

Follow-up Study of Youth who Received EIBI as Young Children

Adrienne Perry

York University, Department of Psychology

4700 Keele Street, Toronto, ON M3J 1P3

(416) 736-2100 ext. 33765

perry@yorku.ca

Julie Koudys

Alice Prichard

Hilda Ho

Abstract

Although early intensive behavioral intervention (EIBI) has been clearly shown to be evidence-based, there is very little information available regarding long-term outcomes, especially from community effectiveness studies. We present data on cognitive, adaptive, and autism severity measures from four time points (pre- and post-EIBI and two follow-up points) for a sample of 21 youth, currently aged 16 years on average (range = 13-20) who received EIBI as young children and who have been out of EIBI for a mean of 10 years (range = 8.5-14). Results show heterogeneous outcomes and a general pattern of stability since the end of EIBI, suggesting gains made in EIBI are maintained.

Follow-up Study of Youth who Received EIBI as Young Children

Despite the abundant evidence supporting the efficacy of Early Intensive Behavioural Intervention (EIBI) for young children with autism (e.g., Eldevik et al. 2009 meta-analysis), as well as its effectiveness in community settings in Ontario (Perry et al., 2008), surprisingly little research has explored whether these positive gains are maintained over time after discontinuation of EIBI (Matson & Smith, 2008). Considering the enormous investment of public and private funds as well as people's hopes and energy, this is a serious limitation in the literature.

Three studies have reported on outcomes following completion of EIBI. One of these studies was conducted by McEachin, Smith, and Lovaas (1993), who examined the outcomes of the original 19 participants of the seminal Lovaas (1987) study about five years after cessation of EIBI, at the age of 9 to 19 years ($M=13$). The nine "best outcome" children who had been average functioning by the end of EIBI (at about age 7) had been out of treatment for a mean of 5 years (range of 3 to 9 years). Those who had not achieved best outcomes, and may have continued longer in treatment, had also been out of treatment for a mean of 5 years (range of 0 to 12 years). Results indicated that, at follow-up, the EIBI treatment group had significantly higher IQ than the control group and group means were similar in both groups from the previous testing at the end of treatment. Thus it seems that intensively treated children, as a group, maintained their gains and continued to have an advantage over children who received the minimal treatment. Individual scores are not reported for this part of the study, however.

The nine best outcome children, who had shown substantial improvement in EIBI, were examined more closely in comparison to a matched group of typically developing children. Examiners were blind to children's history. The best outcome children had IQs that were in the average to high average range and adaptive behavior scores that, though somewhat lower than IQ, were primarily in the average range as well. Furthermore, the majority of clinical ratings (maladaptive behavior, personality) were similar to

those of the typically developing group, with the exception of one individual who exhibited more extreme scores.

More recently, Kovshoff, Hastings, and Remington (2011) conducted a 2-year follow-up study of the sample originally studied by Remington et al. (2007). They were able to contact all 23 participants in the original EIBI group and the majority of the treatment-as-usual comparison group. The 23 EIBI children had received 2 years of intervention via either a University-affiliated program or a private service option. Follow-up data were collected 2 years after the end of the 2-year study period, when the children were about 7 years of age. The time intervals were the same for all participants, which is a strength of this study. A higher proportion of the EIBI group were in mainstream educational settings (61% vs. 22% in the comparison group) although most had support in the classroom. ANCOVAs for follow-up scores, controlling for baseline scores, did not differ between groups for IQ, adaptive, or behaviour ratings. However, there was an unanticipated difference between the two EIBI delivery models such that scores of the children in the University-based program declined following treatment termination whereas the parent-commissioned intervention group remained fairly stable. This is difficult to interpret, however, because there were a number of differences between these two subgroups pre-intervention. For purposes of the present study, the key finding is that some children maintained their gains but others did not, perhaps for a combination of reasons.

Finally, a small study in Ireland (O'Connor & Healy, 2010) provided detailed follow-up data on five children aged 9 to 12 at the time of the follow-up study. These five boys had received school-based EIBI for two or more years, had improved to the point of functioning in the low average to average range, underwent a 1- to 2-year transition period into mainstream classrooms (ending by the age of 9 on average), and had subsequently been receiving mainstream educational services for about 2 years (ranging from 10 months to 3 years, 9 months), with special needs support. At follow-up, three of the

five children showed quite good stability in cognitive and adaptive scores since the end of EIBI, whereas scores for the other two decreased by one to two standard deviations.

In summary, follow-up studies of EIBI are rather sparse and the follow-up period ranges from 2 to 5 years, on average. There has yet to be a study published with longer term results, reporting on youth or young adults who received EIBI as young children. In the present paper, we will present data on a 10-year follow-up (8-14 years). The few studies in the literature also have different methodological constraints such as small samples, concerns regarding the representativeness of their sample, the range of outcome measures (often only IQ and adaptive skills), and the inherent variability within the sample on dimensions of interest (e.g., initial IQ, time in treatment, time out of treatment), some of which are, unfortunately, common to the present study as well, although we report a wider range of outcomes.

An additional methodological concern, common to many reports on EIBI outcomes, is related to the correct interpretation of individual and group test score changes over time. In a group of children, many of whom are functioning in the range of intellectual disability, declining standard scores are to be expected, even though raw scores and age equivalent scores are increasing, because many of the individuals in the sample are learning, but at a slower than average rate. Further, means may be misleading and group comparisons may obscure individual gains or losses. It would be helpful to consider changes on standardized measures using the concept of clinical significance (Kazdin, 2005) in individual data (i.e., meaningful gains or losses relative to the individual's own starting point) as well as statistical significance in group data, both of which are included in the current paper.

The Ontario EIBI Program

The large province-wide, publicly funded Ontario EIBI program has been well researched (Flanagan, Perry, & Freeman, 2012; Freeman & Perry, 2010; Perry et al., 2008; 2011). More specifically, Perry et al. (2008) reported on a retrospective file review of 332 children aged 20 to 86 months at intake, who had

received EIBI for a mean of 18 months, from all nine of the provincial programs. There were statistically significant improvements in cognitive and adaptive behaviour, and significant reductions in autism symptom severity. Simply, results indicated that 25% could be described as having good outcomes (average range or substantially improved), 50% medium outcomes, and 25% poor outcomes (see Perry et al., 2008 for further details).

