects (group x time) are reported below. Note that for all post-hoc pairwise comparisons, Bonferroni correction was applied ( $p \leq .05/3$  or .017 required for statistical significance).

**Language skills (PLS-4 Total Language Standard Scores).** For the 31 children who had a PLS-4 Total Language SS (PLS-4 TL SS) at all three time points, results from the 2 x 3 RM ANOVA revealed a statistically significant main effect for time and a non-significant group x time interaction (see Table 7). Over a two-year period, Group A had achieved a statistically significant mean gain of 17.00 standard points ( $t = -3.10$ ,  $p = .017$ ), whereas Group B had achieved a non-significant mean gain of 5.91 standard points ( $t = -2.26$ ,  $p = .034$ ). For both groups, the greatest gains in PLS-4 TL SS were made in the first year, with Group A benefiting more from each year of intervention than Group B (see Figure 5). Furthermore, it appears that, on average, Group A children benefitted (to some degree) from a second year of intervention, with little (if any) additional benefit for Group B children. However, looking at each year individually, the mean gain of 12.12 standard points in the higher functioning group during the first year was the only one that reached statistical significance ( $t = -3.44$ ,  $p = .011$ ). As for *clinically significant* changes in language skills, four of the eight children (50%) in Group A had achieved an increase  $\geq 15$  standard points by the end of Year 2, whereas four of the 23 children (17.4%) in Group B had achieved this level of improvement. In short, there were clear differences favoring Group A when it came to achieving both clinically significant changes and more “typical” levels of language functioning (e.g., levels comparable to same-age peers in the normative sample). It should be noted that this measure has a high floor (i.e., basal SS of 50), so changes for children with minimal language skills, across both groups, may not have been captured.

**Adaptive functioning skills (SIB-R BI SS).** For the 42 children who had a SIB-R BI SS at all three time points, results from the 2 x 3 RM ANOVA revealed a

statistically significant main effect for time and a non-significant group x time interaction (see Table 7). Over a two-year period, Group B achieved a statistically significant mean gain of 19.84 standard points ( $t = -3.63, p = .001$ ), whereas Group A achieved a non-significant mean gain of 8.80 standard points ( $t = -1.57, p = .152$ ). For both groups, the greatest gains in SIB-R BI SS were made in the first year, with the *lower functioning* group benefiting more from each year of intervention than the higher functioning group (see Figure 6). Furthermore, it appears that (on average) Group B children benefitted, to some degree, from a second year of intervention, with little (if any) additional benefit for Group A children. This being said, looking at each year, the mean gain of 14.31 standard points in the lower functioning group during the first year was the only one that reached statistical significance ( $t = -3.28, p = .003$ ). As for *clinically significant* changes in adaptive functioning, four of the 10 children (40%) in Group A and 16 of the 32 children (50%) in Group B had achieved gains  $\geq 15$  standard points by the end of Year 2.

Although both groups had a similar proportion of children achieving clinically significant gains after two years, the group means at Year 2 show that there are clear differences favoring the higher functioning group when it comes to achieving more “typical” levels of adaptive functioning (e.g., levels comparable to same-age peers in the normative sample).

**Autism severity (PDDBI AC SS).** For the 26 children who had a PDDBI AC SS at all three time points, results from the 2 x 3 RM ANOVA revealed a statistically significant group x time interaction (see Table 7 and Figure 7). After two years of intervention, Group A had achieved a statistically significant mean decrease of 9.67 standard points ( $t = 5.89, p = .000$ ), whereas Group B showed a small mean decrease of

1.29 standard points ( $t = 0.62, p = .544$ ). Looking specifically within each year, Group A achieved a mean decrease of 7.56 standard points at Year 1 that approached statistical significance ( $t = 2.83, p = .022$ ), and achieved an additional mean decrease of 2.11 standard points at Year 2 that was not statistically significant. In contrast, mean changes in autism severity for Group B were small and non-significant after Year 1 or Year 2. As such, it appears that, on average, Group A children benefitted from the first year of intervention, with only a slight additional benefit during the second year; in contrast, scores for Group B tended to remain stable over the 2-year period. As for *clinically significant* changes in symptom severity, six of the nine children (66.7%) from Group A had achieved a decrease  $\geq 10$  standard points (i.e.,  $\geq$  one *SD*) by the end of Year 2, whereas only two of the 17 children (11.8%) from Group B had achieved a decrease of this magnitude. All told, results favored Group A when it came to achieving clinically significant decreases and more “typical” levels of functioning (i.e., fewer challenges compared to other children with PDD).

**Cognitive functioning (WPPSI-III FSIQ SS).** Intake, Year 1, and Year 2 FSIQ SS were available for 11 of the children at all three time points (see Table 7 and Figure 8). Ten of these children were from Group B, whereas only one child came from Group A. For the one higher functioning child, marked gains in FSIQ SS were made at Year 1 (gain of 37 standard points; moving from the *Extremely Low* range to the *Low Average* range). However, there was a subsequent decrease in score from Year 1 to Year 2 (loss of 6 standard points, moving down to the *Borderline* range). Overall, this child had made substantial gains over a 2-year intervention period, but the trajectory was not linear (i.e., large increase in the first year, followed by a small decrease).

For the ten lower functioning children, results from the RM ANOVA revealed a non-significant main effect for time (see Table 7). A non-significant mean gain of 6.70 standard FSIQ points was observed at Year 1 ( $t = -1.95, p = .083$ ), followed by a negligible increase in the second year (mean gain of 0.80 standard points,  $t = -0.19, p = .854$ ). In short, after two years in the program, these children had achieved a mean increase of 7.50 standard FSIQ points ( $t = -1.77, p = .110$ ); however, gains occurred almost exclusively within the first year of intervention, and then were maintained over the following year. As for *clinically significant* changes in FSIQ, three of the 10 children (30%) had achieved an increase  $\geq 15$  standard points by the end of Year 2, with all three children achieving gains  $\geq$  one *SD* within the second year of intervention. Only one child had achieved a FSIQ SS at Year 2  $\geq 85$ . In contrast, the remaining seven children showed no appreciable change in SS from intake.

**Rate of Development.** Rate of development at intake (pre-intervention) was compared with rate of development during each year of intervention, in an attempt to isolate rate of growth occurring between specific time points. Results from the RM ANOVA revealed a non-significant main effect for time, as well as a non-significant overall group x time interaction (see Table 8). However, further examination of the means showed differential gains across groups over time, some of which reached statistical significance. Looking specifically at the first year of intervention, Group A achieved a mean rate of development of 1.23 (mean gain of .57 from intake;  $t = -2.58, p < .001$ ), whereas the lower functioning group had achieved a mean rate of development of 0.64 (mean gain of .23 from intake;  $t = -1.06, p = .016$ ). In contrast, between Year 1 (T1) and Year 2 (T2), the lower functioning children had achieved a mean rate of

development of .76 (mean gain of .12 from T1;  $t = -2.39, p = .076$ ), whereas the higher functioning children showed a mean rate of development of .86 (mean loss of .41 from T1;  $t = 1.88, p = .002$ ). Clearly, rate of development during Year 1 was far greater for the Group A (1.23 vs. .64), almost double that of Group B, with the higher functioning group (on average) exceeding a typical rate of development in the first year. However, during the second year, rate of development was *similar* for both groups (.82 vs. .76). Also, it should be noted that a significant main effect for time *did* emerge when collapsing the results and looking at Rate of Development at Intake vs. Rate of Development During Two Years of Intervention (.47 vs. .78;  $t = -2.60, p = .013$ ). However, the group x time interaction remained non-significant (e.g., both groups achieved a similar magnitude of change during 24 months).

As for individual outcomes, six children (60%) from Group A and 11 children from Group B (34.4%) had rates of development at or above a typical rate ( $\geq 1.0$ ) and had more than doubled their intake rate of development during Year 1. Furthermore, five children (50%) from Group A and nine children (28.1%) from Group B had achieved development at or above a typical rate and had more than doubled their intake rate of development across *both* years of intervention. However, the most dramatic difference between groups lies in the children's *absolute development rate* measured at Year 2: four of the 10 children (40%) from Group A had nearly "caught up" to their same aged peers at Year 2 assessment (rate of development at T2  $\geq .9$ ), whereas only two of the 32 children (6%) from Group B had achieved this degree of improvement. All told, results again favored Group A children when it came to achieving a "typical" rate of development at Year 2 assessment.

**Part B: Summary.** For the smaller subset of children with three data points, results showed that (on average) children achieved statistically and clinically significant reduction in autism symptom severity, as well as statistically and clinically significant improvement in language and adaptive behaviour skills after two years of intervention. As with the *Year 1 Sample*, there was a wide range of outcomes across both groups, as evidenced by the tendency for *SDs* to become larger as intervention progressed (with the exception of the PDDBI, where variability in scores decreased after Year 1 for higher functioning children). Across all four standardized outcome measures, the greatest mean gains occurred during the first year of intervention, with a smaller degree of benefit occurring (in some cases) during the second year. Specifically, in the areas of language development and autism symptom severity, Group A benefitted more from each year of intervention than Group B, with a clear advantage favoring Group A children when it came to achieving both clinically significant changes and more “typical” levels of functioning at Year 2. However, in terms of adaptive functioning, mean comparisons showed that Group B benefitted more from each year of intervention than Group A, with both groups showing a similar proportion of children that achieved clinically significant gains after 2 years. However, the children in Group A were still far more likely to achieve “typical” levels of adaptive functioning (i.e., levels comparable to same-age peers in the normative sample) after two years of intervention.

As for cognitive functioning, there were not enough Group A children to perform a group x time analysis; however, the ten children from Group B had achieved a non-significant mean increase in FSIQ at Year 2 (again, with these gains occurring almost exclusively within the first year of intervention). Three of the ten children achieved

*clinically significant* changes in FSIQ by the end of Year 2, with all three children achieving gains  $\geq$  one *SD* within the second year of intervention, whereas the remaining seven children showed no appreciable change in SS from intake. Finally, rate of development during Year 1 was far greater for Group A (1.23 vs. .64), with Group A children, on average, exceeding a typical rate of development in the first year. However, during the second year of intervention, rate of development was similar for both groups (.82 vs. .76). All told, results again favored Group A children when it came to achieving a “typical” rate of development at Year 2 assessment (40% vs.6%).

### **Part C: Correlates and Predictors of Outcomes at Year 1**

In order to examine correlates and predictors of outcomes for children in the *Year 1 Sample*, the following analyses were performed: (1) correlation analyses, (2) simple linear regression modeling, and (3) hierarchical regression models (simultaneous predictors of outcome). In addition, correlates of outcomes for the children identified as *Best Outcome* were examined and compared to the remaining children from the *Year 1 Sample*. Results from each set of analyses are presented below.

**Correlation Analyses.** For all five outcomes measures, children’s scores at Year 1 were correlated with six main predictors at intake (using Pearson’s *r*): (1) Age at entry (months), (2) PLS-4 TL SS (3) SIB-R BI SS, (4) PDDBI AC SS, (5) FSIQ SS, and (6) Development Rate at Intake (SIB-R AE/chronological age). These predictors were selected because (with the exception of age and Rate of Development) they are all standardized measures, and these factors have all been highlighted in the literature as possible predictors of outcomes for children with ASD (see review in Perry et al., 2011).



