

Early Intensive Behavioral Intervention: Outcomes for Children With Autism and Their Parents After Two Years

Bob Remington

University of Southampton, UK

Richard P. Hastings

University of Wales, Bangor, UK

Hanna Kovshoff and Francesca degli Espinosa

University of Southampton, UK

Erik Jahr

Akershus University Hospital, Norway

Tony Brown, Paula Alsford, Monika Lemaic, and Nicholas Ward

University of Southampton, UK

Abstract

An intervention group ($n = 23$) of preschool children with autism was identified on the basis of parent preference for early intensive behavioral intervention and a comparison group ($n = 21$) identified as receiving treatment as usual. Prospective assessment was undertaken before treatment, after 1 year of treatment, and again after 2 years. Groups did not differ on assessments at baseline but after 2 years, robust differences favoring intensive behavioral intervention were observed on measures of intelligence, language, daily living skills, positive social behavior, and a statistical measure of best outcome for individual children. Measures of parental well-being, obtained at the same three time points, produced no evidence that behavioral intervention created increased problems for either mothers or fathers of children receiving it.

An increasing body of empirical research suggests that early, intensive, structured intervention, based on the principles of applied behavior analysis, is effective in remediating the intellectual, linguistic, and adaptive deficits associated with autism. Lovaas's (1987) original archival study showed that a group of children receiving 40 weekly hours of home-based early intensive behavioral intervention achieved significant gains in IQ and social functioning in comparison with control groups receiving either a less intensive intervention or the standard treatment offered by educational services. McEachin, Smith, and Lo-

vaas's (1993) follow-up study showed that the gains were maintained at age 11.5 years and that 8 of 9 children, previously identified as having achieved "best outcome" status could not be distinguished from typically developing peers by assessors blind to their treatment.

Since 1987, many researchers have conducted evaluation studies attesting to the effectiveness of early intervention with autism, but most have suffered from methodological limitations that threatened their internal validity. For example, in common with Lovaas (1987), several subsequent studies were not truly prospective randomized control

trials because the researchers were unable to assign children to groups randomly (e.g., Anderson, Avery, DiPietro, Edwards, & Christian, 1987; Birnbrauer & Leach, 1993; Eikeseth, Smith, Jahr, & Eldevik, 2002) or used archival data to form a comparison group (Sheinkopf & Siegel, 1998). Others still relied on simple pre–post group comparisons (e.g., Stahmer & Ingersoll, 2004; Weiss, 1999) or controlled single-case studies (e.g., Green, Brennan, & Fein, 2002).

In summary, there are few randomized control trials that meet adequate internal validity criteria and demonstrate the efficacy of early intensive behavioral intervention. Two exceptional studies (Sallows & Graupner, 2005; Smith, Groen, & Wynn, 2000) compared the effects of early intervention implemented using either a clinic- or a parent-directed model. Smith et al. (2000) showed that clinic-based intervention lasting 25 hours per week for 2 to 3 years had greater impact than a less intensive parent training-based intervention (5 hours per week). Group measures of children's intelligence, visual–spatial skills, and language did not differ at age 3 years, but changes in favor of the clinic-directed group were apparent at age 7 to 8 years. In contrast, Sallows and Graupner (2005) found no differences between clinic- and parent-directed programs on similar measures after 4 years of treatment. In this study, however, between-group differences in the intensity of intervention were much less marked.

The paucity of randomized control trials in this area reflects the considerable difficulties of staging them: Unlike drug trials, where patients are, in principle, blind to the intervention, parents are made well-aware in advance of the treatment their children will receive. Moreover, as knowledge accumulates and early intervention is accepted as a treatment of choice for autism (e.g., Surgeon-General, 1999), researchers face ethical difficulties with random assignment, and families become less willing to commit their children to long-lasting treatments of dubious utility. Thus, although a randomized controlled trial approach can, under idealized conditions, produce the strongest evidence establishing the efficacy of an intervention (see, e.g., Whitehurst, 2003), it may be difficult to conduct further evaluative trials of early intensive behavioral intervention unless well-matched, equally credible alternatives can be pitted against standard procedures.