In 2007-8, we conducted a short-term follow-up study of 36 children from the Ontario program (not necessarily participants in the Perry et al. [2008] study). Participants were recruited in several ways: advertisements in various newsletters and websites, direct contact by one program with recently discharged families, and mailings sent out by three other programs. This study formed the basis of Prichard's (2011) unpublished dissertation. The 36 children were between the ages of 5 and 12 and had been discharged from EIBI for about 3 years. Developmental and diagnostic variables (IQ, adaptive skills, autism symptom severity) were compared at three time points: prior to EIBI; upon completion of EIBI; and at follow-up 1 to 6 years after the completion of EIBI.

The Present Study

The current study is a further follow up of Prichard's (2011) group approximately 8 years later, when the participants were between 13 and 20 years old. Thus, data are presented here from four time points: Time 1 was prior to EIBI (retrospective file review conducted by Prichard [2011]); Time 2 was upon completion of EIBI (retrospective file review conducted by Prichard [2011]); Time 3 was at follow-up 1 to 6 years after the completion of EIBI (Prichard [2011] data collection by current authors which took place in 2007-08); and Time 4 was at follow-up a mean of 10 years after the completion of EIBI (data collection by current authors in 2015-16).

We report here on results addressing the following research questions: 1) What is the current profile (i.e., at Time 4) of this group of adolescents on a wide range of developmental and diagnostic

measures?; 2) Do group scores change significantly over time on three primary outcome measures: cognitive level, adaptive skills, and autism severity, and categorical outcome?; and 3) What is the pattern of stability or change in individual scores?

METHOD

Participants

We attempted to contact all participants from the Prichard (2011) study except one with Rett syndrome. Of this maximum sample of 36, seven were lost to follow-up and seven declined to participate. One additional family was impossible to schedule because of distance. Thus, we obtained a final sample of 21 youth (19 male; 2 female) from 18 families (two families with two siblings; one set of twins) from various regions of Ontario. Further child and family characteristics are shown in Table 1.

Insert Table 1 about here

In order to assess how representative the final follow-up sample was of the original larger sample, we examined Time 3 scores on a number of key variables for the 21 participants in the current study compared to the other 15 (of the original 36) who did not participate in the current study. There was a trend for the current sample to have had higher cognitive and adaptive scores and milder autism severity at Time 3, although the only comparison that was statistically significant was for IQ at Time 3 ($t(34)=2.51, p=.02$).

The inclusion criteria for the Prichard (2011) study included the requirement that children had not been receiving EIBI since they had been discharged from the program. In the present study, there

were no such exclusions, but we interviewed parents about the nature of school programs and other services received since the end of the EIBI program. As shown in Table 2, nine children had been in special education classes throughout their schooling, six had been in a regular class (though with some minimal extra resourcing), and six had been in a mixture of placements (some special education class; some regular classroom; either simultaneously or sequentially). As families received educational services from different school districts with different philosophies and types of classroom options, this information is somewhat difficult to interpret.

In terms of other social/mental health services, the majority of children had received occupational therapy at some point and about half had received speech therapy. The majority also obtained some services through a children's mental health centre (primarily social skills groups). None had received intensive behavioural services after leaving the EIBI program. Most parents participated in at least one behavioural parent training program (some did much more), and most families had received some form of respite care at some stage. Many of these services were of minimal intensity and duration (e.g., a 2 hour/week group that lasted 8 weeks).

Insert Table 2 about here

Procedure

Following ethical approval from both Universities employing the first two authors, a letter was sent to families describing the study and then families were contacted by phone or email. Depending on their location, they were invited to come to the University employing the first author or we travelled to their homes in various centres in Ontario. In either case, a team involving one of the first two authors,

together with two or three graduate students, conducted the assessments, beginning with consent by the parent(s) and consent or assent with the youth. Although aware of the child's treatment history generally, and the purpose of the study, assessors were blind to the child's previous scores. None of the authors had played any role in the children's treatment. The parent interview included information on services the child and family had received since EIBI ended, as well as the Vineland Adaptive Behavior Scales and the Child Behavior Checklist. The youth were given an assessment of cognitive level and academic achievement (when possible). The assessment methodology was consistent with that used for the Prichard (2011) study. Following completion of the assessment, with parent consent, we contacted the children's teachers to complete the Teacher Rating Form. Parents received a full psychological report of the findings and had the option of an in-person feedback session.

Measures

Cognitive Level. Cognitive measures used at Time 4 included primarily the Wechsler Intelligence Scale for Children, 4th edition (WISC-IV; Wechsler, 2004; $n=14$), but also a few individuals were assessed with the Wechsler Adult Intelligence Scale ($n=1$), the Stanford-Binet Intelligence Scale ($n=2$), or the Mullen Scales of Early Learning ($n=4$). On this latter test, which was used for individuals with low skill levels who were older than the oldest norms, Mental Age (MA), based on the median of the four subtest age equivalent scores (i.e., Visual Reception, Fine Motor, Expressive Language, Receptive Language), was used to compute Ratio IQs ($MA/CA \times 100$). Various cognitive measures were used across children, and across time in some cases, but we used the same test at Time 4 as was used at Time 3 whenever possible.

Adaptive Behavior. Parents were interviewed using the Vineland Adaptive Behavior Scales 2nd edition, Survey Form (Sparrow, Cicchetti, & Balla, 2005), or the earlier edition for Time 1 and Time 2, which yields domain standard scores for Communication, Daily Living Skills, and Socialization, as well as the

Adaptive Behaviour Composite (ABC) score. Age equivalent scores can also be derived and were used for some analyses.

Autism Symptom Severity. We used the Childhood Autism Rating Scale (CARS; Schopler, Reichler, & Renner, 1988) as a measure of autism severity, as it had been used at the earlier time points. The CARS involves a trained observer rating the individual on 15 items, each on a 7-point scale. Total scores can range from 15 to 60 and scores above 30 are deemed to be in the autism range. A consensus approach was used to obtain ratings by pooling observations by team members who had assessed the child or interviewed the parents, led by one of the first two authors, who had previously achieved inter-rater agreement.

Academic Achievement. As a measure of basic academic skills, we used the Wide Range Achievement Test, 4th edition (WRAT-4; Wilkinson & Robertson, 2006), which is considered an achievement screening test. Specifically, the Reading (decoding), Spelling, and Math subscales were used for participants who were able to complete them.

Social-emotional-behavioural Functioning. The Achenbach System of Empirically Based Assessment (ASEBA; Achenbach, 2009) paper and pencil questionnaires were used to assess the youth's Internalizing, Externalizing, and Total Problems. This included the Child Behavior Checklist (CBCL) for parents and the Teacher Rating Form (TRF).