Results from the correlation analyses are presented in Table 9. Findings showed that language, adaptive functioning, and cognitive scores at intake (T0) were strongly and significantly correlated with all five outcome variables at T1 ( $p < .01$  in all cases). In other words, children with higher language, adaptive functioning and FSIQ scores at intake tended to have better scores on all outcome measures at Year 1. Overall, the strongest correlations occurred with intake language and intake cognitive scores (range: .549 - .838). As for autism symptom severity, while the correlation between PDDBI AC SS at Intake and Year 1 was statistically significant at  $p < .01$  (i.e., children with a lower PDDBI AC SS at intake tended to score better on this measure at Year 1), correlations between PDDBI AC SS at intake and the remaining outcome variables at T1 were considerably smaller in magnitude (range: .078 - .275). Only one of the outcome variables at T1 (PLS-4 TL SS) was significantly related to autism severity at intake ( $p < .05$ ).

Although the correlation between Rate of Development at Intake vs. During Year 1 was *not* statistically significant, correlations between Rate of Development at Intake and the remaining four outcome measures were all statistically significant at  $p < .01$ . In other words, children with a higher Rate of Development at Intake tended to score better on measures of language, adaptive functioning, autism severity, and cognitive functioning at T1. In contrast, none of the outcome variables at T1 were significantly correlated with age at entry.

**Regression Analyses.** Next, simple linear regressions were computed for the five outcome measures at Year 1 (PLS-4 TL SS; SIBR-BI SS; PDDBI AC SS; FSIQ SS, and Rate of Development During Year 1), to see whether any of the six main predictor scores

at intake could account for unique variance in outcomes at Year 1 (over and above initial levels of the outcome variable in question). In keeping with the method of analysis used by Perry et al. (2011), the Intake (T0) score for the variable in question was entered first (as a method of controlling for it) and the  $R^2$  value was reported for the initial step. Next, the initial value of the predictor variable in question was entered (e.g., FSIQ at intake), to see whether it accounted for any additional variance (reported as  $R^2 \Delta$ ).

Results from the regression analyses are reported in Table 10. Looking at the values for FSIQ SS (T0) and PLS-4 TL SS (T0), it is evident that these two predictors account for a statistically significant amount of incremental variance for *all* outcomes at Year 1. Most notably, intake FSIQ SS and PLS-4 TL SS accounted for a statistically significant amount of incremental variance in Rate of Development During Year 1 (38.2% and 26.7%, respectively), as well as PDDBI AC SS at T1 (13.6% and 16.8%, respectively).

Looking next at adaptive functioning, results show that SIB-R BI SS at intake accounted for a statistically significant amount of incremental variance in Rate of Development During Year 1 (24.2%) and PDDBI AC SS at T1 (8.6%). However, unlike the two predictors mentioned above, SIB-R BI SS at intake did not account for a statistically significant amount of additional variance in language or cognitive scores at T1. As for autism symptom severity, while PDDBI AC SS at intake accounted for 31.3% of the variance in PDDBI AC SS at Year 1 (T1), it did not account for a statistically significant amount of incremental variance for any other outcome measure.

Finally, results showed that Development Rate at Intake accounted for 10.8% of the variance in PDDBI AC SS at Year 1, but did not account for a statistically significant

amount of incremental variance for any other outcome measure. And, consistent with the correlation analyses above, age at entry did not account for a significant amount of additional variance for any of the five outcome variables at Year 1.

### **Hierarchical Models: Simultaneous Predictors of Outcomes**

The final step was to examine which combination of predictors would account for the most variance in outcome for these children. Given that most studies have focused on outcome levels of IQ as the main benchmark of intervention success, a final regression analysis was conducted to predict FSIQ SS at Year 1. However, given the results from the simple linear regression analyses, which implicate FSIQ as a potentially powerful predictor of language skills, a final hierarchical regression analysis to predict language skills at Year 1 was conducted as well. The  $R^2$  values presented in Table 10 were used to determine order of entry for each outcome measure.

**Cognitive functioning at T1.** Given that intake cognitive and language skills were found to be important predictors of FSIQ at T1 when examined separately, a hierarchical regression was conducted, so that the role of initial cognitive and language levels could be examined in combination when predicting FSIQ at T1. FSIQ SS at T0 was entered first (accounting for 60.2% of the variance;  $p < .001$ ), followed by PLS-4 SS at T0, which accounted for an additional 4.2% of the variance ( $p = .041$ ). Inclusion of the remaining four intake variables in the model did not offer any additional predictive value that was statistically significant. The final  $R^2$  was .644,  $F(2,38) = 34.32$ ,  $p < .001$ , suggesting that 64.4% of the variance in Year 1 FSIQ SS can be predicted based on the combination of these two variables.

**Language skills at T1.** Intake language and cognitive skills were found to be significant predictors of language at T1 when examined separately. As such, PLS-4 TL SS at T0 was entered first (accounting for 66.7% of the variance;  $p < .001$ ), followed by FSIQ SS at T0, which accounted for an additional 11.2% of the variance ( $p < .001$ ). Inclusion of the remaining four intake variables in the model did not offer any additional predictive value that was statistically significant. The final  $R^2$  was .784,  $F(2,51) = 92.51$ ,  $p < .001$ , suggesting that 78.4% of the variance in Year 1 PLS-4 TL SS can be predicted based on the combination of these two variables.

### **Best Outcome Children**

In order to isolate factors that might be associated with achieving “best outcomes” as defined in the literature, a number of additional analyses were conducted. Children were identified as *Best Outcome* at Year 1 if they met the following criteria: (1) PLS-4 TL SS and/or FSIQ SS in the mid-Low Average range or better (i.e., within one standard deviation of the mean;  $SS \geq 85$ ); (2) SIB-R BI SS in the mid-Low Average range or better (i.e., within one standard deviation of the mean;  $SS \geq 85$ ); and (3) PDDBI AC SS below 40, indicating fewer challenges in comparison to the normative sample of children with PDD-NOS. From the *Year 1 Sample* ( $n = 100$ ), 13 children (13%) met this criteria. Twelve of the children (92%) came from the higher functioning group, and one child (8%) was from the lower functioning group. Four of the children did not have a PDDBI AC SS at Intake or Year 1; however, we were able to confirm that these children had achieved CARS scores at Year 1 that were in the “non-autistic” range.

Table 11 shows the results for *Best Outcome Children* at Intake and Year 1. These children achieved large, statistically significant improvements on all outcome measures.

In terms of language and cognitive functioning, standard scores were notably higher at Year 1 (mean increases of 14.08 and 12.46 points, respectively), and adaptive functioning standard scores had also improved significantly (mean increase of 23.15 points). PDDBI AC SS at Intake and at Year 1 were not available for six of the 13 children, but for the remaining seven children, results showed a significant mean decrease of 15.43 standard points in symptom severity. In addition, for these children, the mean rate of development During Year 1 had increased three-fold compared to intake (mean of 0.75 to a mean of 2.26), which is triple a typical rate of development.

All *Best Outcome* children were males and were similar in age to the remaining children in the *Year 1 Sample* (49.54 vs. 46.40 months); however, the *Best Outcome* group had a lower percentage of children diagnosed with AD compared to the remaining children (30.8% vs. 44.8%, respectively), and were much higher functioning on measures of language, adaptive functioning, cognitive level, and rate of development at intake ( $p \leq .001$ ). The *Best Outcome* children also scored lower on autism symptom severity at intake compared to the remaining children in the *Year 1 Sample* (mean of 46.86 vs. 52.64), but this difference did not reach statistical significance ( $p \leq .136$ ). Overall, it appears that *Best Outcome Children* tend to be less severe diagnostically and tend to be higher functioning at intake. Of note, only one *Best Outcome* child (from the higher functioning group) remained in the EIBI program for a second year of intervention; the remaining 12 children “graduated” at Year 1 or shortly thereafter (i.e., went on to participate in the school-age program, or no longer required services).

**Part C: Summary.** Results from the correlation analyses showed that higher standard scores on language, FSIQ, and (to a lesser degree) adaptive functioning were

associated with better scores on all outcome measures at Year 1. As for autism severity, there was a strong, positive correlation between PDDBI AC SS at Intake vs. Year 1, but only moderate, negative correlations with language skills and adaptive functioning skills at T1. Age at entry was not significantly correlated with any outcome variables at Year 1; however, all correlations between rate of development at intake and outcome measures were statistically significant (except with Rate of Development During Year 1). As for the regression models, results showed that language and cognitive scores at intake accounted for a statistically significant amount of additional variance for all outcome variables at Year 1. The final two-predictor model for FSIQ SS at Year 1 was statistically significant ( $R^2 = .644$ ), suggesting that 64.4% of the variance in Year 1 FSIQ can be predicted based on the combination of intake FSIQ followed by intake PLS-4 TL SS. Perhaps even more impressive, the final two-predictor model for language score at Year 1 was statistically significant ( $R^2 = .784$ ), suggesting that almost 80% of the variance in language at Year 1 can be predicted based on the combination of intake PLS-4 TL SS followed by intake FSIQ SS.

Finally, 13 of the 100 children from the *Year 1 Sample* (13%) met criteria for *Best Outcome* (i.e., language and/or FSIQ SS  $\geq 85$ , adaptive functioning SS  $\geq 85$ , and autism composite SS below 40 criteria). These children achieved statistically significant improvements on all outcome measures (see Table 11), with all children achieving clinically significant gains ( $\geq$  one *SD*) on at least one measure. Twelve of these children came from Group A (92%), whereas one child (8%) was from the lower functioning group. Relative to children who did not meet *Best Outcome* criteria, the *Best Outcome* children tended to be less severe diagnostically and much higher functioning at intake.

## Discussion

The purpose of the current study was to conduct a retrospective analysis of archival data from children participating in the early intervention preschool program at St. Amant, in Winnipeg, Manitoba. The first hypothesis was that children would achieve statistically and/or clinically significant improvement across all standardized outcome measures, and that children identified as relatively “higher functioning” at intake would make greater gains on outcome measures than those identified as “lower functioning”. As predicted, results showed that children in the *Year 1 Sample* achieved statistically and clinically significant reduction in autism symptom severity, as well as statistically and clinically significant improvement in language skills, adaptive behaviour skills, and cognitive level. Higher functioning children (Group A) made more substantial gains relative to lower functioning children (Group B) in language development and reduction in autism symptom severity, whereas both groups showed a similar magnitude of improvement in adaptive functioning and cognitive level. However, when it came to achieving both statistically *and* clinically significant gains (including achieving more “typical” levels of functioning), Group A outperformed Group B on all standardized outcome measures (bearing in mind that Group A was higher functioning at the outset).

