JoDD
Journal on Developmental Disabilities
Le journal sur les handicaps du développement
Volume 16, Number 2, 2010

Outcomes of Intensive Behavioural Intervention in the Toronto Preschool Autism Service

Authors

Nancy Freeman,[1,2]
Adrienne Perry[1,2]

[1] Surrey Place Centre, Toronto, ON

[2] York University, Toronto, ON

Correspondence

nancy.freeman@surreyplace.on.ca

Keywords

intensive behavioural intervention,
autism,
preschool,
outcomes

Abstract

Intensive Behavioural Intervention (IBI) has documented efficacy in small, model programs for young children with autism, but less is known about its effectiveness in large, less prescribed community settings. This paper reports on the outcomes of 89 preschool children (aged 2 to 6) receiving IBI from the Toronto Preschool Autism Service (TPAS). Results indicate substantial improvement in severity of autistic symptoms, cognitive and adaptive behaviour, and rate of development. Outcomes were heterogeneous and were classified into 7 categories, which included some children similar to the "best outcomes" reported in model programs. Predictors of better outcomes included medium to higher initial levels of functioning, younger age at program entry, and relatively longer duration of IBI.

Autism has been considered very difficult to treat and to have a very poor prognosis. A number of interventions have been proposed for autism (see Perry & Condillac, 2003), including some which have been highly sensationalized in the media, but were eventually demonstrated through research to be of limited or no benefit or, indeed, to pose significant risk of harm.

Professional consensus panels (see National Research Council, 2001; New York State Department of Health, 1999; Schreibman, 2000) have strongly recommended early intensive behavioural intervention (IBI) as "best practice" for young children with autism, based on reviews of research literature demonstrating remarkable effectiveness in a significant minority of children studied (e.g., Lovaas, 1987) and improvement in most others. In 1999, the Ontario government launched a province-wide IBI initiative (Ministry of Community and Social Services (MCSS), 2000; Perry, 2002a) based on recent evidence and expert panels.

IBI has a much stronger empirical basis than virtually any other intervention used with children with autism (Perry & Condillac, 2003). However, it is not a panacea. There is a great deal of heterogeneity in outcome in the published studies. In Lovaas' (1987) pioneering work, 19 children with autism (mean age of 3) were given 40 hours per week of IBI for about 2 years and 47% evidenced "best outcomes" (i.e., loss of diagnosis, in regular class with no support), 42% showed moderately good outcomes (i.e., some improvement but still in special classes) and 11% had poor outcomes (i.e., little progress). These results were remarkable compared to the 19 well-matched control children who received 10 (or fewer) hours of IBI per week, plus other community services, only one of whom did very well (45% moderate progress, 53% poor outcomes). These results were essentially maintained at long-term follow-up (McEachin, Smith, & Lovaas, 1993). Methodological critiques of this work have been thoroughly explored and convincingly addressed (e.g., Eikeseth, 2001). Sallows and Graupner (2005) have replicated Lovaas' findings; 48% of children who received either clinic-directed or parent-directed IBI functioned in the average range following intervention.

Subsequent research on IBI has attempted to replicate and extend this work and has examined a number of critical parameters (Perry 2002b). This research indicates that high intensity IBI (20 to 40 hours per week) is associated with clear evidence of improvement in IQ scores and other measures, whereas low amounts of IBI (e.g., 10 hours per week) and special education or eclectic treatments were ineffective or even associated with declines in IQ. Two studies have demonstrated that IBI is significantly more effective than special education of equal intensity (Eikeseth, Smith, Jahr, & Eldevik, 2002; Howard, Sparkman, Cohen, Green, & Stanislaw, 2005).

Prediction of which children will improve and to what degree is still quite uncertain. Children with higher initial IQ and children who begin early (e.g., before age 4 or 5) are likely to do better (e.g., Harris & Handleman, 2000). The literature on the duration of IBI is mixed and hard to interpret, since there are often administrative and clinical criteria for changing intensity or discontinuing treatment. There are some studies showing evidence of significant effectiveness within one year and others which report greater gains in the second year. The studies with the best outcomes typically provided two years or more of IBI (e.g., Lovaas, 1987).

The IBI programs studied are typically small, clinical programs which vary in a number of ways (see Handleman & Harris, 2001). Reviewers of these studies (Anderson & Romanczyk, 1999; Dawson & Osterling, 1997; Green, 1996; Powers, 1992) have attempted to glean the common characteristics in programs with demonstrated efficacy. These can be reduced to six key features: 1) early age at onset of treatment (usually before age 4); 2) a large quantity of intervention—typically 20 to 40 hours/week for 1 to 2 years in duration; 3) a curriculum which is comprehensive in scope, developmental in sequence, individualized for the child, and builds in generalization; 4) positive-oriented, functional approach to problem behaviour; 5) highly trained and well-supervised staff; and 6) parent involvement.

There are some significant methodological issues with this body of literature, both in terms of internal and external validity. First, they are small, clinical samples (fewer than 30 participants) and lack of power precludes many interesting analyses. Second, few studies have control or comparison conditions (only two had random assignment). Third, children are often selected in some way (e.g., above a certain developmental level, absence of co-morbid diagnoses), which limits generalizability. Fourth, the intake and outcome measures (IQ and school placement) are also an issue; IQ has been shown to change significantly, but there are criticisms of its significance as an outcome. Fifth, these measures are often conducted by the same people who are responsible for the treatment. Finally, there is a major issue of treatment specification (i.e., what exactly the treatment is and which components are essential) and implementation (i.e., treatment integrity, fidelity, or quality).

These small studies with selected samples and carefully controlled treatment parameters are considered to be efficacy studies. *Efficacy* refers to whether a treatment or intervention can be shown to work under certain "ideal" conditions. *Effectiveness*, on the other hand, involves whether the intervention actually works when applied in "real life" situations, where there is typically less control over relevant variables (Kazdin, 2005). There have been few large-scale evaluations of IBI in community settings with all their inherent challenges, and it would not be surprising that the results of such effectiveness evaluations would be less positive.

There are three major differences between the Ontario IBI program and the published outcome studies, which may impact upon the results of the current study. First, there have been tremendous capacity-building challenges in recruiting, hiring, training, supervising, and retaining large numbers of staff. Thus, treatment fidelity and quality is much harder to obtain and maintain on such a large scale. Second, the children are not selected in such a way as to bias in favour of good results, as has been done in some model programs (notably Lovaas, 1987). In fact, the opposite may be the case since the Ontario IBI program is intended for more severe and needy children. Therefore, it is important to examine the results carefully in different subgroups within the data. Third, parents are not selected in such a way as to have financial, intellectual, or personal resources which might be associated with better outcomes, as in some of the model programs. Families in the Ontario program represent substantial diversity linguistically, culturally, and socioeconomically and the program has been implemented in a climate of considerable parent pressure, advocacy, media attention, and litigation.