Outcome Categories. Participants were classified into one of three categories of outcome at each time point, based on a combination of available information from the IQ test, Vineland ABC, and CARS scores. The "good" outcome category was defined as: IQ and/or ABC >84 and CARS <30. The "medium" category involved: IQ and/or ABC between 40 and 84 and CARS <37.5; and the "poor" outcome category was defined as: IQ and/or ABC <40 and CARS >30 (anywhere in the autism range). These were the same criteria used by Prichard (2011), but the first two authors of the present study independently re-rated all

children for the present study so that ratings would be as consistent as possible. In order to minimize an expectancy bias, ratings were made for the all the children at one time point, independent of the other time points, rather than rating one individual across all time points. A few sets of scores did not exactly fit these categories (and sometimes one score was missing) so some clinical judgement was required. Inter-rater agreement for these ratings was 90%. Disagreements were resolved by consensus in the direction of the more conservative category.

Demographics. Demographic information regarding the family was gathered during the interview, including the parents' education and occupation, which were used to derive socioeconomic status (SES) scores after the method of Barratt (2006).

RESULTS

1. Characteristics 10 years after EIBI

At Time 4, there was wide variability among the sample on all outcome measures, as shown in Table 3. FSIQ ranged from the profound ID range to the average range. For those who received the WISC-IV and, therefore, had separate verbal and nonverbal IQ, scores were slightly higher for nonverbal IQ but note the large standard deviations for both. Adaptive behaviour scores were similarly variable, ranging from severely delayed to average. Note that the FSIQ and ABC score means were virtually identical at 65. Most of the youth were able to complete some basic academic tasks and these scores ranged from the floor of the test to the high average range. Math scores were, on average, slightly lower than those for reading and spelling. In terms of autism symptom severity, scores on the CARS ranged widely as well, from well below the autism cut off of 30 to the severe autism range. Fifteen children (71%) scored in the non-autism range; five (24%) in the mild/moderate range; and one (5%) in the severe autism range. The social-emotional-behavioural functioning of the participants, as rated by parents (CBCL) and teachers (TRF) was also somewhat variable. Mean *T* scores were predominantly in

the normal range (40 to 60) and tended to be lower (i.e., more typical) from the parents' perspective than from that of the teachers'.

Insert Table 3 about here

Three quite distinct subgroups were seen in the sample at Time 4 (see Table 4). One third of the sample ($n=7$) were considered to have *Good* outcomes. These youth had IQ scores in the average range, but substantially lower adaptive scores, and had mean CARS scores well below the autism range. Almost half the adolescents ($n=10$) fell in the *Medium* group, with cognitive and adaptive scores that were consistent and in the mild ID range and mean CARS scores in the non-autism range. The *Poor* outcome group ($n=4$) had very low IQ and adaptive level (in the severe ID range) and CARS scores that were in the autism range.

Insert Table 4 about here

2. Group Changes over Four Time Points

Changes in scores across time were examined using repeated measures ANOVAs followed by pairwise comparisons. In the case of IQ, there was too much missing data at Time 1 ($n=7$) so only three time points were compared. Results are shown in Table 5. Further, since the comparison of greatest interest in the present paper is the longer term outcomes from Time 3 and Time 4, assessment methodology was more consistent between those times, and we have close to complete data for those

two time points, we also conducted *post hoc* paired *t* tests comparing Time 3 and Time 4 scores, as shown in Table 6.

Insert Table 5 and Table 6 about here

For these analyses, it would be expected that scores would increase from Time 1 to Time 2 (during EIBI) and then, ideally, would remain stable thereafter from Time 2 to Time 3 and from Time 3 to Time 4. However, as noted earlier, in a group progressing at a slower rate of development, standard scores would tend to decline even when age equivalent scores were increasing. Therefore, for these comparisons, we report on both standard scores and age equivalent scores for cognitive and adaptive behaviour. During the approximately 7-year period between the first and second follow-up assessments (i.e., Time 3 to Time 4), standard scores for both cognitive and adaptive measures decreased significantly (with small ES), while age equivalent scores both increased significantly (with large ES). Thus, on average, participants have continued to improve in their skill level but are falling further behind their same-aged peers because they are developing at a slower rate. However, there are several quite significant measurement issues that may be contributing to lower scores at Time 4 (see Discussion). Autism severity became significantly milder in the group as a whole (with small ES).

Table 7 shows the proportions of participants classified into the three outcome categories at each of the four time points. At Time 1 (pre-EIBI) the majority of children were in the Medium category and none was in the Good category. After EIBI, as expected, a significant number of children (38%) moved to the Good category, the same proportion were in the Medium category, and 10% had Poor outcomes (also 14% were n/a because they were missing two of the three scores needed to derive the

category). At Time 3, about 3 years after EIBI had ended, the proportion of children in the Good category remained the same, suggesting that these children had maintained their gains. Another 43% were classified as Medium, and the remaining 19% of children were in the Poor outcome group. At Time 4, approximately 10 years after EIBI, these proportions were roughly similar.

Insert Table 7 about here

Group comparisons may, of course, obscure potential individual changes (in either direction) so we now turn to analyses of clinically significant changes in individual data.

3. Individual Data and Clinically Significant Changes

Figures 1 to 4 illustrate the variability in individual data across time points for IQ, Vineland standard scores (ABC) and age equivalents, and CARS scores respectively. Comparison of the two types of Vineland scores illustrates the statistical finding noted earlier whereby standard scores decrease slightly, on average, whereas age equivalents increase substantially (though variably).

Insert Figures 1, 2, 3, and 4 about here

Clinically significant change was operationalized as approximately 1 SD, i.e., 15 points on IQ and ABC and 5 points improvement on CARS scores, consistent with definitions used in Perry et al. (2008; 2011). Table 8 provides data on the clinically significant change from Time 2 (end of EIBI) to Time 3

(Prichard, 2011) as well as from Time 3 to Time 4. We report the percentage of the sample showing stable scores and the raw number of individual youth who increased or decreased to a clinically significant degree.

Insert Table 8 about here

Looking at clinically significant change from the end of EIBI to the first follow-up (Prichard, 2011), about 3 years later, the vast majority of cognitive scores were stable with only one child showing a clinically significant decline. Two children showed a decline in adaptive scores, but three showed a clinically significant increase. CARS scores showed one child scoring significantly higher (i.e., more severe) and three children significantly improved. From Time 3 (approximately 3 years after EIBI) to Time 4 (approximately 10 years after EIBI), similarly, 90% of the cognitive scores were stable but two showed clinically significant declines. Adaptive scores were 81% stable with four youth showing significant declines. Interestingly, CARS scores were least stable (71%) and most likely to show improvement (five youth clinically significantly improved) though one participant was clinically significantly worse.