Mean gains for the *Year 1 Sample* as a whole were relatively modest compared to those reported for children in the numerous efficacy studies (e.g., Lovaas, 1987; Sallows & Graupner, 2005; see review by Howlin et al., 2009). However, when looking specifically at the higher functioning children (Group A), who were similar to the samples in the efficacy studies and similar to the higher functioning children in Perry et al. (2008), some notable comparisons emerged. For example, Group A children achieved

a statistically significant decrease in autism symptom severity, as did the higher functioning children in Perry and colleagues. However, different measures were used (PDDDBI vs. CARS), making direct comparisons difficult. With regards to adaptive functioning, gains in the present study were similar to those reported by Perry and colleagues for the higher functioning children, whereas Group B children here also achieved statistically significant gains (mean gain of over 13 standard points at Year 1 in the present study vs. essentially no change or a slight worsening of score in their study). Note, however, that the VABS has a higher floor than the SIB-R, so changes for lower functioning children in the Perry et al. study may not have always been captured. Furthermore, it should be noted that we might expect lower functioning (Group B) children to make greater gains in adaptive functioning and fewer gains in cognitive and/or language development, as this may reflect priorities in programming based on different needs identified at intake.

As for cognitive functioning, Perry and colleagues reported a 21-point IQ gain for their higher functioning children (similar to the average reported in other efficacy studies), whereas Group A children here achieved a much smaller mean gain of 7.8 standard points. However, it should be noted that the present study reported on a much smaller sample of children, all receiving the WPPSI-III; in contrast, Perry and colleagues combined scores from different cognitive tests to create one “cognitive level” (and different measures were often used at intake vs. exit). All told, differences noted above between the two sets of “higher functioning” children may, in part, result from the fact that treatment duration varied in length in the Perry et al. study (range of 4 to 47 months,



with a mean of approximately 18-months), whereas in the present study, all children in the *Year 1 Sample* were assessed after one year in the program.

Finally, standardized measures of language skills have been included less often in these types of studies, and when they have been included, many children have been unable to attain a score, often owing to higher norms (see Remington et al., 2007). The present study is the first *effectiveness* study that has included a standardized language measure at both intake and yearly time intervals, representing an important contribution to the EIBI effectiveness literature. In the *Year 1 Sample*, there was a large mean gain in PLS-4 language skills for Group A children at Year 1 (mean gain of almost 11 standard points), whereas the gains for Group B were much smaller (mean gain of 3.4 standard points). Significant gains in language development for Group A are consistent with the results from efficacy studies that have included a language measure (e.g., Hayward et al., 2009; see Howlin et al., 2009). As mentioned previously, the PLS-4 has a high floor (i.e., basal SS of 50), so changes for children with minimal language skills, across both groups, may not have been captured.

In addition to the valuable contributions from the *Year 1 Sample*, the present study is the first *effectiveness* study to report on children with three data points taken at regular intervals (Intake, Year 1, and Year 2). While there are clear limitations due to low sample sizes (i.e., many children lacked scores at all three assessment points), this smaller subset of children offered a “first glimpse” of the developmental trajectory for children who remained in EIBI for two full years. Consistent with the first hypothesis, children in *Year 2 Sample* achieved statistically and clinically significant reductions in autism symptom severity, as well as statistically and clinically significant improvement

in language skills and adaptive behaviour skills. In terms of language development and autism symptom severity, Group A benefitted more from each year of intervention than Group B, whereas the reverse was true when it came to gains in adaptive functioning. However, when it came to achieving both statistically *and* clinically significant gains (including achieving more “typical” levels of functioning), Group A outperformed Group B on all three standardized outcome measures at Year 2 (again bearing in mind that Group A was higher functioning at the outset). The greatest mean gains occurred during the first year of intervention, with a smaller degree of benefit occurring (in some cases) during the second year. This trend towards a “plateau” after Year 1 is similar to findings reported in other studies with multiple data points (see review by Howlin et al., 2009).

Although there were not enough Group A children in the *Year 2 Sample* to conduct a group x time analysis for cognitive level, the ten children from Group B had achieved a mean increase of 7.5 FSIQ standard points at Year 2 (again, with gains occurring almost exclusively within the first year of intervention), with three of the ten children achieving clinically significant changes in FSIQ by the end of Year 2. The one child from Group A made marked gains in FSIQ SS at Year 1 (moving from the *Extremely Low* range to the *Low Average* range), followed by a modest decrease in SS between Year 1 and Year 2 (down to the *Borderline* range). Although results here are for a very small sample of children ( $n = 11$ ), findings are consistent with studies reporting IQ changes over three or more time points, where changes between baseline and first follow-up are generally larger than between first and second follow up (e.g., Lovaas, 1987; Cohen et al., 2006; Remington et. al., 2007; see review by Howlin et al., 2009).

Furthermore, given the sample size restriction for this measure, gains in FSIQ SS over two years may have reached statistical significance with a larger number of children available for analysis.

The second hypothesis was that EIBI offered through St. Amant would significantly accelerate the developmental rates of most children participating in the program. It was also predicted that during treatment, the greatest gains would be made during the first year (see Howard et al., 2005). As expected, findings for *Year 1 Sample* revealed that (on average) children in both groups more than doubled their intake rate of development during the first year of intervention. Furthermore, consistent with the findings noted above, the magnitude of improvement was much greater for Group A, with these higher functioning children far more likely to achieve a “typical” rate of development during Year 1 (67.6% vs. 35.7%, respectively). In the present study, both groups doubled their initial rate of development (similar to findings reported by Perry et al., 2008), but initial rates and rates during Year 1 here were notably higher, particularly for Group A children (.74 at intake vs. 1.49 during Year 1, compared to pre-post rates of .47 and 1.16 for their higher functioning group).

The first part of Hypothesis 2 was also supported in the *Year 2 Sample*, with children achieving statistically significant increases in developmental rate across two years of intervention (mean of .47 vs. .78 over two years for the total sample). Results showed that (on average) both Group A and Group B achieved a similar magnitude of change during 24 months of EIBI. However, looking specifically at the first year of intervention, rate of development during Year 1 was far greater for Group A (1.23 vs. .64), almost double that of Group B, with the higher functioning group (on average)

exceeding a typical rate of development in the first year. As for the prediction that the greatest gains would occur during Year 1, this was true for Group A children only (large gains in the first year followed by more modest gains during Year 2), whereas the rate of development for Group B was comparable in each year of intervention. Of note, during the second year of intervention, rate of development was similar for both groups (.82 vs. .76). This being said, results still favored Group A when it came to achieving a “typical” rate of development at Year 2 assessment (40% vs.6%).

It should be noted that, at Year 1 assessment, Group A children from *Year 2 Sample* had not achieved the same levels of functioning as Group A children from *Year 1 Sample*. Furthermore, higher functioning children from *Year 2 Sample* had not reached the same levels of improvement in language, adaptive functioning, or developmental rate even after two years of intervention. This being said, the children with three data points were generally more delayed (i.e., had greater skill deficits) at intake relative to children with only Intake and Year 1 scores. As well, 12 of the 13 *Best Outcome Children* from *Year 1 Sample* exited the program shortly after Year 1, and therefore did not contribute to group means for the *Year 2 Sample*. (Of note, only two additional children had achieved *Best Outcome* status at Year 2 assessment.) All told, most of children who remained in the program for two full years achieved meaningful benefits, and it appears that children who stay in the program longer do so because they are continuously progressing and need more time to achieve the full benefits of the program. Previous studies have shown that, while the largest gains often occur during the first year of EIBI, children (on average) continue to achieve gains during the second year and beyond (e.g.,

Cohen et al., 2006; Howard et al., 2005; Sallows & Graupner, 2005; see review by Hayward et al., 2009).

The third hypothesis was that skill level at entry would be associated with level of skill acquisition in the later stages of treatment (Sallows & Graupner, 2005; Howard et al., 2005), and that certain key variables at entry (e.g., age, cognitive level, language skills) might predict which children would make the greatest gains during intervention. As such, a number of analyses were conducted to explore which intake variables might predict better outcomes at Year 1. Results showed strong significant correlations between cognitive and language scores at intake and all outcome measures at Year 1; furthermore, these two predictor variables accounted for a statistically significant amount of unique incremental variance for all outcome variables at Year 1. The final two-predictor model for FSIQ SS at Year 1 was statistically significant ( $R^2 = .644$ ), suggesting that 64.4% of the variance in Year 1 FSIQ can be predicted based on the combination of intake FSIQ followed by intake PLS-4 TL SS (with no additional predictive value from the remaining three outcome measures). This finding is somewhat consistent with Perry et al. (2011), who also reported a hierarchical regression model accounting for 64% of unique variance in IQ. However, in their model, all four of their predictors accounted for significant unique variance in IQ, with age making the largest contribution (5.3%) beyond initial IQ (54%). As for language, the final two-predictor model for Total Language SS at Year 1 was statistically significant ( $R^2 = .784$ ), suggesting that almost 80% of the variance in language at Year 1 can be predicted based on the combination of intake PLS-4 SS followed by intake FSIQ SS (again, with no additional predictive value from the remaining three outcome measures). Clearly, in the *Year 1 Sample*, language and

cognitive functioning offered the strongest predictive value for outcomes after one year of intervention. While adaptive functioning, autism severity, and rate of development showed some degree of predictive value in correlation and regression analyses, the values reported tended to be less substantial and not always significant. It should be further noted that (1) initial cognitive and language skills likely predict later levels in children not receiving EIBI, and (2) there was still some unexplained variance in both hierarchical regression models, suggesting that other, unexplored factors (e.g., parent factors, comorbid diagnoses, priorities in programming) may also make a significant contribution to outcomes.

Perry and colleagues (2011) found moderate significant negative correlations between age at entry and outcome variables (e.g., cognitive and adaptive functioning, severity of autism, rate of development), and also found that earlier age at entry was associated with better outcomes at exit when comparing younger vs. older children (under 4 vs. over 4 years of age). They also found that age accounted for a significant (but modest) amount of unique variance for most outcome variables (beyond the initial value of the variable in question; for example, age at entry accounted for an additional 6% of the variance in IQ outcomes at exit). Their findings were consistent with several other studies that included samples with a wide range of ages (e.g., Anderson et al., 1987; Bibby et al., 2002; Fenske et al., 1985). However, in the present study, age at entry was not significantly correlated with any of our outcome variables, and did not account for a significant amount of additional variance for any outcome variables at Year 1. A number of other studies have failed to show a relationship with age (e.g., Hayward et al., 2009; Lovaas, 1987; Smith et al., 2000), and it has been suggested that this may be due to

narrower age ranges, different types of analyses, and/or differences in sample characteristics (Perry et al., 2011), although these reasons may not apply to the present study. As such, the all-important question of age, and the notion of an early “sensitive period” when EIBI is likely to be most effective, warrants further investigation. Interestingly, while age did not emerge in this study as an important predictor of outcomes, all correlations between *Rate of Development at Intake* and all outcome measures were statistically significant (except for Rate of Development at T1). Perhaps a (relatively) stronger predictor of outcomes is a child’s intake age equivalent score in relation to their chronological age (i.e., how close they are to a “typical” rate of development), rather than chronological age *per se*. For example, in the present study, *Development Rate at Intake* accounted for 10.8% of the variance in PDDBI AC SS at Year 1 (but did not account for a statistically significant amount of incremental variance for any other outcome measure).