The present Program Evaluation was conducted in 2005 (Freeman & Perry, 2005), and subsequently expanded to the other eight Regional Programs; the data reported here were included in that provincial study (Perry et al., 2008). This paper reports on the outcomes of children based on pre-post psychological assessment data from Toronto Preschool Autism Service (TPAS), which is the largest of the nine publicly-funded Regional Programs in Ontario providing IBI. This report examines the following three main research questions:

1) Do children show significant improvement in autism severity, adaptive behaviour, cognitive level, and rate of development?

2) Given the heterogeneity in the population and in outcomes in previous studies, what is the range of progress/outcomes of children in the program?

3) What factors are related to their progress and outcome?

Method

Ethics review was provided by Surrey Place Centre.

Participants

Psychological assessment files from 89 children with intake and exit assessments (within 3 months) were examined. Not all files included all measures or variables of interest, especially at intake. There were 73 (82%) boys and 16 (18%) girls, a sex ratio which is typical of this population (4:1). There was considerable linguistic, ethnic, and socioeconomic diversity in the sample. Children's age at program entry ranged from 20 to 83 months ($M = 53.64$, $SD = 13.12$), and duration in the program ranged from 5 to 47 months ($M = 19.39$, $SD = 9.76$). Age at entry and duration were highly correlated ($r = -.86$); children who started at younger ages received IBI for longer durations and those who started at an older age received it for shorter durations.

The children's initial status on all diagnostic and developmental variables is shown in Table 1. All children had a diagnosis of autism or falling towards the severe end of the autism spectrum. On the CARS over half fell in the mild/moderate autism range and a third in the severe autism range. VABS Standard Score means were in the 50s, with a wide range. Age Equivalent Scores were extremely variable, with mean age equivalents about: 13 months for Communication, 20 months for Daily Living, 12 months for Socialization, and 27 months for Motor, a VABS profile commonly seen clinically in this population (Perry, Flanagan, Dunn Geier, & Freeman, 2009). Note that the CARS and VABS are described more fully below. The estimated Full Scale IQ scores (when available) ranged widely, from the profound intellectual disability range to the borderline range, with a mean Mental Age (MA) of about 18 months. Initial Developmental Rate (see Measures) averaged .30 ($SD = .12$), indicating that their rate of development prior to involvement in the TPAS program had been less than one-third of a typical rate of development.

Three Subgroups

Children were classified into one of three *Initial Level of Functioning Groups*, defined *a priori* based on the VABS ABC Standard Scores at intake. The groupings were defined as follows: a) *"higher" functioning* (ABC 60 to 74); b) *medium functioning* (ABC between 50 and 59); and c) *lower functioning* (ABC 49 or lower). "Higher" is a relative term, and must be distinguished from the term "high functioning" which is sometimes used to mean the average intellectual range. These children were still in the mild to borderline range developmentally and at the severe end of the autism spectrum. Developmental and diagnostic scores for the three subgroups are found in Table 2.

These subgroups were compared using one-way ANOVAs on the initial variables, with signifi-

Table 1. Developmental and Diagnostic Status of Participants at Intake (n = 89 except as noted)

	M (SD) **or** *n (%)*	***Range***
DSM-IV Diagnosis		
• Autistic Disorder	55 (60.7%)	
• PDD-NOS	28 (31.4%)	
• PDD or ASD (unspecified)	7 (7.9%)	
CARS Total Score	35.49 (4.58)	27–45
Category:		
• not autism	10 (11.2%)	
• mild/moderate	48 (53.9%)	
• severe	31 (34.8%)	
VABS Standard Scores		
• Communication (n = 84)	51.64 (9.46)	40–96
• Daily Living (n = 84)	53.76 (10.69)	26–82
• Socialization (n = 84)	55.25 (6.31)	43–79
• Motor (n = 81)	58.98 (15.30)	16–101
• ABC (n = 84)	50.65 (8.56)	31–74
VABS Age Equivalents (months)		
• Communication (n = 85)	12.56 (5.80)	1–32
• Daily Living (n = 85)	19.72 (5.38)	11–35
• Socialization (n = 85)	11.75 (4.71)	1–25
• Motor (n = 82)	26.73 (7.10)	12–47
• ABC (mean of first 3) (n = 85)	14.83 (4.60)	7–25
FSIQ Estimate (n = 31)	36.65 (14.83)	15–77
Mental Age (months) (n = 31)	18.03 (9.06)	8–55
Initial Rate of Development (n = 86)	.30 (.12)	.11–.84

cant ANOVAs ($p < .001$) followed by post hoc tests to compare groups. For CARS Total score, the lower and medium group did not differ and both were significantly different from the "higher" group. All VABS standard scores followed the pattern of overall significant ANOVAs (F significant $p < .001$) and each group being different from every other on post hoc tests, indicating good separation of the three groups (as would be expected based on the way the groups were defined). For cognitive level (when available) the "lower" group had a mean IQ estimate which was significantly lower than either the medium or higher groups, who did not differ significantly from one another. However, Age Equivalents on the VABS and on cognitive level (i.e., MA) did not differ across groups (not shown). This indicates that the three groups entered the program at roughly the same developmental level (mean age equivalents were 16, 15, and 14 months) but they varied in chronological age and thus they differed in standard scores which are corrected for age. The Initial Developmental Rate for children in the three subgroups also differed (each group differed from the others), indicating that these are three groups of children who, prior to entering the IBI program, had been progressing at different rates.

Measures

Autism severity was measured using the *Childhood Autism Rating Scale (CARS*; Schopler, Reichler, & Renner, 1988) at intake and discharge. Scores range from 15 to 60, with higher scores indicating greater severity. Previous research has demonstrated excellent psychometric properties in older samples (Perry & Freeman, 1996) and in a diagnostically heterogeneous preschool-aged sample (Perry, Condillac, Freeman, Dunn Geier, & Belair, 2005).