Looking at the consistency of category placement after EIBI across time within individual data, shown in the last row of Table 8, 88% of participants remained in the same category from Time 2 to Time 3 and 76% from Time 3 to Time 4. In each case, two children in the Good category moved to the Medium category. A few individuals in the Medium category changed categories and this happened in both directions with one child moving to the Good category and one to the Poor category in each time

interval comparison. Children in the Poor category tended to remain there, although one improved to the Medium category between Time 3 and Time 4.

DISCUSSION

In this study, we followed up a group of 21 adolescents with ASD who had received publicly funded community-based EIBI as young children. We reported both group and individual data for a range of outcomes, principally focusing on cognitive, adaptive, and autism severity measures taken at four time points: prior to EIBI, at the completion of EIBI, about 3 years later, and about 10 years after the completion of EIBI. There was substantial heterogeneity in the sample on all measures. Considering group data, although children's age equivalent scores increased substantially over time, there were, on average, small but significant decreases in standard scores (IQ and ABC) since the end of EIBI, but this is not surprising given the nature of standard scores. Children were classified into three outcome categories (good, medium, and poor) and their classifications were quite stable over time, although a few individuals changed categories (in each direction). When we examined individual data more closely, we saw that the general pattern was one of stability (within 1 SD) in scores for the majority of participants, i.e., no clinically significant changes in either direction. Thus, in general, scores and classifications based on post-EIBI scores tended not to change very much over time up to 10 years later. There was no evidence for a precipitous decline in scores when EIBI ended, as is sometimes feared. Furthermore, children's autism symptom severity tended to be milder over time and their social-emotional scores at Time 4 were generally similar to those of normative samples.

There are a number of limitations to the study. The sample in the present study was not large ($n=21$) although it was larger than many of the EIBI study groups in the literature. The sample may have been somewhat biased in that it was a volunteer sample and may not have been completely representative of the original sample, or of children in the Ontario EIBI program generally. However, the

sample characteristics did include a very wide range of child characteristics and some demographic diversity in the families. Thus, there is no reason to suppose the outcomes of the individuals in this study are necessarily different from those of other children served by the Ontario EIBI program at the same time. The sample and research design were based mainly on convenience. We relied on data collected from clinical settings retrospectively for the Time 1 and 2 scores. There was considerable variability in the age starting EIBI, the age ending EIBI, the interval until the first follow-up and the interval to the current follow-up. All of these factors contribute to "noise" in the dataset. Most importantly, the study was uncontrolled and, hence, "outcomes" cannot strictly be attributed to the children having received EIBI as young children. Finally, the assessors were not unaware of children's treatment history.

Several important cautions are in order regarding the interpretation of specific test scores over time. Recall that Time 1 and Time 2 data were obtained by Prichard (2011) by file review (with parents' consent) and so the availability and quality of these data are less than ideal. For IQ there was considerable missing data (only seven children had scores at T1 and 14 at Time 2). There was a mixture of IQ tests across children and across time, and combination of deviation IQs and ratio IQs, although at Time 3 and Time 4 (when we assessed all children ourselves) we used consistent tests wherever possible. Because the ratio IQs were the only option for the poor outcome individuals (because there were no appropriate norms for their age and functioning level), this could have contributed to the significantly lower mean IQ scores over time. However, recall that over 90% of individual participants had stable IQ scores from Time 2 to Time 3 to Time 4. The Vineland Adaptive Behavior Scale was used consistently as the adaptive measure but the older edition was used at Time 1 and Time 2 and the second edition (Vineland-II) was used at Time 3 and Time 4. Even more important, standard scores on the Vineland-II may appear to decrease over time only because the floor of the test is lower in older normative groups (Perry, Flanagan, Dunn Geier, & Freeman, 2009; Flanagan et al., 2012). For example, the lowest standard score at age 5 may be in the 60s, whereas the lowest score at age 18 may be in the

20s. Furthermore, the Vineland has rather modest reliability and is a parent report measure with the inherent subjectivity that implies. The CARS was used at all time points but the raters who did the CARS ratings at Time 1 and Time 2 may have been more inconsistent in their methods, whereas all raters at Time 3 and Time 4 were similarly trained and did ratings in a consistent manner. The CARS was designed for and has been used primarily with younger children (Schopler et al., 1988) but has also been used reliably in samples with wide age ranges (e.g., Perry, Condillac, Freeman, Dunn Geier, & Belair, 2005). However, it was our clinical impression, for the good outcome youth in the present study, that the CARS was not very sensitive to more subtle manifestations of autism. Thus, improvements in CARS scores and low autism severity at follow-up may be somewhat overstated. Other ASD measures were also administered and agreement was very poor across measures (Esteves, Taheri, Perry, & Koudys, 2016). However, since the CARS has been used in this sample previously and in the literature, and seemed to be suitable for the medium and poor outcome subgroups, we chose to report it here.

Notwithstanding these limitations, this is one of the very few studies that has tracked children who received EIBI as young children and it is, by far, the longest follow-up study to date. It is all the more relevant to the question of long-term maintenance of EIBI gains since the original treatment was conducted in various community-based programs across Ontario's publicly funded provincial program.

In terms of clinical and policy implications, the fact that a significant number of these youth do demonstrate substantial gains during EIBI and appear to maintain these gains, provides good justification for the value of the program for these children and their families. The study also provides some data relevant to the question of eligibility and suitability of the program for children with very severe levels of intellectual disability whose scores essentially do not change as illustrated in the graphs. Unfortunately, EIBI may not be very effective for these individuals, at least in terms of these standardized outcome measures.

More research is needed to further explore the long-term effectiveness of EIBI. In particular, larger, prospective controlled studies would be ideal, using well-specified samples with independent assessors blind to group membership, using standardized measures, at regular time intervals (e.g., every 2 years). The nature of the treatment should also be well characterized including measuring and taking into account duration, intensity, treatment fidelity or quality, staff qualifications, family involvement, additional treatments and educational programming. This is an ambitious endeavor but one that merits researchers' attention.

In summary, EIBI is considered a well-established treatment for young children with autism, resulting in significant gains in cognitive and adaptive functioning, but very little information is available regarding long-term outcomes. We found generally stable scores and classifications of outcome about 10 years after cessation of EIBI. Such long-term benefits have not previously been documented and are a crucial piece of evidence justifying the extensive resources required to mount this intervention. Clearly, more research is needed on the long-term effectiveness of EIBI across a range of other samples and settings, ideally studies that will address some of the methodological issues inherent in the present study.