The final hypothesis was that approximately 10-15% of children in this study would resemble the “best outcome” children described in the literature after one year of intervention. Thirteen of the children (13%) for *Year 1 Sample* met criteria for *Best Outcome* (i.e., language and/or cognitive skills in the mid-Low Average range, adaptive functioning in the mid-Low Average range, and autism symptom severity below the “typical” level for children diagnosed with PDD). Furthermore, as predicted, these children tended to be less severe diagnostically and much higher functioning at intake. However, in the present study, *Best Outcome Children* were similar in age to the remaining children in the *Year 1 Sample* at intake (49.54 vs. 46.40 months), whereas the Average Functioning children in Perry et al. (2008) were, on average, more than a year

younger than the remaining children at intake (42 months vs. 55 months, respectively). Again, given the disparity in findings when it comes to the importance of age at entry, future research needs to clarify possible reasons for these reported discrepancies.

When compared to other studies of EIBI in the literature, the present research offers a number of strengths. First of all, this study contributes to the limited body of research examining community-based effectiveness of EIBI for children diagnosed with autism or ASD. The overall sample size is quite large ( $n = 100$  for our *Year 1 Sample*), smaller only to the Perry et al. study (2008), with two large subgroups and considerable heterogeneity on all outcome variables. As such, we were able to conduct a number of statistical analyses with less concern regarding power limitations (although this remained an issue with *Year 2 Sample*). Furthermore, similar to Perry and colleagues, *clinically significant* change was examined on an individual level, as well as statistical significance based on group mean changes (see Kazdin, 2005). In addition, the larger sample size allowed for examination of possible predictors of treatment outcomes at *Year 1*, which is often limited in studies with smaller samples.

Second, the present study examined results from standardized measures of cognitive, language, adaptive functioning, and symptom severity, as well as rate of development, which was based on age and SIB-R AE scores. Of particular note, standardized measures of both language skills *and* autism severity were included for analysis here, which has rarely been done in prior studies. Perry and colleagues (2008) included an autism severity measure (CARS), but this measure is based on a pre-DSM-IV conceptualization of autism, whereas the PDDBI is a current, more comprehensive measure normed on children diagnosed with PDD. As for language, few *effectiveness*



studies have included a standardized measure of language acquisition. Given that marked developmental delays in communication is one of the three DSM-IV diagnostic criteria for autism, examination of language skill development during EIBI is clearly warranted. Overall, the measures examined in this study address, to some degree, most (if not all) key changes of interest for children with ASD.

Third, this is the first large-scale, community-based effectiveness study where measures were taken for each child at *regular intervals* (annually) over a two-year period. First and foremost, this eliminated the possible confound of treatment duration when assessing outcomes, since all children were assessed at yearly intervals. Furthermore, preliminary findings from our *Year 2 Sample* offered a “first glimpse” of the yearly developmental trajectories for children with three consecutive data points. In addition, each child received the *same* four standardized measures at all time points. This reduces the impact of potential confounds when assessing outcomes, allowing for direct comparisons between intake and annual outcome scores, and avoids the difficulties involved when combining scores from different tests to produce one “global” measure of interest (e.g., “cognitive level” based on several different IQ measures). More often than not, the combining of measures is unavoidable in large, community-based samples (Matson, 2006); therefore, consistency of measurement instruments here represents a notable strength. Finally, all children in the present study came from one EIBI program, whereas Perry and colleagues (2008) examined outcomes for children participating in nine EIBI programs across different regions of Ontario, Canada. While the latter yielded a much larger sample ( $n = 332$ ), the intervention in the present study was likely more homogeneous.

In addition to the many strengths noted above, there are also a number of limitations which must be considered. Arguably, the biggest limitation is the lack of a comparison group of similar children receiving either no treatment (i.e., wait-list control group) or an alternative treatment. Without a comparison sample, gains made during EIBI cannot be conclusively attributed to the intervention (e.g., we must consider maturation effects, expectancy effects, other unknown factors, etc.). However, similar to Perry et al. (2008), the calculation of pre- and post developmental trajectories and the comparisons made between Group A and Group B offered us some degree of confidence that the intervention in question was, indeed, effective. Furthermore, in the present study, age was factored into each outcome measure (i.e., four standardized scores and rate of development), allowing more confidence that the results are not confounded by maturation effects. All this being said, the inclusion of a control group would address the majority of concerns regarding possible confounds. It should be further noted that, without a comparison group, results from predictor analyses here only demonstrate the relative relationship of various scores at two different time points – these predictors may be operating regardless of intervention (see Perry et al., 2011).

A second limitation of the current study involves the presence of missing data. Due to a variety of factors, only 100 children were available for analysis at Year 1, and many children only had complete data for adaptive functioning and one other measure (most often language functioning). In particular, lower sample size was of most concern when analyzing changes in autism symptom severity and especially changes in cognitive level. This being said, the magnitude of the time effect in our *Year 1 Sample* tended to be quite large (statistically significant mean change score, for both groups, on all outcome

measure), with quite large effect sizes, suggesting that chance findings were not likely over-interpreted. Of note, in the *Year 1 Sample*, the interaction effect for autism severity approached significance and the effect size was relatively small; this interaction may well have reached significance with a larger sample. Also, given the degree of missing data, we elected to perform correlation and regression analyses for the *Year 1 Sample* only.

While there may have been some drawbacks when analyzing results for the *Year 1 Sample*, the major impact of missing data occurred when analyzing outcomes for the *Year 2 Sample* ( $n = 50$ ). Although we proceeded with 2 x 3 RM ANOVAs, statistical values need to be interpreted with caution, due to the low sample sizes (particularly in Group A). For autism symptom severity, there were very few children in Group A, making statistical group comparisons “questionable” at best; for cognitive functioning, group comparisons were not possible ( $n = 1$  from Group A). Although a main effect for time was observed for three of the outcome measures in the *Year 2 Sample* (language, adaptive functioning, and autism symptom severity), group and time effects may have been more pronounced with larger sample sizes in each subgroup.

Other limitations include the fact that we have no information about whether gains achieved during EIBI were improved upon, maintained, or showed decline after exiting from the program. Many children from this study who achieved clinically meaningful gains went on to participate in the School-Age Program offered through St. Amant. Future research is needed on long-term outcomes for all children who participated in the early intervention program. Yet another limitation is that the people conducting the assessments were aware of the children’s participation in the program

(and were not independent of St.Amant), resulting in a potential bias towards demonstrating effectiveness of the program.

Despite these limitations, results from the present study contribute significantly to the existing literature on evaluating effectiveness of EIBI when delivered in “real world” community settings. To date, evaluations of EIBI delivered in the community have been sparse, and researchers in the field have emphasized the need for additional EIBI outcome research, particularly in “real world” community settings, where treatment is most often delivered. Similar to Perry and colleagues (2008) results from this study demonstrate that children diagnosed with ASD can achieve statistically and clinically significant reduction in autism symptom severity, as well as statistically and clinically significant improvement in language skills, adaptive behaviour skills, cognitive level, and rate of development while participating in a large, community-based intervention program. More importantly, the present study provided insights on the robustness of EIBI under less-controlled, real-world conditions, and contributed to a growing body of literature aimed at identifying individual and program variables that may be critical to EIBI success.

## References

- Achenbach, T. M. (1991a). *Child Behavior Checklist*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T. M. (1991b). *Manual for the Teacher's Report Form and 1991 Profile*. Burlington: University of Vermont Department of Psychiatry.
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4<sup>th</sup> ed., text revision). Washington, DC: American Psychiatric Association.
- Anderson, S. R., Avery, D. L., DiPietro, E. K., Edwards, G. L., & Christian, W. P. (1987). Intensive home-based intervention with autistic children. *Education and Treatment of Children, 10*, 352–366.
- Barlow, D. H. (1996). Health care policy, psychotherapy research, and the future of psychotherapy. *American Psychologist, 51*(10), 1050-1058.
- Bayley, N. (1993). *Bayley Scales of Infant Development (2nd ed.)*. San Antonio, TX: Psychological Corporation.
- Bibby, P., Eikeseth, S., Martin, N. T., Mudford, O. C., & Reeves, D. (2002). Progress and outcomes for children with autism receiving parent-managed intensive interventions. *Research in Developmental Disabilities, 23*, 81–104.
- Birnbrauer, J. S., & Leach, D. J. (1993). The Murdoch Early Intervention Program after 2 years. *Behavior Change, 10*, 63–74.
- Boyd, R. D., & Corley, M. J. (2001). Outcome survey of early intensive behavioral intervention for young children with autism in a community setting. *Autism, 5*, 430–441.

- Bruininks, R. H., Woodcock, R. W., Weatherman, R. F., & Hill, B. K. (1996). *Scales of Independent Behaviour-Revised (SIB-R)*. Chicago, IL: The Riverside Publishing Company.
- Cohen, I. L., Schmidt-Lackner, S., Romanczyk, R., & Sudhalter, V. (2003). The PDD Behavior Inventory: A rating scale for assessing response to intervention in children with PDD. *Journal of Autism and Developmental Disorders*, *33*, 31-45.
- Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Developmental and Behavioral Pediatrics*, *27*, 145-155.
- Cohen, I., & Sudhalter, V. (2005). *Pervasive Developmental Disorders Behavior Inventory-Parent Rating Form (PDDBI)*. Florida: Psychological Assessment Resources, Inc.
- Eikeseth, S. (2009). Outcome of comprehensive psycho-educational interventions for young children with autism. *Research in Developmental Disabilities*, *30*, 158-178.
- 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.
- Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2007). Outcome for children with autism who began intensive behavioral treatment between ages 4 and 7: A comparison controlled study. *Behavior Modification*, *31*, 264-278.
- Eldevik, S., Eikeseth, S., Jahr, E., & Smith, T. (2006). Effects of low-intensity behavioral treatment for children with autism and mental retardation. *Journal of Autism and Developmental Disorders*, *36*, 211-224.

- Eldevik, S., Hastings, R. P., Hughes, C. J., Jahr, E., Eikeseth, S., & Cross, S. (2009). Meta-analysis of early intensive behavioral intervention for children with autism. *Journal of Clinical Child & Adolescent Psychology, 38*, 439-450.
- Fenske, B. C., Zalenski, S., Krantz, P. J., & McClannahan, L. E. (1985). Age at intervention and treatment outcome for autistic children in a comprehensive intervention program. *Analysis and Intervention in Developmental Disabilities, 5*, 49-58.
- Fombonne, E. (2003). Epidemiological surveys of autism and other pervasive developmental disorders: An update. *Journal of Autism and Developmental Disorders, 33*, 365-382.
- Freeman, N. L., & Perry, A. (2010). Outcomes of intensive behavioural intervention in the Toronto Preschool Autism Service. *Journal on Developmental Disabilities, 16*(2), 17-32.
- Gordon, B. (2004). The Wechsler Preschool and Primary Scale of Intelligence: Test review. *Canadian Journal of School Psychology, 19*, 205-220.
- Harris, S. L., & Handleman, J. S. (2000). Age and IQ at intake as predictors of placement for young children with autism: A four- to six- year follow up. *Journal of Autism and Developmental Disorders, 30*, 137-142.
- Hayward, D., Eikeseth, S., Gale, C., & Morgan, S. (2009). Assessing progress during treatment for young children with autism receiving intensive behavioural interventions. *Autism, 13*, 613-633.

- Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatment for young children with autism. *Research in Developmental Disabilities, 26*, 359-383.
- Howell, David. C. (2007). *Statistical Methods for Psychology* (6<sup>th</sup> ed.). Belmont, CA: Wadsworth.
- Howlin, P., Magiati, I., & Charman, T. (2009). Systematic review of early intensive behavioral interventions for children with autism. *American Journal on Intellectual Abilities, 114*, 23-41.
- Kasari, C. (2002). Assessing change in early intervention programs for children with autism. *Journal of Autism and Developmental Disorders, 32*, 447-461.
- Kazdin, A. E. (2005). Clinical significance: Measuring whether interventions make a difference. In E. D. Hibbs & P. S. Jensen (Eds.), *Psychosocial treatments for child and adolescent disorders: Empirically-based strategies for clinical practice* (2<sup>nd</sup> ed., pp. 691-710). Washington, DC: American Psychological Association.
- Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: The Guilford Press.
- 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., 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.



- Makrygianni, M. K., & Reed, P. (2010). A meta-analytic review of the effectiveness of behavioural early intervention programs for children with Autistic Spectrum Disorders. *Research in Autism Spectrum Disorders, 4*, 577–593.
- Martin, G., & Pear, J. (2011). *Behavior modification: What it is and how to do it* (9<sup>th</sup> ed.). Upper Saddle River, NJ: Pearson Education.
- Matson, J. L. (2006). Determining treatment outcome in early intervention programs for autism spectrum disorders: A critical analysis of measurement issues in learning based interventions. *Research in Developmental Disabilities, 28*, 207-218.
- Matson, J. L., & Mishawaka, N. F. (2006). *Early intervention for autism spectrum disorders: A critical analysis*. Oxford, England: Elsevier Science Inc.
- 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.
- Msall, M. E. (2005). Measuring functional skills in preschool children at risk for neurodevelopmental disabilities. *Mental Retardation and Developmental Disabilities Research Reviews, 11*, 263-273.
- Mullen, E. M. (1995). *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service.
- Nathan, P. E., Stuart, S. P., & Dolan, S. L. (2000). Research on psychotherapy efficacy and effectiveness: Between Scylla and Charybdis? *Psychological Bulletin, 126*, 964-981.
- Partington, J. W., & Sundberg, M. L. (1998). *The Assessment of Basic Language and Learning Skills (The ABLLS)*. Pleasant Hill, California: Behavior Analysts Inc.

- Perry, A., & Condillac, R. (2003). *Evidence-based practices for children and adolescents with autism spectrum disorders: Review of the literature and practice guide*. Children's Mental Health Ontario.
- Perry, A., Condillac, R. A., Freeman, N. L., Dunn-Geier, J., & Belair, J. (2005). Multi-site study of the childhood autism rating scale (CARS) in five clinical groups of young children. *Journal of Autism and Developmental Disorders*, 35, 625-634.
- Perry, A., Cummings, A., Geier, J., Freeman, N. L., Hughes, S., LaRose, L., Managhan, T., Reitzel, J., & Williams, J. (2008). Effectiveness of intensive behavioral intervention in a large, community-based program. *Research in Autism Spectrum Disorders*, 2, 621-642.
- Perry, A., Cummings, A., Geier, J., Freeman, N. L., Hughes, S., Managhan, T., Reitzel, J., & Williams, J. (2011). Predictors of outcome for children receiving intensive behavioral intervention in a large, community-based program. *Research in Autism Spectrum Disorders*, 5, 592-603.
- Remington, B., Hastings, R.P., Kovshoff, H., degli Espinosa, F., Jahr, E., Brown, T., et al. (2007). Early Intensive Behavioral Intervention: Outcomes for children with autism and their parents after two years. *American Journal on Mental Retardation*, 112, 418-438.
- 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.
- Sattler, J. M. (2002). *Assessment of children: Behavioral and clinical applications* (4<sup>th</sup>ed). San Diego, CA: Jerome M. Sattler, Publisher, Inc.

- Schopler, E., Reichler, R. J., & Renner, B. (1988). *The Childhood Autism Rating Scale (CARS)*. Los Angeles, CA: Western Psychological Services.
- Sheinkopf, S. J., & Siegel, B. (1998). Home-based behavioral treatment of young children with autism. *Journal of Autism and Developmental Disorders*, 28, 15-23.
- Smith, T., Groen, A., & Wynn, J. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal on Mental Retardation*, 105, 269-285.
- Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales* (Interview Ed.). Circle Pines, MN: American Guidance Service.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. S. (1986). *The Stanford-Binet Intelligence Scale: Fourth edition (SB:FE)*. Chicago, IL: Riverside.
- Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, metaregression and dose-response meta-analysis of multiple outcomes. *Clinical Psychology Review*, 30, 387–399
- Wechsler, D. (2002). *The Wechsler Preschool and Primary Scale of Intelligence – Third Edition*. San Antonio, TX: The Psychological Corporation.
- Wells, K., Condillac, R., Perry, A., & Factor, D. C. (2009). A comparison of three adaptive behaviour measures in relation to cognitive level and severity of autism. *Journal on Developmental Disabilities*, 15(3), 55-63.
- Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool Language Scale – Fourth Edition (PLS-4)*. San Antonio, TX: The Psychological Corporation.

## Appendix

Table A

Age, gender, and diagnosis for *All Children Served* and *Year 1 Sample*

	<i>All Children Served</i> <sup>a</sup> n = 217	<i>Year 1 Sample</i> <sup>b</sup> n = 100
<b>Age (months)</b>		
Mean	47.38	46.75
Range	25-68	26-63
<b>Gender</b>		
Males	185 (84%)	85 (85%)
Females	32 (16%)	15 (15%)
Ratio	5.7:1	5.7:1
<b>Diagnosis</b>		
Autism/AD	90 (41%)	43 (43%)
ASD	120 (55%)	54 (54%)
PDD-NOS	7 (4%)	3 (3%)

Abbreviations: AD = Autistic Disorder; ASD = Autism Spectrum Disorder; PDD-NOS = Pervasive Developmental Disorder Not Otherwise Specified.

<sup>a</sup> *All Children Served*: All children served by the St. Amant ABA Preschool Program between 2003-2011.

<sup>b</sup> *Year 1 Sample*: Children with an intake SIB-R Broad Independence Standard Score and Intake/Year 1 data for at least one other outcome measure.

Table B

Intake scores for *All Children Served* and *Year 1 Sample*

	<i>All Children Served</i>		<i>Year 1 Sample</i>	
	<i>M (SD)</i>	Range	<i>M (SD)</i>	Range
<b>Language Functioning</b>	<b><i>n</i> = 190</b>		<b><i>n</i> = 94</b>	
PLS-4 TL SS	59.05 (16.14)	50-128	57.73 (13.40)	50-110
<b>Adaptive Functioning</b>	<b><i>n</i> = 151</b>		<b><i>n</i> = 100</b>	
SIB-R BI SS	53.71 (27.54)	1-131	49.54 (27.91)	1-131
<b>Autism Severity</b>	<b><i>n</i> = 118</b>		<b><i>n</i> = 69</b>	
PDDBI AC SS	51.54 (10.14)	26-74	51.33 (8.83)	29-74
<b>Cognitive functioning</b>	<b><i>n</i> = 94</b>		<b><i>n</i> = 60</b>	
WPPSI-III FSIQ SS	61.27 (21.65)	40-128	60.15 (19.73)	40-110
<b>Rate of Development</b>	<b><i>n</i> = 151</b>		<b><i>n</i> = 100</b>	
Broad Independence AE/CA	0.56 (0.24)	0.04-1.74	0.55 (0.25)	0.16-1.74

Abbreviations: PLS-4 TL SS = Preschool Language Scale (4<sup>th</sup> Edition) Total Language Standard Score; SIB-R BI SS = Scales of Independent Behavior–Revised Broad Independence Standard Score; PDDBI AC SS = Pervasive Developmental Disability Behavior Inventory Autism Composite Standard Score; WPPSI-III FSIQ SS = Wechsler Preschool and Primary Scale of Intelligence (3rd Edition) Full Scale IQ Standard Score; AE = Age Equivalent; CA = Chronological Age.

Table 1

Age, gender, and diagnosis for *Year 1 Sample* and the two subgroups<sup>a,b,c</sup>

	<i>Year 1 Sample</i> <i>n</i> = 100	<i>Group A</i> <i>n</i> = 39	<i>Group B</i> <i>n</i> = 61
<b>Age (months)</b>			
Mean	46.75	47.66	46.22
Range	26-63	26-63	26-63
<b>Gender</b>			
Males	85 (85%)	36 (88%)	51 (86%)
Females	15 (15%)	5 (12%)	8 (14%)
Ratio	5.7:1	7.3:1	6.1:1
<b>Diagnosis</b>			
Autism/AD	43 (43%)	9 (22%)	33 (56%)
ASD	54 (54%)	29 (70%)	26 (44%)
PDD-NOS	3 (3%)	3 (8%)	0 (0%)

Abbreviations: AD = Autistic Disorder; ASD = Autism Spectrum Disorder; PDD-NOS = Pervasive Developmental Disorder Not Otherwise Specified.

<sup>a</sup> *Year 1 Sample*: Children with an intake SIB-R Broad Independence Standard Score and Intake/Year 1 data for at least one other outcome measure

<sup>b</sup> *Group A* (Higher functioning): Children from *Year 1 Sample* with an intake SIB-R Broad Independence Standard Score of 60 and above

<sup>c</sup> *Group B* (Lower functioning): Children from *Year 1 Sample* with an intake SIB-R Broad Independence Standard Score of 59 or lower

Table 2

Intake scores for *Year 1 Sample*, *Group A* (Higher functioning), and *Group B* (Lower functioning)

	<i>Year 1 Sample</i> <i>M (SD)</i>	<i>Group A</i> <i>M (SD)</i>	<i>Group B</i> <i>M (SD)</i>	<i>Group A vs. B</i> <i>t value</i>
<b>Language Functioning</b>	<b><i>n = 94</i></b>	<b><i>n = 37</i></b>	<b><i>n = 57</i></b>	
PLS-4 TL SS	57.73 (13.40)	66.41 (17.36)	52.11 (4.73)	+4.89***
<b>Adaptive Functioning</b>	<b><i>n = 100</i></b>	<b><i>n = 39</i></b>	<b><i>n = 61</i></b>	
SIB-R BI SS	49.54 (27.91)	78.18 (12.76)	30.69 (16.50)	+16.16***
<b>Autism Severity</b>	<b><i>n = 70</i></b>	<b><i>n = 29</i></b>	<b><i>n = 41</i></b>	
PDDBI AC SS	51.33 (8.83)	48.28 (9.39)	53.93 (8.06)	-2.70**
<b>Cognitive functioning</b>	<b><i>n = 60</i></b>	<b><i>n = 31</i></b>	<b><i>n = 29</i></b>	
WPPSI-III FSIQ SS	60.15 (19.73)	71.90 (20.23)	49.16 (11.22)	+5.33***
<b>Rate of Development</b>	<b><i>n = 100</i></b>	<b><i>n = 39</i></b>	<b><i>n = 61</i></b>	
SIB-R AE/age at entry	0.55 (0.25)	0.74 (0.17)	0.42 (0.20)	+8.33***

\*\*\*  $p < .001$ , \*\*  $p < .01$ 

Abbreviations: PLS-4 TL SS = Preschool Language Scale (4<sup>th</sup> Edition) Total Language Standard Score; SIB-R BI SS = Scales of Independent Behavior-Revised Broad Independence Standard Score; PDDBI AC SS = Pervasive Developmental Disability Behavior Inventory Autism Composite Standard Score; WPPSI-III FSIQ SS = Wechsler Preschool and Primary Scale of Intelligence (3rd Edition) Full Scale IQ Standard Score; AE = Age Equivalent.