Adaptive behaviour levels were assessed using the *Vineland Adaptive Behavior Scales (VABS*; Sparrow,

Table 2. Three Initial Functioning Level Subgroups

	"Higher" Functioning (n = 11) *M (SD)*	**Medium Functioning (n = 32)** *M (SD)*	**Lower Functioning (n = 42)** *M (SD)*	***post hoc*** **Group Differences**
CARS Total	30.91 (3.15)	35.03 (3.99)	37.05 (4.61)	1 < 2,3
VABS Standard Scores				
• Communication	68.64 (11.96)	53.66 (4.82)	45.51 (3.27)	1 > 2 > 3
• Daily Living Skills	70.27 (6.45)	58.91 (4.28)	45.32 (6.57)	1 > 2 > 3
• Socialization	65.00 (6.94)	57.22 (4.57)	51.10 (2.73)	1 > 2 > 3
• Motor	83.00 (11.05)	61.88 (12.74)	49.58 (8.24)	1 > 2 > 3
• ABC (Mean of 3)	66.64 (4.93)	53.59 (3.37)	44.07 (3.93)	1 > 2 > 3
Cognitive Level				
• FS IQ Estimate (n = 4,14, 12)	52.50 (12.23)	41.14 (13.55)	26.83 (10.62)	1,2 > 3
Rate of Development pre-IBI	.46 (.08)	.33 (.11)	.23 (.07)	1 > 2 > 3
Age	36.46 (10.76)	48.78 (11.27)	61.19 (8.62)	1 < 2 < 3

Note. All ANOVAs significant at $p < .001$

Balla, & Cicchetti, 1984) semi-structured parent interview at intake and discharge. Age Equivalents (in months) and Standard Scores ($M = 100$; $SD = 15$) were used in different analyses for three principle domains (i.e., Communication, Daily Living Skills, Socialization), the overall Adaptive Behavior Composite (ABC) score, and Motor if under 6 years.

The child's *development rate* was also assessed. The *Initial Developmental Rate* was calculated by dividing the VABS ABC age equivalent score by the child's age at the time (e.g., a 24-month-old child with an ABC of 12 months would have an initial developmental rate of .50). The *Developmental Rate during Intervention* was calculated by taking the difference between the exit and intake age equivalents, divided by the IBI duration interval between them. For example, if the 24-month-old child initially at the 12-month level achieved a 30-month score after 18 months of IBI, his/her developmental rate during intervention would be: (30–12)/18 = 1.0.

Information on *cognitive levels* was available for 35% of children at intake and 66% at discharge, due to resource limitations (especially during TPAS start-up). As is usual in autism research, different tests were used to accommodate children's varying age and level of ability. The tests included the Mullen Scales of Early Learning (Mullen, 1995; n = 16 at intake; n = 36 at exit); the Bayley Scales of Infant Development (Bayley, 1993; n = 14 at intake); the Stanford-Binet Intelligence Scale: Fourth Edition (SB:IV; Thorndike, Hagen, & Sattler, 1986; n = 1 at intake; n = 13 at exit); and the Wechsler Preschool and Primary Scale of Intelligence (3rd ed.) (WPPSI-3; Wechsler, 2002; n = 9 at exit). For the Mullen and Bayley, a Mental Age (MA) was derived to generate a Ratio IQ, since children were typically too old to obtain a standard score. For the SB:IV and WPPSI-3, a Full Scale IQ was used, and an MA was calculated. For most analyses, it was necessary to use these estimates of IQ and MA, in order to have larger numbers.

Results

1. Did Children Improve?

Autism Severity

Statistically significant improvement in Total CARS scores was clearly apparent from intake to discharge, as shown in Table 3, decreasing from about 36 to 31, on average. Looking at individual data for CARS Categories, of those children who scored in the mild/moderate autism range at intake (n = 48); 44% were in the non-autism range at discharge; 46% were still in the same mild/moderate range; and 10% were now in the severe range. Of those who had been in the severe autism range at intake (*n* = 31); 13% had improved so substantially that they were now in the non-autism range; 65% had improved and were now in the mild/moderate range; and 23% were still severe. A McNemar Exact test for these frequency data was highly significant ($p < .001$) as well as a chi-square test ($X^2(4) = 15.93$ $p = .003$). Thus, 77% of the children in the severe range could be said to have improved versus 44% of those in the mild/moderate range.

Adaptive Behaviour

VABS Age Equivalents for all domains increased significantly from intake to discharge (see Table 3), as did raw scores (not shown). This was true of Communication, Daily Living Skills, Socialization, Motor, and overall ABC age equivalent (all $p < .001$). These results clearly indicate that children had significantly more skills in all areas of adaptive behaviour when they left the program. However, the examination of VABS standard scores, which control for age, is a more stringent test and, as shown in Table 3, were fairly stable overall. The Communication domain was marginally statistically significantly higher at exit than at intake, but the Daily Living Skills domain was significantly lower at exit, and the other scores did not differ from intake to exit. In general, the differences in standard scores were quite small, well within the standard error of the test. It is not uncommon, when reviewing individual clinical data in these children, for raw scores and age equivalents to increase while standard scores remain stable or even decrease because, although children are gaining skills the rate of their developmental acquisition is slower than the rate of their chronological age maturation. In other words, they are aging faster than they are learning.

The lack of a difference between the overall group mean standard scores from intake to exit may, however, obscure substantial subgroup or individual differences. A repeated measures ANOVA for the three groups (between) and

Table 3. Comparison of Scores at Intake and Exit

	Intake *M (SD)*	**Exit** *M (SD)*	*t*	*p*
CARS Total (*n* = 88)	35.56 (4.56)	30.93 (4.99)	9.45	< .001
VABS Standard Scores				
• Communication (n = 81)	51.79 (9.59)	55.41 (22.25)	-1.92	.06
• Daily Living Skills (n = 81)	53.90 (10.83)	47.94 (16.16)	4.49	< .001
• Socialization (n = 81)	55.23 (6.23)	56.30 (10.26)	-1.16	ns
• Motor (n = 46)	65.22 (16.32)	63.70 (21.20)	0.55	ns
• ABC (Mean of 3) (n = 81)	50.79 (8.63)	49.88 (16.05)	0.72	ns
VABS Age Equivalents (months)				
• Communication (n = 82)	12.59 (5.82)	28.37 (20.20)	-7.62	< .001
• Daily Living Skills (n = 82)	19.71 (5.42)	30.63 (11.30)	-9.61	< .001
• Socialization (n = 82)	11.67 (4.38)	20.73 (12.00)	-7.29	< .001
• Motor (n = 63)	25.97 (6.94)	41.00 (14.66)	-9.10	< .001
• ABC overall Age Equiv. (n = 82)	14.82 (4.54)	26.60 (13.67)	-8.28	< .001
Cognitive Level				
• FS IQ Estimate (n = 20)	38.10 (16.42)	49.45 (28.93)	-2.67	.015
• MA (n = 20)	18.50 (10.65)	33.40 (17.49)	-4.43	< .001
Rate of Development (n = 83)	.30 (.12)	.60 (.65)	-4.19	< .001