In any case, it is likely that the effectiveness in practice of early intensive behavioral interven-

tion would be overestimated by any putative randomized trial. In general, the external validity of such trials is compromised by tight control of variables, including co-morbidity, treatment fidelity, treatment adherence, and self-selection into and out of trials (Kendall, Chu, Gifford, Hayes, & Nauta, 1998; Persons & Silberschatz, 1998; Seligman, 1995). Absence of control of such factors is commonplace in typical service settings so the long-term clinical benefit of any intervention depends on its remaining effective in conditions that are less than optimal. Considerations of this kind have given rise to field effectiveness research, in which random assignment to groups and the most rigorous experimental control are traded against a more naturalistic evaluation of service delivery in context. Two recent evaluations of early behavioral intervention for autism (H. Cohen, Amerine-Dickens, & Smith, 2006; Howard, Sparkman, Cohen, Green, & Stanislaw, 2005) have adopted this approach.

Using the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)* (American Psychiatric Association, 1994) criteria rather than the “gold standard” research tool, namely, the Autism Diagnostic Interview-Revised (Lord, Rutter, & Le Couteur, 1994), Howard et al. (2005) identified 61 children who met criterion either for autistic disorder or for pervasive developmental disorder—not otherwise specified (PDD-NOS). They compared 29 children who received intensive clinic-directed behavior analytic intervention (25 to 40 hours per week) with two comparison groups, one ($n = 16$) that received equally intensive eclectic intervention and the other ($n = 16$) whose members were not enrolled in any intensive public intervention programs. Assignment to groups was not randomized but depended on the advice of practitioners, with “parental preferences weighted heavily” (Lord et al., 1994, p. 364). Unusually, Howard et al. eschewed direct group comparison using ANOVA models, opting instead for a multiple regression-based analysis, with group membership treated as a categorical variable. This showed that prior to treatment there were no differences between the behavior analytic intervention group and the two comparison groups combined.

In a second analysis of functioning 14 months later, Howard et al. (2005) found that children in the intensive behavior analytic intervention group had higher scores than those in the combined comparison groups on standardized tests of cognitive, linguistic, and adaptive func-

tioning. Although the effects implied by these analyses were confirmed in a similar test of the absolute change scores on all measures, no analysis taking into account conditional change (i.e., relative to baseline scores) was presented.

In a 3-year prospective outcome study carried out in a community setting, Cohen et al. (2006) compared 21 children receiving early intensive behavioral treatment with an equal number of children participating in public school special education classes. Random assignment to groups was not attempted; instead, assignment was based on parental preference and a file review process used to identify an IQ- and CA-matched child for each child receiving intensive intervention. In this way, it was possible to form a group of children “who met participation criteria . . . and whose parents chose other services” (p. S147). Both groups included some children with a diagnosis of autism and others with a PDD-NOS diagnosis, but the proportion of the latter was lower in the intervention group. Analysis of covariance (ANCOVA), using baseline scores as the covariates, and comparing performance after 1, 2, and 3 years revealed that the intensive group was superior on measures of IQ and adaptive behavior, but not on measures of language or nonverbal skills. Moreover, the absence of a Group \times Time interaction indicated that between-group performance differences achieved after 12 months did not increase throughout the treatment. The number of children scoring in the normal range on the primary outcome measure (IQ) was higher in the intensive intervention group after 3 years, but this difference was not statistically significant.

Results of the Howard et al. (2005) and H. Cohen et al. (2006) studies suggest that early intensive behavioral intervention can be effective when delivered in more typical community settings and when compared with treatment as usual—the typical mix of interventions available to children with autism. However, in common with almost all research in this area, these researchers did not consider two crucial questions that we sought to address in the present research. First, does early intensive behavioral intervention have an impact beyond the cognitive, language, and adaptive behavior deficits associated with autism, additionally affecting the characteristic diagnostic symptoms of the disorder? In the present study, we included rating scale measures of autistic presentation, behavior problems, and prosocial behavior, as well as an observational measure of joint attention

(Mundy & Crowson, 1997). The second issue we addressed concerns the impact of intensive intervention on family members. This has been explored only minimally, and although existing data suggest that the mothers and siblings of participating children are not adversely affected (Birnbauer & Leach, 1993; Hastings, 2003a; Hastings & Johnson, 2001; Smith, Buch, & Gamby, 2000; Smith, Groen, & Wynn, 2000), there is as yet no published controlled study of a range of measures of both maternal and paternal well-being.