Table 3

Age, gender, and diagnosis for *Year 1 Sample* vs. *Year 2 Sample*

	<i>Year 1 Sample</i> <i>n</i> = 100	<i>Year 2 Sample</i> <i>n</i> = 50
<b>Age (months)</b>		
Mean	46.75	45.14
Range	26-63	26-63
<b>Gender</b>		
Males	85 (85%)	39 (78%)
Females	15 (15%)	11 (22%)
Ratio	5.7:1	3.5:1
<b>Diagnosis</b>		
Autism/AD	43 (43%)	26 (52%)
ASD	54 (54%)	24 (48%)
PDD-NOS	3 (3%)	0 (0%)

<sup>a</sup> *Year 1 Sample*: Children with an intake SIB-R Broad Independence Standard Score and Intake/Year 1 data for at least one other outcome measure

<sup>b</sup> *Year 2 Sample*: Children with an intake SIB-R Broad Independence Standard Score and all three data points on at least one other outcome measure



Table 4

Intake scores for *Year 1 Sample* vs. *Year 2 Sample*

	<i>Year 1 Sample</i> <i>M (SD)</i>	<i>Year 2 Sample</i> <i>M (SD)</i>
<b>Language Functioning</b>	<i>n = 94</i>	<i>n = 47</i>
PLS-4 TL SS	57.73 (13.40)	52.70 (7.18)
<b>Adaptive Functioning</b>	<i>n = 100</i>	<i>n = 50</i>
SIB-R BI SS	49.54 (27.91)	37.88 (24.45)
<b>Autism Severity</b>	<i>n = 70</i>	<i>n = 32</i>
PDDBI AC SS	51.33 (8.83)	52.97 (9.34)
<b>Cognitive functioning</b>	<i>n = 60</i>	<i>n = 29</i>
WPPSI-III FSIQ SS	60.15 (19.73)	50.90 (14.56)
<b>Rate of Development</b>	<i>n = 100</i>	<i>n = 50</i>
SIB-R AE/age at entry	0.55 (0.25)	0.47 (0.24)

Abbreviations: PLS-4 TL SS = Preschool Language Scale (4<sup>th</sup> Edition) Total Language Standard Score; SIB-R BI SS = Scales of Independent Behavior–Revised Broad Independence Standard Score; PDDBI AC SS = Pervasive Developmental Disability Behavior Inventory Autism Composite Standard Score; WPPSI-III FSIQ SS = Wechsler Preschool and Primary Scale of Intelligence (3rd Edition) Full Scale IQ Standard Score; AE = Age Equivalent.

Table 5

*Year 1 Sample: Outcome Measures at Intake (T0) and Year 1 (T1)*

	<b>Intake</b> <i>M (SD)</i>	<b>Year 1</b> <i>M (SD)</i>	<b>T1-T0</b> $\Delta$ score
<b>Language Skills (PLS-4 TL SS)<sup>a</sup></b>			
Group A ( <i>n</i> = 34)	66.85 (17.96)	77.71 (19.19)	10.86***
Group B ( <i>n</i> = 53)	52.26 (4.87)	55.64 (13.06)	3.38*
<b>Adaptive Functioning (SIB-R BI SS)<sup>b</sup></b>			
Group A ( <i>n</i> = 37)	78.59 (12.74)	88.95 (20.14)	10.35***
Group B ( <i>n</i> = 56)	29.66 (16.60)	43.41 (26.86)	13.75***
<b>Autism Severity (PDDBI AC SS)<sup>c</sup></b>			
Group A ( <i>n</i> = 28)	48.07 (9.50)	40.50 (9.35)	-7.57***
Group B ( <i>n</i> = 38)	54.37 (8.17)	51.08 (8.38)	-3.29*
<b>Cognitive Functioning (FSIQ SS)<sup>d</sup></b>			
Group A ( <i>n</i> = 23)	73.83 (18.28)	81.61 (16.03)	7.78*
Group B ( <i>n</i> = 20)	50.85 (13.34)	57.65 (20.56)	6.80*

\*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ .

<sup>a</sup> Repeated Measures ANOVA (group x time) – Significant main effects of time: Wilks'  $\Lambda = .699$ ;  $F(1,85) = 36.54$ ;  $p < .001$ ; partial  $\eta^2 = .301$  and significant interaction: Wilks'  $\Lambda = .926$ ;  $F(1,85) = 6.77$ ;  $p = .011$ ; partial  $\eta^2 = .074$ .

<sup>b</sup> Repeated Measures ANOVA (group x time) – Significant main effect of time: Wilks'  $\Lambda = .765$ ;  $F(1,91) = 27.90$ ;  $p < .001$ ; partial  $\eta^2 = .235$  and non-significant interaction: Wilks'  $\Lambda = .994$ ;  $F(1,91) = 0.56$ ;  $p = .490$ ; partial  $\eta^2 = .006$ .

<sup>c</sup> Repeated Measures (group x time) – Significant main effect of time: Wilks'  $\Lambda = .731$ ;  $F(1,64) = 23.50$ ;  $p < .001$ ; partial  $\eta^2 = .269$  and non-significant interaction: Wilks'  $\Lambda = .946$ ;  $F(1,64) = 3.65$ ;  $p = .060$ ; partial  $\eta^2 = .054$ .

<sup>d</sup> Repeated Measures (group x time) – Significant main effect of time: Wilks'  $\Lambda = .822$ ;  $F(1,42) = 8.63$ ;  $p = .005$ ; partial  $\eta^2 = .178$  and non-significant interaction: Wilks'  $\Lambda = .990$ ;  $F(1,42) = 0.41$ ;  $p = .526$ ; partial  $\eta^2 = .010$ .

Table 6

*Year 1 Sample: Rate of Development at Intake and During Year 1*

	<b>Intake</b> <i>M (SD)</i>	<b>During Year 1</b> <i>M (SD)</i>	<b>12 month</b> $\Delta$ score
<b>Rate of Development</b>			
Group A ( <i>n</i> = 37)	0.74 (0.17)	1.49 (0.96)	0.75***
Group B ( <i>n</i> = 56)	0.39 (0.20)	0.79 (0.97)	0.38*

\*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ .

Repeated Measures ANOVA (group x time) – Significant main effect of time: Wilks'  $\Lambda = .774$ ;  $F(1,91) = 26.58$ ;  $p < .001$ ; partial  $\eta^2 = .226$  and non-significant interaction: Wilks'  $\Lambda = .970$ ;  $F(1,91) = 2.83$ ;  $p = .096$ ; partial  $\eta^2 = .030$ .

Note: *Rate of Development at Intake* = SIB-R AE SS at Intake/Age at entry; *Rate of Development During Year 1* = (SIB-R AE SS at Year 1 - SIB-R AE SS at Intake)/12 months.

Table 7

Year 2 Sample: Outcomes for Children with Scores at Intake, Year 1, and Year 2

	<b>Intake</b> <i>M (SD)</i>	<b>Year 1</b> <i>M (SD)</i>	<b>Year 2</b> <i>M (SD)</i>	<b>Overall</b> $\Delta$ score	<b>1<sup>st</sup> Year</b> $\Delta$ score	<b>2<sup>nd</sup> Year</b> $\Delta$ score
<b>Language Skills (PLS-4 TL SS)<sup>a</sup></b>						
Group A ( <i>n</i> = 8)	58.38 (14.67)	70.50 (20.61)	75.38 (26.61)	17.00*	12.12*	4.88
Group B ( <i>n</i> = 23)	51.35 (3.16)	56.04 (12.67)	57.26 (13.04)	5.91	4.70	1.22
<b>Adaptive Functioning (SIB-R BI SS)<sup>b</sup></b>						
Group A ( <i>n</i> = 10)	71.80 (4.69)	81.50 (13.54)	80.60 (15.09)	8.80	9.70	-0.90
Group B ( <i>n</i> = 32)	25.88 (17.61)	40.19 (26.39)	45.72 (30.01)	19.84**	14.31**	5.53
<b>Autism Severity (PDDBI AC SS)<sup>c</sup></b>						
Group A ( <i>n</i> = 9)	50.67 (9.47)	43.11 (10.22)	41.00 (7.48)	-9.67***	-7.56	-2.11
Group B ( <i>n</i> = 17)	55.59 (7.53)	53.29 (7.05)	54.29 (10.53)	-1.29	-2.29	1.00
<b>Cognitive Functioning (FSIQ SS)<sup>d</sup></b>						
Group A ( <i>n</i> = 1)	45.00 ---	82.00 ---	76.00 ---	---	---	---
Group B ( <i>n</i> = 10)	46.20 (8.40)	52.90 (18.09)	53.70 (18.00)	7.50	6.70	0.80

\*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .017$ 

<sup>a</sup> Repeated Measures (group x time) – Significant main effect of group:  $F(1,29) = 7.24, p = .012$ , partial  $\eta^2 = .200$ ; significant main effect of time: Wilks'  $\Lambda = .604$ ;  $F(2,28) = 9.20, p < .001$ , partial  $\eta^2 = .396$ ; and non-significant interaction: Wilks'  $\Lambda = .871$ ;  $F(2,28) = 2.07, p = .145$ , partial  $\eta^2 = .129$ .

<sup>b</sup> Repeated Measures (group x time) – Significant main effect of group:  $F(1,40) = 35.80, p = .000$ , partial  $\eta^2 = .472$ ; significant main effect of time: Wilks'  $\Lambda = .820$ ;  $F(2,39) = 4.27, p = .021$ , partial  $\eta^2 = .180$ ; and non-significant interaction: Wilks'  $\Lambda = .961$ ;  $F(2,39) = .788, p = .462$ , partial  $\eta^2 = .039$ .

<sup>c</sup> Repeated Measures (group x time) – Significant main effect of group:  $F(1,24) = 9.27, p = .006$ , partial  $\eta^2 = .279$ ; significant main effect of time: Wilks'  $\Lambda = .650$ ;  $F(2,23) = 6.19, p = .007$ , partial  $\eta^2 = .350$ ; and significant interaction: Wilks'  $\Lambda = .768$ ;  $F(2,23) = 3.48, p = .048$ , partial  $\eta^2 = .232$ .

<sup>d</sup> Only Group B subjects were analyzed. For this RM analysis, there was a non-significant main effect for time: Wilks'  $\Lambda = .645$ ;  $F(2,8) = 2.20, p = .173$ , partial  $\eta^2 = .355$ .