two time periods (within) for Communication Standard Scores indicated a main effect for time and a group x time interaction, with post hoc tests indicating each group differed significantly from the others. Both the initially "higher" functioning group and the medium group did show significant increases in VABS Communication Standard Scores, from 68.64 (*SD* = 11.96) to 89.00 (*SD* = 25.31) and from 53.78 (*SD* = 4.26) to 58.78 (*SD* = 21.31) respectively, whereas the scores for the lower functioning group decreased from 47.33 (*SD* = 2.50) to 43.91 (*SD* = 5.79), indicating that the lower functioning group (though their age equivalent scores increased) were not keeping pace with the age-referenced scores and thus were falling somewhat further behind. This type of analysis was not significant for the other VABS Standard Scores.

Cognitive Levels

There were only 20 children who had cognitive scores of some kind at both intake and discharge. They showed a significant increase in IQ of about 11 points, on average. Nine children made clinically significant gains (defined as 15 points or more) and 3 of these made particularly noteworthy gains (37, 47 and 54 IQ points, respectively). Similarly, mean Mental Age increased significantly. Clinically significant cognitive gains (defined as one standard deviation or more [10 months]) were seen in 10 of these 20 children.

Rate of Development

For the total sample, the rate of development during IBI was .60 (*SD* = .65); this is double the initial rate of .30 (*SD* = .12) and represents a substantial alteration in developmental trajectory. However, note that there was a considerably larger standard deviation during IBI, indicating substantial individual variation.

The three subgroups were also examined separately. The "higher" functioning children had an initial developmental rate of .46 (*SD* = .08) and a rate during IBI of .98 (*SD* = .47), approximately double (2.13 times) and virtually a typical rate of development. Six of 11 children in this group (55%) had rates of development at or above a typical rate. The medium functioning group, who started at a developmental rate of .33 (*SD* = .12) increased to a rate of .58 (*SD* = .55). Although this is not as good a rate as the first group in absolute terms, proportion-

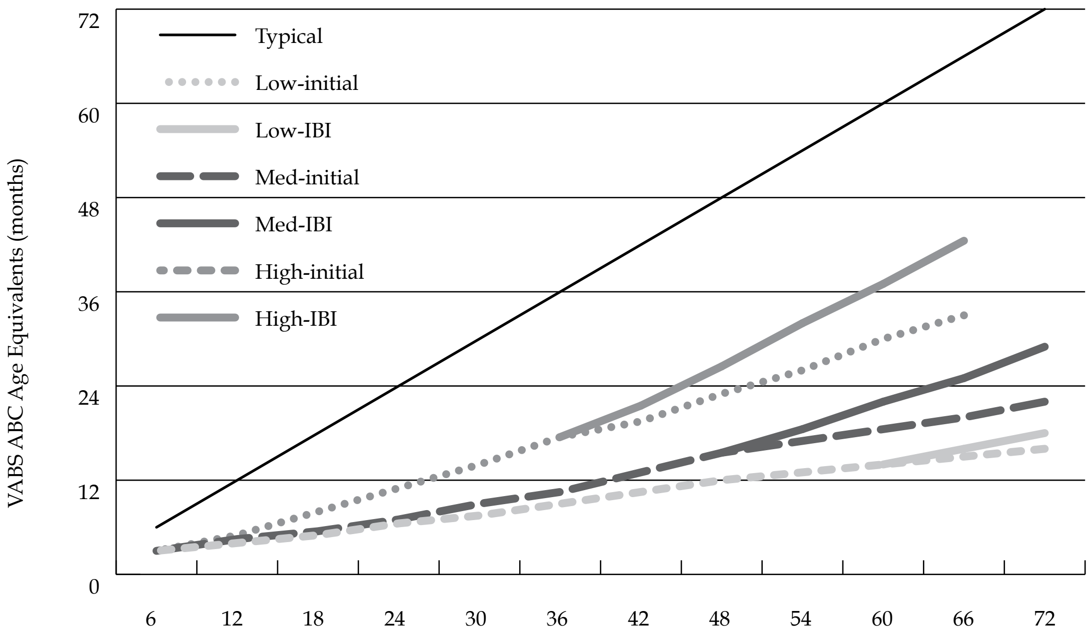

Figure 1. Developmental Trajectories for children in the three initial functioning level groups prior to (dotted line) and during (solid line) IBI

ately it is close (1.76 times the initial rate). Seven of 30 (23%) children for whom these rates were available in this medium functioning group were developing at or above a typical rate of development during IBI. The lower functioning group with an initial rate of development of .23 (*SD* = .07) showed a rate during intervention of .50 (*SD* = .74), which is 2.17 times their own initial rate of development. Four of these children (10% of the 41 with these rates available) were developing at or above a typical rate of development. The standard deviations are substantially larger for the rates of development during intervention which provides further evidence of the variability in response to IBI within the sample. Figure 1 illustrates the hypothetical developmental trajectory of the three subgroups relative to a typical trajectory.

2. What was the Range of Progress or Outcomes?

Seven categories of progress/outcome were generated to describe the progress and outcome of children in a quantifiable, meaningful way that was also clinically sensitive to progress in the heterogeneous sample. The categories were based on pre-post information including: Developmental rates, VABS Adaptive Behavior Composite (standard scores), cognitive test standard scores (if available), VABS Adaptive Behavior Composite Age Equivalent scores, CARS Total score and CARS category (see Perry et al., 2008 for a full description of categories). The mutually exclusive categories of progress were labelled as follows:

1) *Average Functioning:* standard scores at exit on cognitive and/or adaptive behaviour in the low average range or better, non-autism CARS scores.

2) *Substantial Improvement:* improved 10 points or more in standard scores on the VABS and/or a cognitive measure, clinically significant decreases on the CARS, typical rates of development during IBI, did not reach the low average range.

3) *Clinically Significant Improvement:* developmental rate .50 or greater and higher than their rate prior to IBI, improvements in VABS Adaptive Behavior Composite Age Equivalents and CARS scores.

4) *Less Autistic:* improved significantly on the CARS Total scores and/or category indicating improvement in diagnostic severity, but did not progress developmentally.