References

- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Barratt, W. (2006). *The Barratt Simplified Measure of Social Status (BSMSS) Measuring SES*. Unpublished Report. Indiana State University.
http://wbarratt.indstate.edu/socialclass/Barratt_Simplified_Measure_of_Social_Status.pdf
- Eldevik, S., Hastings, R. P., Hughes, J. C., 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.
- Esteves, J., Taheri, A., Perry, A., & Koudys, J. (2016, October). Examining diagnostic measures used for the diagnosis of autism spectrum disorder. Poster presentation at the Ontario Psychological Association annual convention. Toronto, ON.
- Flanagan, H.E., Perry, A., & Freeman, N.L. (2012). Effectiveness of large-scale, community-based intensive behavioural intervention: A waitlist comparison study exploring outcome and predictors. *Research in Autism Spectrum Disorders, 6*, 673-682.
- Freeman, N. L., & Perry, A. (2010). Outcomes of Intensive Behavioural Intervention in the Toronto Preschool Autism Services. *Journal on Developmental Disabilities, 16*, 17-32.
- 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.

- Kovshoff, H., Hastings, R.P., & Remington, B. (2011). Two-year outcomes for children with autism after cessation of early intensive behavioral intervention. *Behavior Modification, 35*, 427-450.
- Lovaas, O.I. (1987). Behavioral treatment and normal intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*, 3-9.
- Matson, J.L., & Smith, R.M. (2008). Current status of intensive behavioral interventions for young children with autism and PDD-NOS. *Research in Autism Spectrum Disorders, 2*, 60-74.
- 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*, 359-372.
- O'Connor, A. B., & Healy, O. (2010). Long-term post-intensive behavioral intervention outcomes for five children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 4*, 594-604. doi: 10.1016/j.rasd.2009.12.002
- Perry, A., Condillac, R. A., Freeman, N. L., Dunn Geier, J., & Belair, J. (2005). Multisite 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., Hughes, S., LaRose, L., ...& Williams, J. (2008). Effectiveness of Intensive Behavioral Intervention in a large, community-based program. *Research in Autism Spectrum Disorders, 2*, 621-642. doi:10.1016/j.rasd.2008.01.002
- Perry, A., Cummings, A., Dunn Geier, J., Freeman, N.L., Hughes, S., Managan, 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.

- Perry, A., Flanagan, H.E., Dunn Geier, J., & Freeman, N.L. (2009). 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.
- Prichard, E.A. (2011). *Short-term follow-up of children with autism who have received Intensive Behavioural Intervention*. Unpublished doctoral dissertation, York University, Toronto, ON, Canada.
- Remington, B., Hastings, R., Kovshoff, H., degli Espinosa, F., Jahr, E., Brown, T.,... & Ward, N. (2007). Early intensive behavioral intervention: Outcomes for children with autism and their parents after two years. *American Journal on Mental Retardation*, 112, 418-438.
- Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *The Childhood Autism Rating Scale (CARS)*. Los Angeles, CA: Western Psychological Services.
- Sparrow, S.S., Cicchetti, V.D., & Balla, A.D. (2005). *Vineland Adaptive Behavior Scales (2nd ed.)*. Circle Pines, MN: American Guidance Service.
- Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test 4 professional manual*. Lutz, FL: Psychological Assessment Resources.

Acknowledgements

Financial support for this study was provided by: Brock University (CRISS grant); York University Faculty of Health; and the Social Sciences and Humanities Research Council of Canada (SSHRC) Minor Research Grant, York University. In addition, we are most grateful for use of space in the York University Psychology Clinic.

We would like to thank the following individuals who assisted with the assessments: Ksusha Blacklock, Meisha Charles, Stella Dentakos, Jeffrey Esteves, Hilda Ho, Nidhi Luthra, Jennifer MacMullin, Nancy Marchese, Busi Ncube, Kimberley Saldanha, Rebecca Shine, Azin Taheri, Krysten Thompson, Ami Tint, Odette Weiss, and Katherine Wincentak.

Finally, we would like to express our sincere thanks to the parents, youth, and teachers who participated in this study.

Table 1

Sample Characteristics (n=21)

	<i>M(SD) or n (%)</i>	<i>Range or n (%)</i>
Child Characteristics (n=21)		
Sex		male: 19 (90.5%)
Age at Time 1 (start EIBI)	3.41 (1.05)	1.83 - 5.58
Age at Time 2 (end EIBI)	5.77 (0.97)	3.17 - 7.58
Age at Time 3 (first follow-up)	8.81 (1.77)	5.69 - 12.58
Age at Time 4 (current follow-up)	16.02 (1.79)	13.25 -20.17
Interval T1--> T2 (duration EIBI)	2.23 (0.82)	1.08 - 3.42
Interval T2--> T3	3.01 (1.61)	1.08 - 6.42
Interval T3--> T4	7.21 (0.35)	6.59 - 7.92
Total time since end of EIBI	10.22 (1.64)	8.50 - 14.01
Family Demographics at T4 (n=18)		
Barratt SES	39.43 (13.64)	16-63
Family composition	<i>n (%) Adults</i>	<i>n (%) Children</i>
1 (single-parent)	3 (16.7)	5 (27.8)
2	12 (66.7)	11 (61.1)
3	3 (16.7)	2 (11.1)
Born in Canada; First language English (n=18 mothers; 15 fathers)	15 (83.3)	11 (73.3)
Employment (n=16 mothers; 13 fathers)	Not currently 5 (31.2)	Not currently 1 (7.7)
	Part-time 3 (18.8)	Part-time 0
	Full-time 8 (50.0)	Full-time 12 (92.3)

Table 2

Services received since end of EIBI (n=21)

Service or Treatment	n (%)
School Program	
Special education throughout	9 (42.9%)
Mixed	6 (29.6%)
Regular class throughout	6 (28.6%)
Other Services/Treatments	
Occupational therapy	16 (76.2%)
Children's mental health services (e.g., social skills group)	15 (71.4%)
Parent behavioural training program	15 (71.4%)
Family support services (e.g., respite care)	15 (71.4%)
Speech-Language Pathology services	12 (57.1%)
Behaviour management for problem behaviour	7 (33.3%)
Medication (e.g., risperidone)	7 (33.3%)
Other special diet/supplement	6 (28.6%)
Other treatment/service (e.g., service dog, horseback therapy, neuropsych training)	6 (28.6%)
Augmentative & Alternative Communication (AAC) services	5 (23.8%)
Gluten/casein free diet	3 (14.3%)
Applied Behaviour Analysis	2 (9.5%)
Other direct services (e.g., floortime)	2 (9.5%)
Medication for seizures	2 (9.5%)