Table 8

*Year 2 Sample: Rate of Development at Intake, During Year 1, During Year 2, and During 24 Months*

	<b>Intake</b> <i>M (SD)</i>	<b>During Year 1</b> <i>M (SD)</i>	<b>During Year 2</b> <i>M (SD)</i>	<b>During 24 Months</b> <i>M (SD)</i>	<b>Intake vs. 24 months</b> $\Delta$ score
<b>Rate of Development</b>					
Group A ( <i>n</i> = 10)	0.66 (0.05)	1.23 (0.66)*	0.82 (0.69)	1.02 (0.52)	0.36
Group B ( <i>n</i> = 32)	0.41 (0.26)	0.64 (0.95)*	0.76 (0.71)	0.70 (0.69)	0.29

\*  $p < .017$

Repeated Measures (2 groups x 3 time) – Significant main effect of group:  $F(1,40) = 23.48$ ,  $p = .000$ , partial  $\eta^2 = .370$ ; non-significant main effect of time: Wilks'  $\Lambda = .889$ ;  $F(2,39) = 2.45$ ,  $p = .100$ , partial  $\eta^2 = .111$ ; and non-significant interaction: Wilks'  $\Lambda = .940$ ;  $F(2,39) = 1.25$ ,  $p = .299$ , partial  $\eta^2 = .060$ .

Intake vs. During 24 months: Repeated Measures (2 groups x 2 time) – Significant main effect of group:  $F(1,41) = 7.41$ ,  $p = .010$ , partial  $\eta^2 = .153$ ; significant main effect of time: Wilks'  $\Lambda = .886$ ;  $F(1,41) = 5.28$ ,  $p = .027$ , partial  $\eta^2 = .114$ ; and non-significant interaction: Wilks'  $\Lambda = .999$ ;  $F(1,41) = 0.54$ ,  $p = .818$ , partial  $\eta^2 = .001$ .

Note: *Rate of Development During Year 2* = (SIB-R AE SS at Year 2 - SIB-R AE SS at Year 1)/12 months.

*Rate of Development During 24 Months* = (SIB-R AE Score at Year 2 – SIB-R AE Score at Intake)/24 months.

Table 9

*Year 1 Sample: Correlations of outcome variables with predictors*

Outcome variables at T1	Predictors at Intake					
	Age	Rate of Dev.	PLS-4 TL SS	SIB-R BI SS	PDDBI AC SS	FSIQ SS
<b>Language Skills</b>						
PLS-4 TL SS	.039	.404**	.779**	.507**	-.275*	.838**
<b>Adaptive Functioning</b>						
SIB-R BI SS	-.088	.653**	.600**	.766**	-.190	.659**
<b>Autism Severity</b>						
PDDBI AC SS	-.026	-.465**	-.578**	-.463**	.558**	-.549**
<b>Cognitive Functioning</b>						
FSIQ SS	-.135	.520**	.731**	.534**	-.244	.752**
<b>Rate of Development</b>						
(AE T1 – AE T0)/12 months	.065	.061	.478**	.358**	.078	.550**

\*\*  $p < .01$ , \*  $p < .05$

Note: For intake scores: age ( $n = 100$ ), PLS-4 SS ( $n = 94$ ), SIB-R BI SS ( $n = 100$ ), PDDBI Autism SS ( $n = 69$ ), and FSIQ SS ( $n = 60$ ), and Rate of Development at Intake ( $n = 100$ ). For outcome variables at T1,  $n$  varies.

Table 10

*Year 1 Sample* Simple regression: Variance accounted for by T0 predictor variables on dependent variables at T1

Outcomes Variables (T1)	Age (T0)		Rate of Dev (T0)		PLS-4 (T0)		SIB-R (T0)		PDDBI (T0)		FSIQ (T0)	
	R <sup>2</sup>	R <sup>2</sup> Δ	R <sup>2</sup>	R <sup>2</sup> Δ	R <sup>2</sup>	R <sup>2</sup> Δ	R <sup>2</sup>	R <sup>2</sup> Δ	R <sup>2</sup>	R <sup>2</sup> Δ	R <sup>2</sup>	R <sup>2</sup> Δ
<b>Language skills</b>												
PLS-4 TL SS	.606	.001	.606	.002	.606	---	.606	.010	.528	.002	.629	.152***
<b>Adaptive functioning skills</b>												
SIB-R BI SS	.586	.012	.582	.006	.591	.037**	.586	---	.557	.008	.593	.081***
<b>Autism Severity</b>												
PDDBI AC SS	.311	.007	.311	.108***	.314	.168***	.311	.086**	.311	---	.273	.136**
<b>Cognitive Functioning</b>												
WPPSI-III FSIQ SS	.565	.000	.609	.004	.602	.042*	.609	.009	.472	.000	.565	---
<b>Rate of Development (Year 1)</b>												
(AE T1 – AE T0)/12 months	.004	.004	.004	---	.003	.267***	.004	.242***	.035	.001	.001	.382***

\*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$

Note: For each comparison,  $n$  varies (pairwise analyses).

Table 11

*Year 1 Sample: Results for Best Outcome Children at Intake and Year 1*

	<b><u>Best Outcome Children</u></b>		<b><math>\Delta</math> Score</b>
	Intake <i>M (SD)</i>	Year 1 <i>M (SD)</i>	Intake vs. Year 1
<b>Language Functioning</b>	<b><i>n</i> = 12</b>		
PLS-4 TL SS	83.50 (17.25)	97.58 (8.94)	14.08**
<b>Adaptive Functioning</b>	<b><i>n</i> = 13</b>		
SIB-R BI SS	78.15 (22.05)	101.31 (16.91)	23.15**
<b>Autism Severity</b>	<b><i>n</i> = 7</b>		
PDDBI AC SS	46.86 (8.83)	31.43 (6.13)	-15.43**
<b>Cognitive functioning</b>	<b><i>n</i> = 11</b>		
WPPSI-III FSIQ SS	80.55 (18.29)	93.00 (12.69)	12.46**

\*\*  $p < .01$



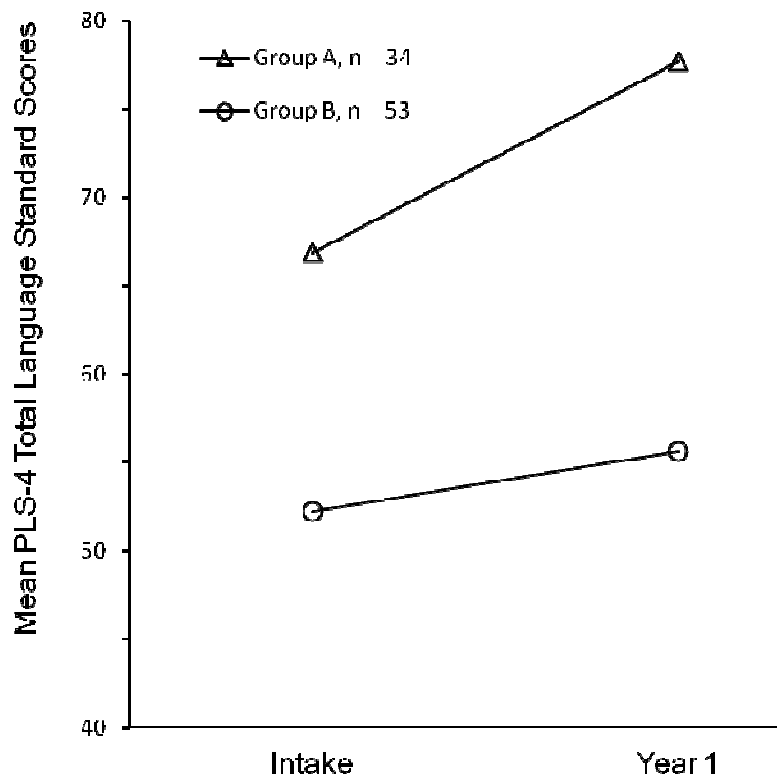


Figure 1. Mean Preschool Language Scale-4 Total Language Standard Scores at Intake and Year 1 for higher (Group A) and lower (Group B) functioning children.

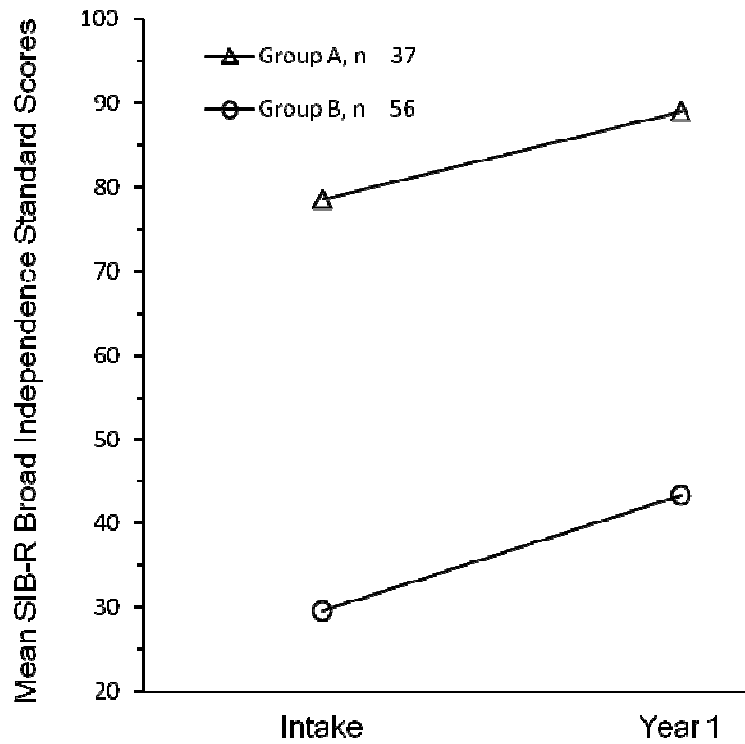


Figure 2. Mean Scales of Independent Behavior Broad Independence Standard Scores at Intake and Year 1 for higher (Group A) and lower (Group B) functioning children.

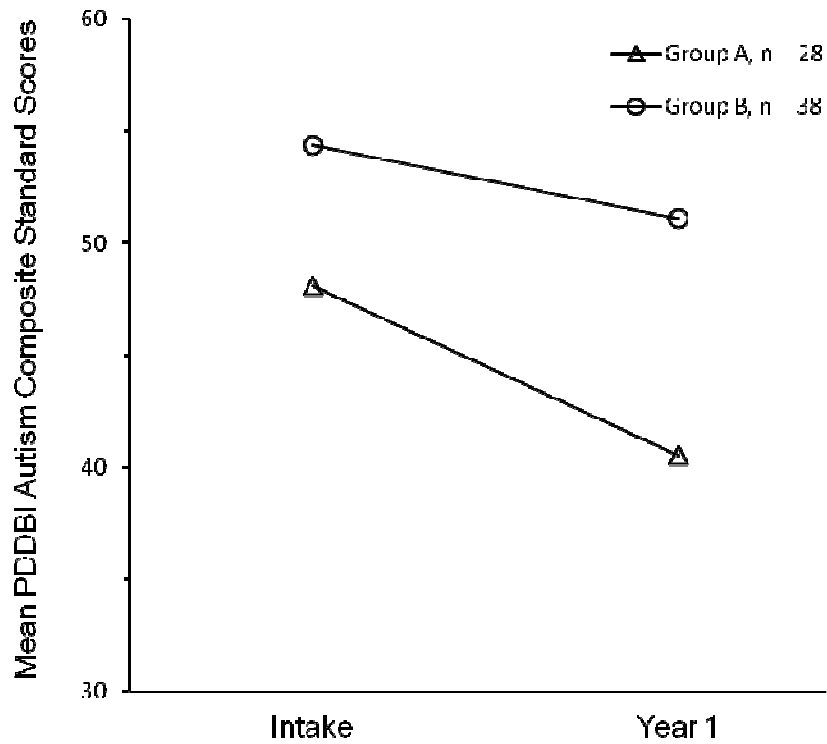


Figure 3. Mean Pervasive Developmental Disorders Behavior Inventory Autism Composite Standard Scores at Intake and Year 1 for higher (Group A) and lower (Group B) functioning children.