5) *Minimal Improvement:* developmental rate of .25 or greater during intervention, improved somewhat developmentally, based on VABS Adaptive Behavior Composite Age Equivalents.

Table 4. Categories of Progress/Outcome for Total Sample and as a Function of Initial Functioning Level

	Initial Functioning Level Subgroup			
	"Higher" **Functioning (*n*=11)**	**Medium Functioning (*n*=31)**	**Lower Functioning (*n*=42)**	**Total Sample (*n*=84)**
1) Average Functioning	5 (45.5%)	4 (12.9%)	0	9 (10.7%)
2) Substantial Improvement	1 (9.1%)	3 (9.7%)	2 (4.8%)	6 (7.1%)
3) Clinically Significant Improvement	4 (36.4%)	9 (29.0%)	16 (38.1%)	29 (34.5%)
4) Less Autistic	0	4 (12.9%)	4 (9.5%)	8 (9.5%)
5) Minimal Improvement	0	6 (19.4%)	11 (26.2%)	17 (20.2%)
6) No change	0	3 (9.7%)	5 (11.9%)	8 (9.5%)
7) Worse	1 (9.1%)	2 (6.5%)	4 (9.5%)	7 (8.3%)

Note. Progress/outcome categories could be determined for only 84 children. Initial functioning level was available for 85 children. No outcome category could be determined for one child in the medium functioning group. Percentages are % of children in initial functioning category (column) of the total available who had particular outcome (rows).

6) *No change:* rate of development very low (< .25) and/or unchanged during IBI, Age Equivalents on the VABS and CARS scores essentially unchanged.

7) *Worse:* zero or negative rate of progress during intervention and a worse rate during than prior to IBI, VABS Adaptive Behavior Composite standard scores were significantly lower (by 10 points or more), and stayed the same or became more severe on the CARS.

There was sufficient information to classify 84 children into these categories (Table 4). The majority of children (82%) improved in some measurable way during the time they were enrolled in TPAS; only 18% did not. Approximately 18% showed very successful outcomes, i.e., average functioning or substantial improvement (category 1 or 2). Another third of the group showed clinically significant improvements (category 3). Children in categories 4 and 5 (about a third of the sample) might be said to have showed some more modest benefits. An additional 8 children showed no change, and 7 were worse (at least on the measures used in the study, although severe self injury was eliminated in some).

The seven categories of outcome/progress were examined in each of the three initial levels of functioning subgroups. Table 4 shows that 10 of 11 children who were initially in the *"higher"* functioning group improved in some way (categories 1, 2, or 3), where 46% achieved average functioning. The group who started out in the *medium level* of functioning showed a wide range of outcomes, with some children falling in every category of progress, including 4 in the average functioning category. Almost 84% of these children could be said to have shown some progress, with over half the children (52%) showing at least clinically significant improvement (categories 1, 2, or 3). The *lower* functioning group also displayed considerable heterogeneity in outcome. Although none of these children achieved average functioning, 43% experienced substantial or clinically significant improvement (categories 2 and 3). An additional third (36%) showed more limited improvement in either development or diagnostic severity (categories 4 and 5), and 21% did not show any benefit (where 5 showed no change and 4 seemed to be doing worse at discharge).

Finally, a subgroup of 25 children was selected who roughly approximated Lovaas' sample—i.e., they were under age 4 at intake, received 2

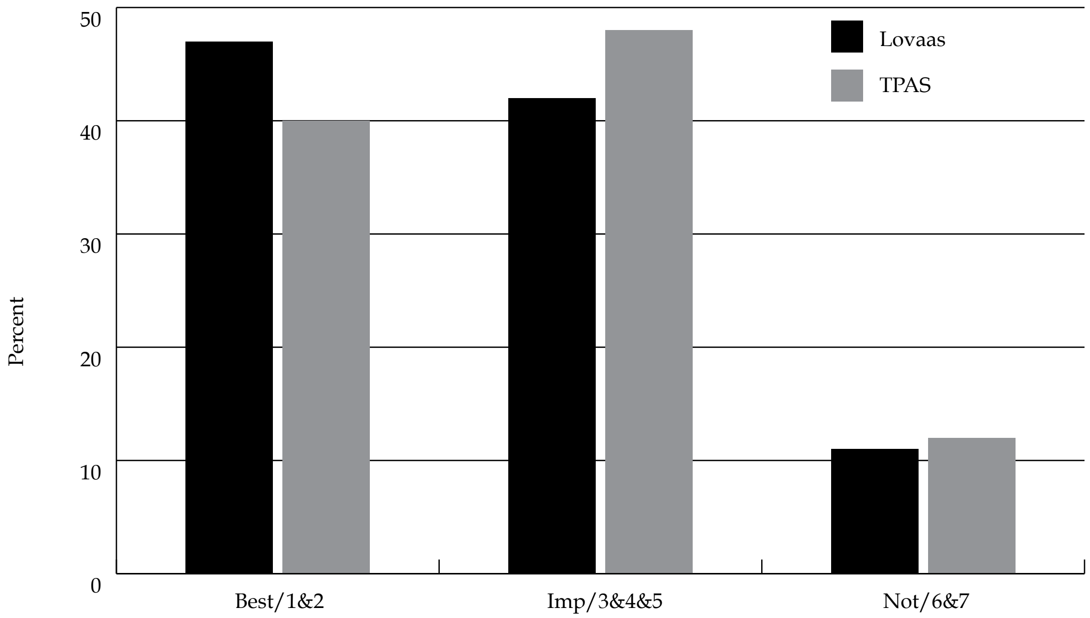

Figure 2. Percentage of children in Lovaas versus TPAS sample in three outcome groupings (first set of bars: Lovaas' best outcomes, present study categories 1 & 2; second set of bars: Lovaas' improved, present study categories 3, 4, & 5; third set of bars: Lovaas' not improved, present study categories 6 & 7)

years or more of IBI, and had a developmental level of at least the "medium functioning" level initially. Specific information was not available in terms of intensity of IBI for these children, though in principle, they were receiving 20 to 40 hours per week, compared to Lovaas' study where children reportedly received 40 hours per week. Of these 25 children, 9 were in the average functioning group following IBI (they constituted the entire group, in fact) and an additional 1 made substantial progress. Together these 10 children (40%) may be considered roughly comparable to Lovaas' 47% best outcome children (who were in regular classes, had IQs in the average range and retained no diagnosis). An additional 6 children in the TPAS subgroup were clinically significantly improved, 3 were less autistic, and 3 were minimally improved. Together, these 12 (48%) may be roughly similar to Lovaas' improved group (42% of his sample, children who were in special education classes for mildly delayed or language impaired children). Two children in the TPAS subgroup showed no progress and 1 was worse. These 3 (12%) may be roughly similar to Lovaas' 11% who were considered not improved (in classes for severely handicapped/autistic children). As shown in Figure 2, the proportions of children achieving various outcomes are quite comparable. However, this comparison includes only about one quarter of the children in the study (the others were older and lower functioning).