Table 3

Scores on all Outcome Measures at Time 4 approximately 10 years after EIBI

	<i>n</i>	<i>M (SD)</i>	Range
Cognitive (WISC-IV or other) Standard scores			
VIQ (if WISC-IV)	17	79.88 (22.40)	43 - 108
PIQ (if WISC-IV)	17	87.06 (23.37)	47 - 118
FSIQ (all tests)	21	65.52 (28.87)	<20 - 108
Adaptive (Vineland-II) Standard scores			
Communication	21	68.76 (18.25)	30 - 100
Daily Living Skills	21	65.33 (19.60)	25 - 99
Socialization	21	68.38 (18.32)	37 - 97
Adaptive Behaviour Composite	21	65.43 (17.97)	27 - 96
Academic (WRAT-4) Standard scores			
Word Reading	19	86.79 (16.91)	55 - 116
Spelling	19	87.11 (18.10)	55 - 120
Mathematics	19	78.68 (14.97)	55 - 107
Autism Severity (CARS)			
CARS Total score	21	26.21 (6.36)	18 - 43
Social-emotional (CBCL/TRF) T scores			
Parent Internalizing	19	56.89 (9.72)	44 - 79
Parent Externalizing	19	52.16 (9.95)	34 - 71
Parent Total Problems	19	57.21 (9.84)	42 - 73
Teacher Internalizing	16	61.94 (9.50)	46 - 87
Teacher Externalizing	16	59.00 (8.82)	42 - 74
Teacher Total Problems	16	61.38 (8.54)	50 - 82

Note. WISC-IV=Wechsler Intelligence Scale for Children (4th ed.); WRAT-4=Wide Range Achievement Test (4th ed.); CARS=Childhood Autism Rating Scale (higher scores indicate more severe autism); CBCL=Child Behavior Checklist; TRF=Teaching Rating Form

Table 4

Scores in three Outcome Subgroups at Time 4

Outcome Category	FSIQ	ABC	CARS
Good (<i>n</i> =7)	94.86 (6.41)	79.57 (9.57)	22.14 (2.38)
Medium (<i>n</i> =10)	63.50 (13.77)	66.40 (12.19)	25.40 (4.82)
Poor (<i>n</i> =4)	19.25 (5.56)	22.14 (2.38)	35.38 (5.99)
	$F(2, 18)=64.35, p<.001$	$F(2, 18)=18.77, p<.001$	$F(2, 18)=11.77, p=.001$
	Post hoc G>M>P	Post hoc G>M>P	Post hoc G=M<P

Table 5

Group Comparisons across Four Time Points

	<i>M(SD)</i>	Overall ANOVA	Pairwise comparisons
IQ (<i>n</i> =14)		$\lambda=.49$, $F(2,12)=6.17$, $p=.014$, partial $\eta^2=.51$	Pairwise comparisons: T1 vs T2 n/a * T2>T3 ($p=.08$) T3>T4 ($p=.052$)
Time 2	81.43 (29.90)		
Time 3	76.14 (28.60)		
Time 4	71.71 (28.17)		
MA (years) (<i>n</i> =14)		$\lambda=.15$, $F(2,12)=53.07$, $p<.001$, partial $\eta^2=.80$, after Greenhouse-Geisser adjustment	Pairwise comparisons: T1 vs T2 n/a * T2<T3 ($p=.002$) T3<T4 ($p<.001$)
Time 2	4.56 (1.70)		
Time 3	6.57 (2.72)		
Time 4	11.25 (4.35)		
ABC SS (<i>n</i> =20)		$\lambda=.64$, $F(3,17)=5.65$, $p=.005$, partial $\eta^2=.23$, after Greenhouse-Geisser adjustment	Pairwise comparisons: T1<T2 ($p=.018$) T2=T3 ($p=.59$) T3>T4 ($p<.001$)
Time 1	63.45 (8.95)		
Time 2	73.05 (20.47)		
Time 3	74.60 (15.09)		
Time 4	66.85 (17.18)		
ABC Age Eq (years) (<i>n</i> =18)		$\lambda=.24$, $F(3,15)=40.17$, $p<.001$, partial $\eta^2=.70$, after Greenhouse-Geisser adjustment	Pairwise comparisons: T1<T2 ($p<.001$) T2<T3 ($p=.004$) T3<T4 ($p<.001$)
Time 1	1.54 (0.66)		
Time 2	3.48 (1.37)		
Time 3	4.73 (2.05)		
Time 4	8.22 (4.09)		
CARS (<i>n</i> =19)		$\lambda=.36$, $F(3,16)=16.75$, $p<.001$, partial $\eta^2=.48$	Pairwise comparisons: T1>T2 ($p<.001$) T2=T3 ($p=.31$) T3>T4 ($p=.03$)
Time 1	34.16 (5.49)		
Time 2	29.58 (5.23)		
Time 3	28.74 (5.41)		
Time 4	26.63 (6.40)		

Note. MA=Mental Age; ABC SS=Vineland Adaptive Behavior Composite Standard Score; ABC Age Eq=age equivalent score (mean of 3 domains); CARS= Childhood Autism Rating Scale (higher scores indicate more severe autism) * unable to calculate because of missing data at Time 1

Table 6

Comparison of Outcomes at Time 3 and Time 4 (n=21)

	Time 3	Time 4	Comparison and ES		
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t(df)</i>	<i>p</i>	<i>d</i>
IQ	71.25 (30.29)	65.52 (28.87)	<i>t (20)=2.63, p=.02, d= -.19</i>		
MA (years)	6.03 (2.56)	10.26 (4.31)	<i>t (20)= -8.91, p<.001, d= 1.19</i>		
ABC SS	73.29 (15.89)	65.43 (17.97)	<i>t (20)=4.75, p<.001, d= -.46</i>		
ABC Age Eq (years)	4.58 (2.03)	7.90 (3.98)	<i>t (20)= -6.46, p<.001, d= 1.05</i>		
CARS	28.69 (5.17)	26.21 (6.36)	<i>t (20)=2.66, p=.02, d= -.43</i>		

Note. MA=Mental Age; ABC SS=Vineland Adaptive Behavior Composite Standard Score; ABC Age Eq=age equivalent score (mean of 3 domains); CARS= Childhood Autism Rating Scale (higher scores indicate more severe autism)

Table 7

Number of participants classified into Three Outcome Categories at each Time Point

	Time 1	Time 2	Time 3	Time 4
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Good	0	8 (38.1)	8 (38.1)	7 (33.3)
Medium	15 (71.4)	8 (38.1)	9 (42.9)	10 (47.6)
Poor	6 (28.6)	2 (9.5)	4 (19.0)	4 (19.0)
N/A	0	3 (14.3)	0	0