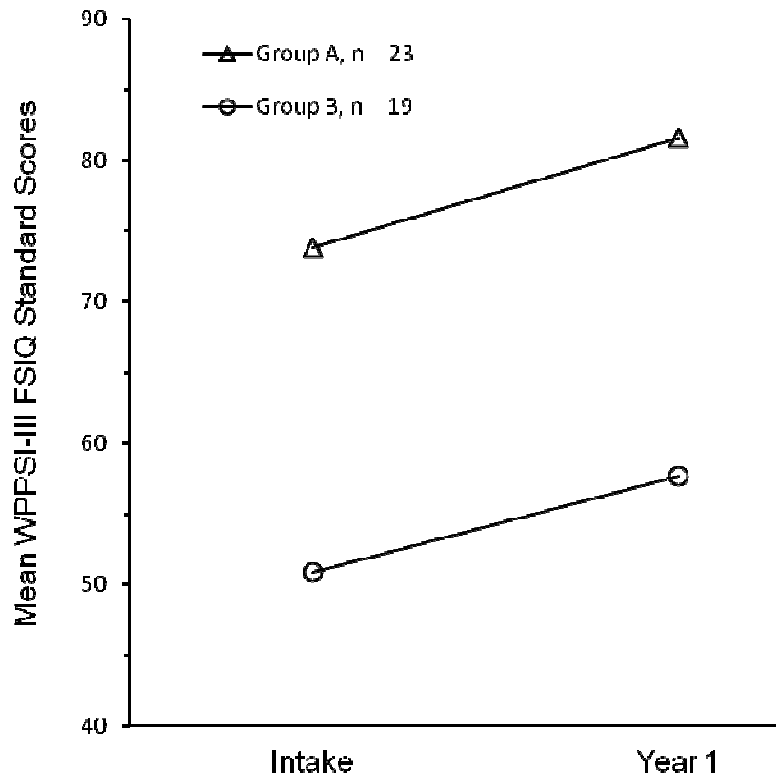


Figure 4. Mean WPPSI-III Full Scale IQ Standard Scores at Intake and Year 1 for higher (Group A) and lower (Group B) functioning children.

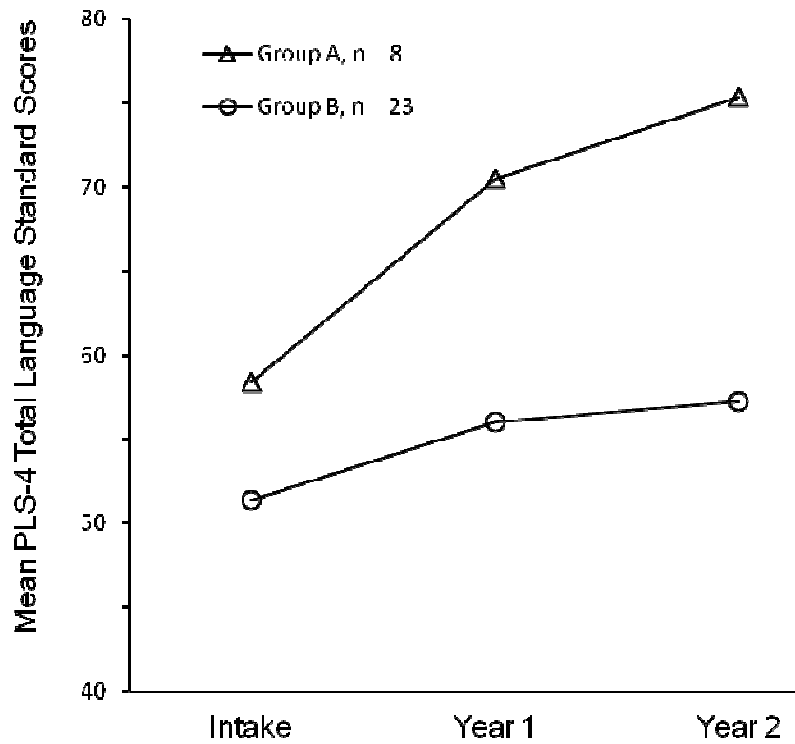


Figure 5. Mean Preschool Language Scale-4 Total Language Standard Scores at Intake, Year 1, and Year 2 for higher (Group A) and lower (Group B) functioning children.

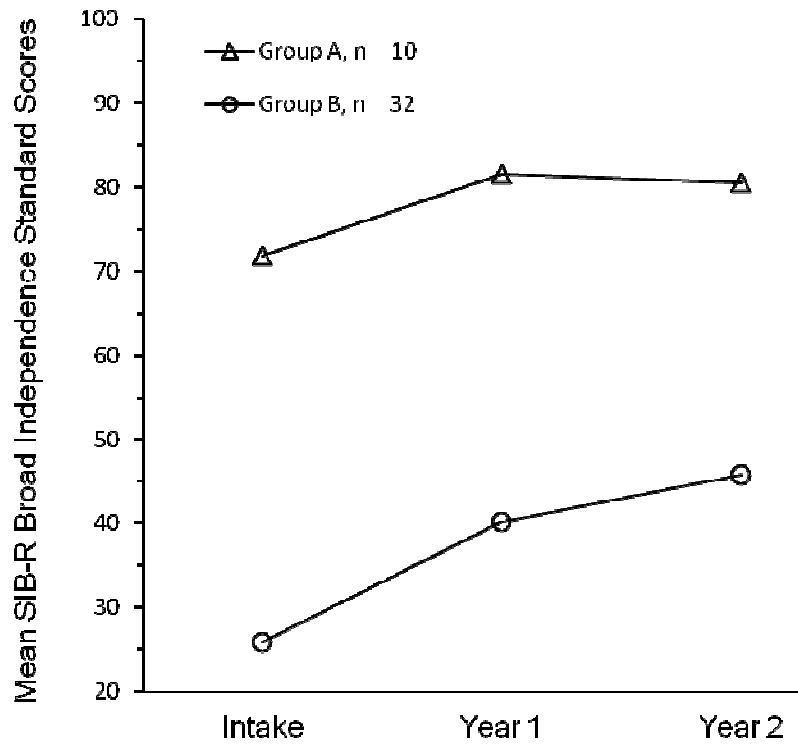


Figure 6. Mean Scales of Independent Behavior Broad Independence Standard Scores at Intake, Year 1, and Year 2 for higher (Group A) and lower (Group B) functioning children.

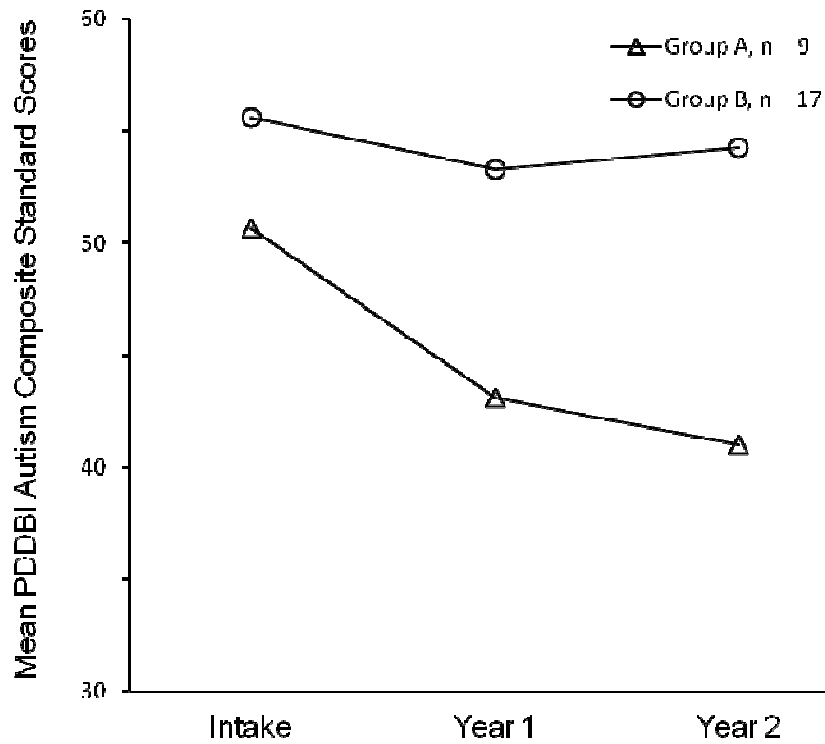


Figure 7. Mean Pervasive Developmental Disorders Behavior Inventory Autism Composite Standard Scores at Intake, Year 1, and Year 2 for higher (Group A) and lower (Group B) functioning children.

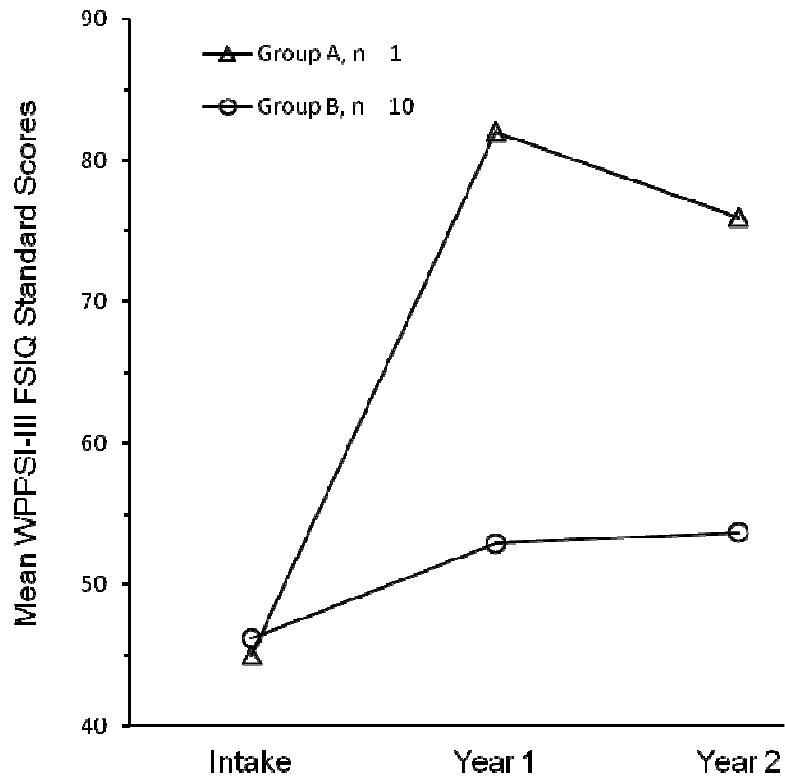


Figure 8. Mean WPPSI-III Full Scale IQ Standard Scores at Intake, Year 1, and Year 2 for higher (Group A) and lower (Group B) functioning children.