3. What Predicted Progress/Outcome?

Initial Developmental Level

Children's scores at exit were examined as a function of their initial scores on all measures. Scores on most assessment measures were significantly correlated between intake and discharge. Correlations ranged from .52 to .72 for the VABS scores, .54 for the CARS Total score, and .79 for FSIQ and .52 for MA (when available). Still, these correlations are not so high as to preclude factors other than initial level being predictive of outcome. In fact, in a series of regression analyses, initial levels of developmental variables accounted for the largest proportion of variance in all cases when age and duration were included in the regression. However, at most 50% of the variance was accounted for (Freeman & Perry, 2005).

Age at Entry

The importance of children's age when they began IBI was examined in several ways. Independent *t* tests were computed examining absolute levels of the exit assessment scores for two subgroups divided by age (before and after 4 years at entry) (see Table 5). Results indicate that younger children (who began

Table 5. Scores on Outcome Variables (Means and SDs) for Younger versus Older Age at Entry

	Younger *Age < 48*	*Older* *Age > = 48*		
CARS	**(*n* = 28)**	**(*n* = 60)**	*t*	*p*
• Total Score	28.86 (5.25)	31.89 (4.60)	2.76	.007
VABS Standard Scores	**(*n* = 28)**	**(*n* = 58)**		
• Communication	69.96 (28.00)	47.91 (13.00)	-5.01	< .001
• Daily Living	57.50 (16.97)	43.00 (12.90)	-4.40	< .001
• Socialization	62.61 (13.91)	53.00 (5.16)	-4.66	< .001
• Motor	65.14 (20.19)	59.43 (21.96)*	-0.94	ns
• ABC	59.86 (19.52)	44.64 (10.37)	-4.73	< .001
Cognitive Level	**(*n* = 20)**	**(*n* = 38)**		
• FS IQ Estimate	70.15 (34.33)	41.34 (20.57)	-4.00	< .001
• MA (months)	43.90 (21.26)	30.81 (15.95)	-2.63	.01

* *n = 21*

Table 6. Categories of Outcome for Younger vs. Older Children and Mean Age of Each Outcome Group

	n	*Younger (< = 48 months)* **n (%)**	*Older (> 48 months)* **n (%)**	*Mean Age at Entry M (SD)*
1) Average functioning	9	9 (30.0%)	0	33.33 (8.65)*
2) Substantial Improvement	7	2 (6.7%)	5 (8.6%)	60.43 (12.88)
3) Clinically Significant Improvement	29	8 (26.7%)	21 (36.2%)	55.76 (13.03)
4) Less Autistic	11	4 (13.3%)	7 (12.1%)	55.64 (11.99)
5) Minimal Improvement	17	3 (10.0%)	14 (24.1%)	53.94 (9.93)
6) No change	8	3 (10.0%)	5 (8.6%)	54.63 (10.90)
7) Worse	7	1 (3.3%)	6 (10.4%)	58.29 (9.55)
Total	88	30	58	53.57 (13.18)

** Group 1 significantly different from all other groups, none of which differ from each other*

IBI before age 4) scored significantly better at exit than older children on the CARS, VABS Communication, Daily Living, Socialization, ABC, IQ Estimate, and MA. Table 6 shows the numbers of children in each of the seven progress/outcome categories who were in the younger and older groups at program entry. Generally, younger children tended to fall into more optimal outcome categories (two-thirds falling in category 1, 2, or 3). Finally, the mean ages at entry were compared for children who ended up in each category of progress/outcome. The one-way ANOVA was significant ($F(6,81) = 5.487$, $p < .001$), and post-hoc tests indicated that Average Functioning children started significantly earlier (33 months) than all other outcome groups (54 months for the whole sample) but, apart from that, there was no particular trend for age at intake to be related to progress/outcome category.

Duration of IBI

Like age at entry, the duration of IBI was explored in several ways and showed a similar pattern of results (Table 7). First, looking at absolute outcomes on the developmental and diagnostic variables, a series of independent t tests comparing two groups in terms of duration (more than 2 years versus less than 2 years) on the discharge assessment values indicated that children who had longer durations scored significantly better at exit than children who had shorter ones on all variables except Motor. Table 8 indicates the numbers of children in each of the seven progress/outcome categories by duration of IBI. Generally, children with longer durations tended to experience more optimal outcome categories. Finally, the mean IBI durations (also shown in Table 8) were compared for children who ended up in each category of progress/outcome using a one-way *ANOVA*, which was significant ($F(6,81) = 6.505$, $p < .001$). As with the age analyses above, post-hoc tests indicated the Average Functioning children received IBI significantly and substantially longer than all other outcome groups and there were no other differences.

Predictors of Average Functioning

Given that nine children in the sample resembled the “best outcomes” group in Lovaas’ study, closer examination of these 9 children (8 boys and 1 girl) seemed warranted (see Table 9). The children were all in or very close to the autism range on the CARS initially (and were not “misdiagnosed”); at discharge, they were all in the non-autism range with a mean score of 23, which is well below the cut-off of 30. This represents a substantial change of more than 3 standard deviations in total score. Adaptive behaviour scores improved significantly for Communication, Socialization, and overall ABC, but not for Daily Living or Motor. Gains were quite dramatic in the Communication domain, in particular (more than 2 SDs). However, Adaptive Behaviour outcomes were considerably more variable and not quite as good overall as their cognitive outcomes (as has been reported by others in the literature) with a mean Adaptive Behavior Composite score of

Table 7. Scores on Outcome Variables (Means (SDs) as a Function of Duration of IBI

CARS	*Longer (> = 24 months)* **(*n* = 30)**	*Shorter (< 24 months)* **(*n* = 58)**	*t*	*p*
Total Score	29.02 (5.72)	31.91 (4.30)	-2.67	.009
VABS Standard Scores	**(*n* = 30)**	**(*n* = 56)**		
• Communication	68.17 (28.48)	48.09 (12.48)	4.54	< .001
• Daily Living	56.47 (17.13)	43.04 (12.94)	4.09	< .001
• Socialization	61.77 (13.95)	53.11 (4.99)	4.19	< .001
• Motor	64.62 (21.15)	59.90 (20.85)*	0.77	ns
• ABC	58.67 (19.82)	44.73 (10.09)	4.33	< .001
Cognitive	**(*n* = 21)**	**(*n* = 37)**		
• FS IQ	68.05 (34.91)	41.76 (20.61)	3.62	.001
• MA (months)	43.81 (21.39)	30.50 (15.55)	2.71	.009