Table 8

Stability/Change since the end of EIBI

	Time 2--> Time 3			Time 3--> Time 4		
	% stable	<i>n</i> Clinically significantly better	<i>n</i> Clinically significantly worse	% stable	<i>n</i> Clinically significantly better	<i>n</i> Clinically significantly worse
IQ	93	0	1	90	0	2
ABC	75	3	2	81	0	4
CARS	79	3	1	71	5	1
Outcome Category	88	1	3	76	2	3

Note. ABC=Vineland Adaptive Behavior Composite Standard Score; CARS= Childhood Autism Rating Scale

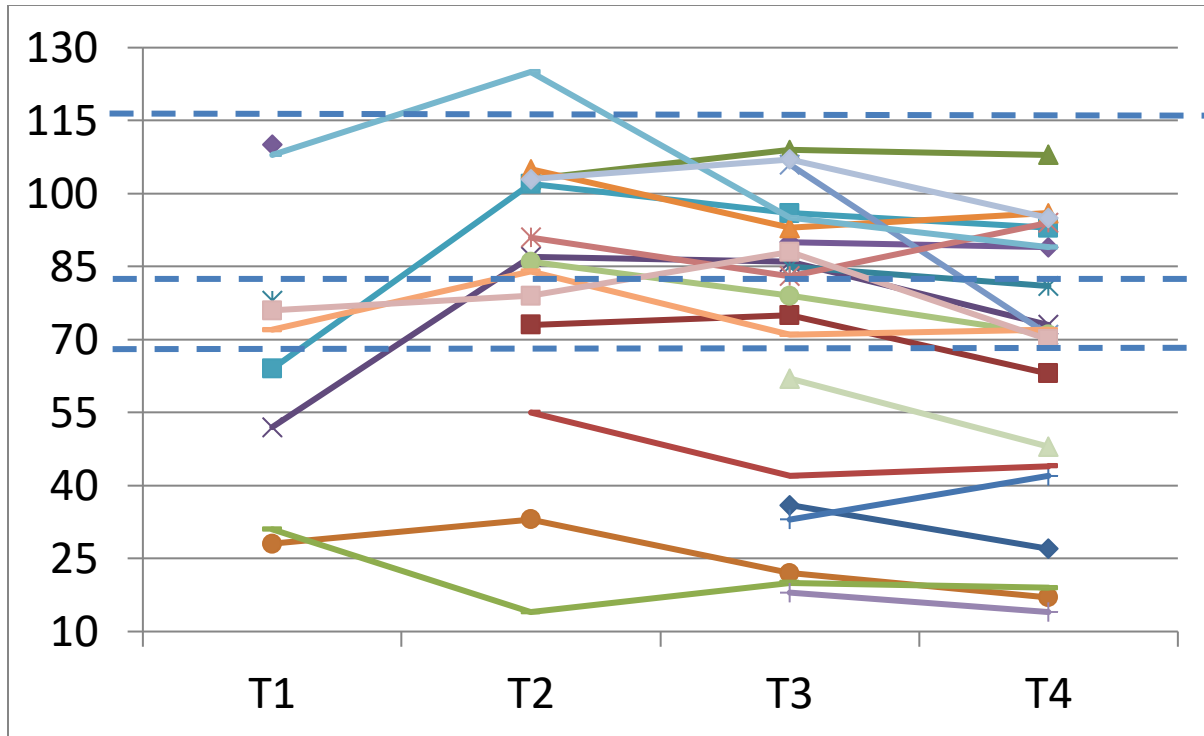


Figure 1. Individual FSIQ Scores at Four Time Points

Note. Average range=85-115; 70-85=borderline; <70=intellectual disability range. Clinically significant change=15 points. Note that time intervals are not equal as they appear.

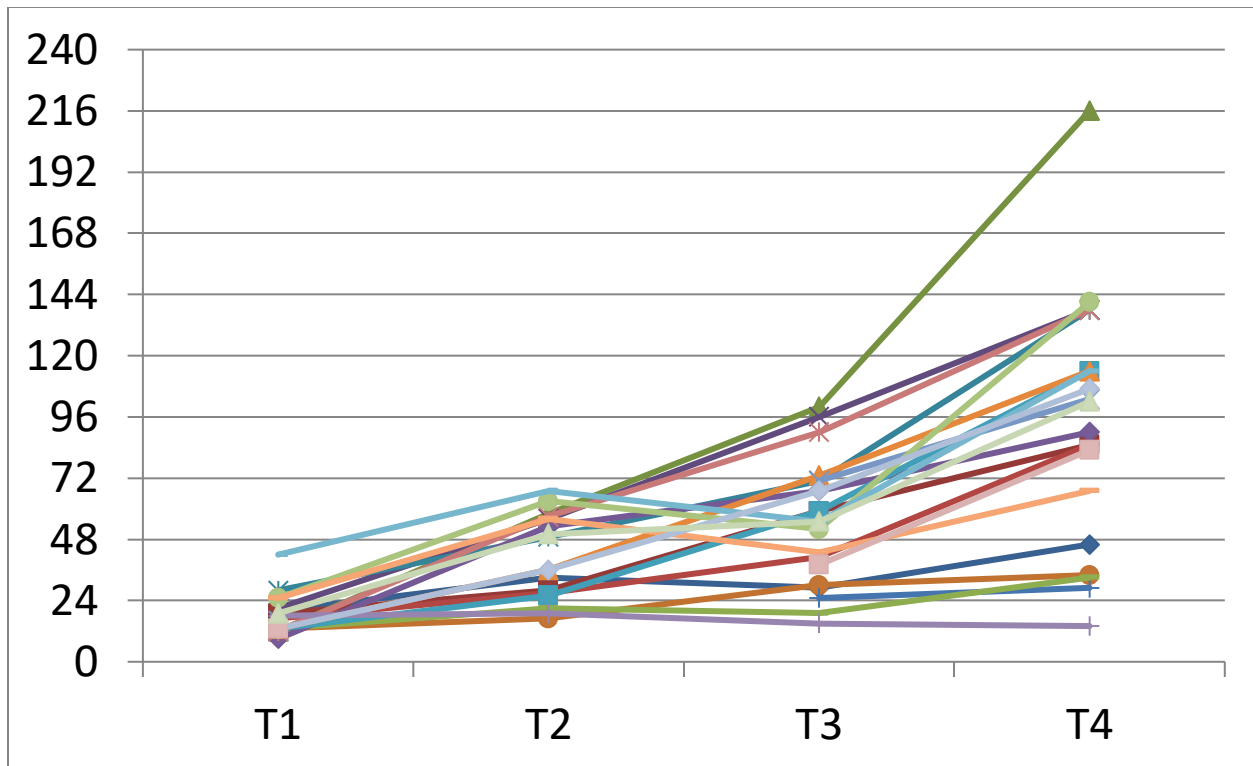


Figure 2. *Individual Vineland Age Equivalent Scores (months) at Four Time Points.*

Note. Scores are M of Communication, Daily Living, and Socialization age equivalents (in months). Note that time intervals are not equal as they appear.