** n = 20*

Table 8. Categories of Outcome for Longer vs. Shorter Duration of IBI and Mean Duration of Each Outcome Group

	n	*Longer (> = 24 months)* ***n* (%)**	*Shorter* ***(< 24 months)*** ***n* (%)**	*Mean Duration* ***M (SD)***
1) Average functioning	9	9 (30.0%)	0	35.56 (4.59)*
2) Substantial Improvement	7	2 (6.7%)	5 (8.6%)	16.29 (7.80)
3) Clinically Significant Improvement	29	7 (23.3%)	22 (37.9%)	17.21 (9.34)
4) Less Autistic	11	4 (13.3%)	7 (12.1%)	18.36 (10.00)
5) Minimal Improvement	17	3 (10.0%)	14 (24.1%)	18.18 (6.40)
6) No change	8	4 (13.3%)	4 (6.9%)	20.25 (9.18)
7) Worse	7	1 (3.3%)	6 (10.3%)	15.43 (8.12)
Total	88	30	58	19.47 (9.79)

** Group 1 significantly different from all other groups, none of which differ from each other*

81. Similar results were found for most of the Domain scores, though the Communication standard score was considerably higher with a mean of 99. Unfortunately, cognitive scores were available for only 3 children at intake, precluding pre-post comparisons. At discharge, however, their Full Scale IQ was in the average range (or above) with a mean of 105 with Performance (PIQ) substantially higher than Verbal (VIQ) (means of 118 versus 93 respectively) although verbal scores were in the average range for all but one child. Thus, at the time of discharge, these 9 children could be said to be in the non-autism range on the CARS and have no diagnosis of intellectual disability.

Finally, although these nine children typically had higher VABS scores than children whose outcomes were not as good, they were not all in the *"higher"* functioning group initially (5 were and 4 were in the *medium* functioning category). Looking at their initial diagnostic status, 3 had Autistic Disorder and 6 had PDD-NOS (a different proportion to the total sample) and they had an initial CARS score which was about 32 (somewhat milder than the 35.5 for the whole sample, but this difference was non-significant). Thus, it could be argued that they were slightly "milder" diagnostically than other children in the sample. In addition, as noted above, they were significantly and substantially younger when they entered the program (most were 3 years or younger on intake) and they received IBI for a significantly and substantially longer duration (about 3 years of IBI) relative to children with less optimal outcomes.

Discussion

This paper reported on the effectiveness of IBI in a large community setting, specifically, in the TPAS program. The results indicated that children showed statistically significant reduction in autism symptom severity on the CARS, where many fell into a milder category on the CARS by discharge; this was particularly true for those who scored more severely initially.

Table 9. Developmental and Diagnostic Results for Average Functioning Outcome Group (n = 9)

	Intake Mean (SD)	*Range*	*Exit Mean (SD)*	*Range*	*t*	*p*
Autism Severity						
• CARS Total	32.28 (2.77)	29.5–37.5	23.06 (2.14)	20.0–26.5	6.29	< .001
• Category	2 Not autism 6 Mild/Moderate 1 Severe		9 Not autism			
Adaptive Level						
• Communication	67.56 (12.88)	55–96	98.67 (27.29)	67–132	-3.17	.01
• Daily Living	66.33 (5.15)	60–75	72.67 (18.59)	57–114	-1.08	ns
• Socialization	64.00 (5.88)	58–72	76.89 (14.14)	59–105	-2.50	.04
• Motor	70.44 (23.68)	16–101	87.67 (11.66)	71–104	-1.81	ns
• ABC	63.44 (6.35)	55–74	80.56 (19.48)	60–117	-2.48	.04
Cognitive Level	*(n = 3)*		*(n = 8)*			
• Verbal IQ			93.25 (8.28)	83–105		
• Performance IQ			118.25 (10.26)	105–133		
• FSIQ	75.67 (33.82)	50–114	105.50 (7.41)	97–117		
Rate of development	.49 (.15)	.36–.84	.97 (.51)	.46–2.08	2.45	.04
Age/Duration	Age	Range	Duration	Range		
(months)	33.33 (8.65)	20–47	35.55 (4.59)	27–42		

Children also gained significantly in developmental skills (increased age equivalents) in all areas of adaptive behaviour; however, standard scores, which are corrected for age, were generally stable. Significant improvement was also found in cognitive level among those children for whom this information was available. Children's rate of development during IBI was approximately double the rate prior to IBI, and this was true for all three initial rate groupings.

Children were classified into seven categories of progress/outcome, and the vast majority of children showed some benefit or improvement during IBI. Further, 11% achieved average functioning, similar to the term "best outcomes" in the literature (i.e., scoring in the non-autism range and average range cognitively). A subset of children selected as being roughly equivalent to Lovaas' sample prior to IBI were found to have a comparable outcome pattern. Children's progress/outcomes were clearly related to their initial functioning levels (though not totally); children who started IBI before age 4 did better than those who started after age 4, and age at entry was a strong predictor of children achieving average functioning (but there were no clear patterns of prediction for other progress/outcome groups); and children who received 2 or more years of IBI did better than those who received less than 2 years. Duration was a strong predictor of children achieving average functioning but, again, there were no simple patterns of prediction for other progress/outcome groups, and duration was confounded with initial age.

Relative to studies of IBI in the literature, the present study has certain strengths, including a large sample size (i.e., 89 children), and considerable heterogeneity on virtually all variables examined. Thus a number of statistical analyses were possible which have been precluded in other published studies. Furthermore, it is noteworthy for being a study of real world effectiveness in a large-scale community application of an intervention which has documented efficacy in model programs.