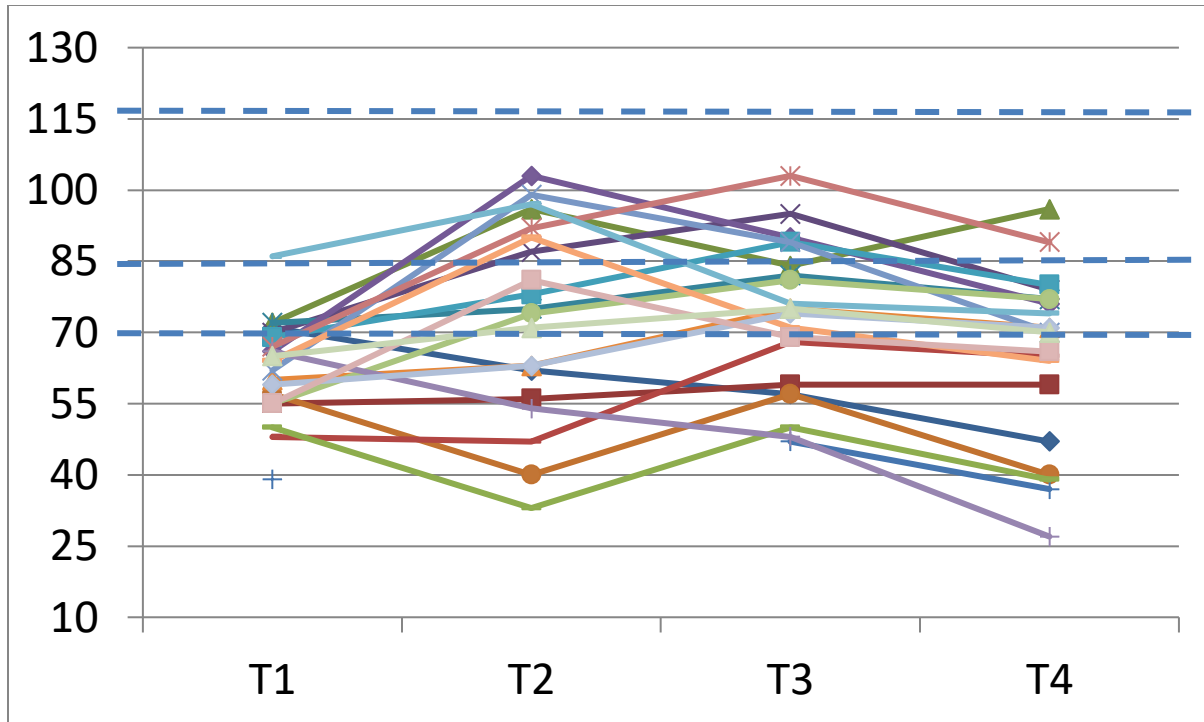


Figure 3. Individual Vineland ABC Standard Scores at Four Time Points

Note. Average range=85-115; 70-85=borderline; <70=intellectual disability range. Clinically significant change=15 points. Note that time intervals are not equal as they appear.

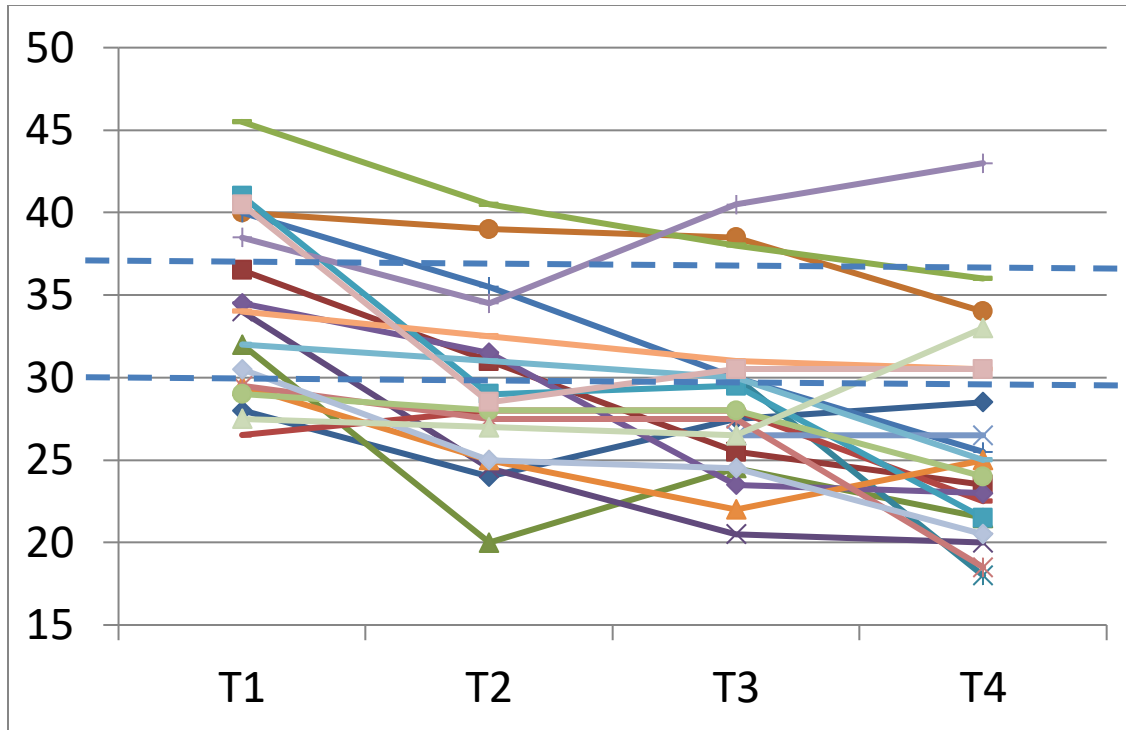


Figure 4. Individual CARS Autism Severity scores at Four Time Points

Note. <30=non-autism range; 30-36=mild/moderate autism; >37=severe autism. Clinically significant change=5 points. Note that time intervals are not equal as they appear. Dashed lines represent cut-scores of 30 for Mild/moderate Autism and 37 for Severe Autism range.