However, there are also several limitations which should be noted. First, there is no comparison group of similar children who received no treatment or a different treatment. Thus, any changes seen during IBI cannot be conclusively attributed to the IBI. However, the comparison of the developmental trajectories pre-IBI versus during IBI and the subgroup comparisons within the data compensate for this to a certain extent. It should be noted that a comparison group study is currently underway. Second, there is no measure of treatment integrity or fidelity, meaning that there is no evidence that children were receiving IBI of a similar quantity and quality to that in the model program studies reported in the literature. However, it is known that the TPAS program had standards in terms of staff qualifications and training, staff-child ratios, and staff-supervisor ratios. Third, the measures are limited and do not necessarily tap all possible changes of interest (e.g., problem behaviour), have imperfect reliability and validity, and were not always available for all children. For example, cognitive scores were only available for about one-quarter of the sample, and these scores resulted from the use of several different IQ tests (which often differed between intake and exit). Finally, the people responsible for both intake and discharge assessments were not blind to the children's participation in IBI, nor were they independent of the organization providing the IBI. Further, the research team could be considered biased toward demonstrating the effectiveness of the program.

Nevertheless, the results of this study are important because they show that IBI can be effective in a community setting, under less-than-ideal conditions. In this study, the vast majority of children showed some measurable progress or improvement during their time in the TPAS program, and remarkably, their rate of development doubled. By the time of discharge, some children even showed average functioning, which is comparable to children from model programs referred to as "best outcomes." Although the rate of good outcomes for the present sample as a whole is not as high as in efficacy studies of model programs, when parameters such as children's age, duration of IBI, and developmental level are explored in relevant subgroups of the sample, outcomes approximate many of the published efficacy studies.

Acknowledgements

We would like to thank John Flannery, Chief Executive Officer, of Surrey Place Centre and Elizabeth Scott, Vice President Autism Services, for their support and encouragement of this project. Surrey Place Centre is the lead agency for TPAS, and thanks go to the partner agencies: Adventure Place, Aisling Discoveries, Etobicoke Children's Centre, and Geneva Centre, and all the children and families. Thank you to Sally Bond, Alissa Levy, Abbie Solish, and April Sullivan for assistance with data collection, entry, and verification.

References

Anderson, S. R., & Romancyzk, R. G. (1999). Early intervention for young children with autism: Continuum-based behavioral models. *Journal of the Association for Persons with Severe Handicaps, 24*, 162–173.

Bayley, N. (1993). *Bayley Scales of Infant Development (2nd ed.)*. San Antonio, TX: Psychological Corporation.

Dawson, G., & Osterling, J. (1997). Early intervention in autism. In M.J. Guralnick (Ed.), *The effectiveness of early intervention* (pp. 307–326). Baltimore, MD: Brookes.

Eikeseth, S. (2001). Recent critiques of the UCLA Young Autism Project. *Behavioral Interventions, 16,* 249–264.

Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive behavioral treatment at school for 4- to 7-year-old children with autism: A 1-year comparison controlled study. *Behavior Modification, 26*(1), 49–68.

Freeman, N. L., & Perry, A. (2005). Child outcomes of Intensive Behavioural Intervention (IBI) in the Toronto Preschool Autism Service (TPAS): A program evaluation report. Unpublished manuscript. Surrey Place Centre, Toronto, ON.

Green, G. (1996). Early behavioral intervention for autism: What does the research tell us? In C. Maurice, G. Green, & S. C. Luce (Eds.), *Behavioral intervention for young children with autism: A manual for parents and professionals* (pp. 29–44). Austin, TX: PRO-ED.

Handleman, J. S., & Harris, S. L. (Eds.). (2001). Preschool education programs for children with autism (2nd ed.). Austin, TX: PRO-ED.

Harris, S. L., & Handleman, J. S. (2000). Age and IQ at intake as predictors of placement for young children with autism: A 4 to 6 year follow-up. *Journal of Autism and Developmental Disorders, 30,* 137–142.

Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities, 26*, 359–383.

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 (2nd ed.)* (pp. 691–710). Washington, DC: American Psychological Association.

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55(1)*, 3–9.

McEachin, J. J., Smith, T., & Lovaas, O. I. (1993). Long-term outcome for children with autism who received early intensive behavioural treatment. *American Journal on Mental Retardation, 97*, 359–372.

Ministry of Community and Social Services (MCSS) (2000). *Program guidelines for the Ontario Intensive Early Intervention Program.* Toronto, ON: Author.

Mullen, E. M. (1995). *Mullen Scales of Early Learning.* Circle Pines, MN: American Guidance Service.

National Research Council (2001). *Educating children with autism.* Committee on Education and Interventions for Children with Autism. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

New York State Department of Health (1999). *Autism/Pervasive Developmental Disorders. Clinical practice guidelines technical report.* New York: Author.

Perry, A. (2002a). Intensive early intervention program for children with autism: Background and design of the Ontario preschool autism initiative. *Journal on Developmental Disabilities, 9* (2), 121–128.

Perry, A. (2002b). *Autism and intensive behavioural intervention: The empirical background to the Ontario Intensive Early Intervention Program.* Report prepared for the Ministry of the Attorney General, Government of Ontario.

Perry, A., & Condillac, R. A. (2003). *Evidence-based Practices for Children with Autism Spectrum Disorders*. 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., Dunn Geier, J., Freeman, N. L., Hughes, S., LaRose, L., et al. (2008). Effectiveness of intensive behavioural intervention in a large, community-based program. *Research in Autism Spectrum Disorders, 2,* 621–642.

Perry, A., Flanagan, H. E., Dunn Geier, J., & Freeman, N. L. (2009). Brief report: The Vineland Adaptive Behavior Scales in young children with autism spectrum disorders at different cognitive levels. *Journal of Autism and Developmental Disorders, 39,* 1066–1078.

Perry, A., & Freeman, N. L. (1996, March). Psychometric properties of the Childhood Autism Rating Scale (CARS). Poster presentation at the Ontario Association on Developmental Disabilities Research Day, Toronto, ON.

Powers, M. D. (1992). Early intervention for children with autism. In D. E. Berkell (Ed.), *Autism: Identification, education, and treatment* (pp. 225–252). Hillsdale, NJ: Lawrence Erlbaum.

Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioural treatment for children with autism: Four-year outcome and predictors. *American Journal on Mental Retardation, 110*, 417–438.

Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *The Childhood Autism Rating Scale (CARS).* Los Angeles: Western Psychological Services.

Schreibman, L. (2000). Intensive behavioural/psychoeducational treatments for autism: Research needs and future directions. *Journal of Autism and Developmental Disorders, 30,* 373–378.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales.* Circle Pines, MN: American Guidance Service.

Thorndike, R. L., Hagen, E. P., & Sattler, J. S. (1986). *The Stanford-Binet Intelligence Scale: Fourth Edition (SB:IV)*. Chicago: Riverside.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence-III.* San Antonio, TX: Psychological Corporation.