We also explored a key methodological issue relating to intervention effectiveness by adopting a more precise approach to identifying “best outcome” children based on Jacobson and Truax’s (1991) objective criteria for establishing whether a particular child has benefited meaningfully from an intervention. These criteria are (a) *reliable change* (the extent to which statistical factors can be ruled out as an explanation for apparent change) and (b) *clinically significant change* (the extent to which change is also clinically meaningful). Although in earlier research investigators have used a criterion of IQs moving to within the normal range (Birnbauer & Leach, 1993; Eikeseth et al., 2002; Lovaas, 1987; McEachin et al., 1993; Sallows & Graupner, 2005; Smith et al., 2000), to the best of our knowledge this is the first study simultaneously to apply statistical criteria for both reliable and clinical change to the outcomes for early intensive behavioral intervention programs.

We explored these three key issues within the United Kingdom educational system, where in previously published research, based on an uncontrolled survey of the impact of home programs, Bibby, Eikeseth, Martin, Mudford, and Reeves (2001) reported only minimal outcomes and wide variations in the quality and intensity of service delivery. In contrast, we sought to construct the most rigorously controlled field effectiveness study achievable within the constraints of the prevailing culture. This involved a prospective 2-year longitudinal design, comparing children with autism whose families had chosen intensive behavioral intervention from a range of different service providers in England with children whose parents were not seeking this type of intervention and were receiving typical statutory services (treatment as usual).

In summary, we designed this study as a rigorous test of whether early intensive behavioral intervention for children with autism can be beneficial in routine use, incorporating a wide range

of outcome measures for both children with autism and their parents. We used objective criteria to identify children achieving “best outcome.” Following previous effectiveness studies, we expected intervention to lead to improvements in children’s cognitive, language, and social functioning when compared with treatment as usual. Existing family research suggests that parents’ psychological well-being would not be adversely affected by engagement with intensive intervention, although it was unclear whether positive outcomes could be expected. Given the lack of published data, we had no expectations as to whether there would be positive changes in ratings of autistic symptoms, behavior problems, or measures of joint attention behaviors following early intensive behavioral intervention.

Method

Design Overview

Two groups of preschool children with a formal diagnosis of autism were identified. Parents of children in the intervention group had opted for early intensive behavioral intervention, either provided from public funds or purchased privately; parents of children in the comparison group were not actively seeking behavioral intervention, and instead were receiving publicly funded standard provision offered by their Local Education Authority (i.e., treatment as usual). Assessments of the children’s cognitive functioning, adaptive behavior, autistic behaviors, and social and communicative skills were undertaken at three data-collection points: prior to intervention (baseline); after 1 year, and again after 2 years of intervention (12- and 24-month assessments). Measures of parental mental health, stress, and positive perceptions of their child were obtained at the same time points.

Participants

Children with autism. Children were recruited through referrals from local education authorities, through advertisements placed with the United Kingdom National Autistic Society, its regional branches, and through parent groups or charities. Demographic data relating to families appear in Table 1 and to children, in Table 2 (for baseline information, see *Results*). To meet the inclusion requirements for this study, all children in both the intervention and comparison groups had to

meet the following criteria. First, we required a diagnosis of autism based on the Autism Diagnostic Interview-Revised carried out by an assessor (the last author), who was fully trained to administer and score this instrument for research purposes. All children had also either previously been diagnosed with autism by a clinician independent of the research program or had a suspected diagnosis of autism. Second, children were required to be between 30 and 42 months of age at time of their induction. Third, they were required to be free of any other chronic or serious medical condition that might interfere with the ability to deliver consistent intervention or might otherwise adversely affect development. Finally, all the children lived in the family home.

We identified 44 children who met these criteria. The families of 23 of them, constituting the intervention group, had opted for early intensive behavioral intervention, either receiving provision from the University of Southampton and funded through their local education service ($n = 13$) or through a private service provider ($n = 10$). In the latter cases, services were either paid for by the parents themselves or by their local education service. The remaining 21 families, the comparison group, were receiving various forms of publicly funded educational provision for their children. The groups differed slightly on chronological age (CA), with the comparison group children ($M = 38.4$ months, $SD = 4.4$) being on average approximately 3 months older than the children in the intervention group ($M = 35.7$ months, $SD = 4.0$), $t(42) = 2.14$, $p < .05$. None of the other demographic variables assessed for the children differed between the two groups at baseline assessment. Chronological age was explored as a control variable in the main statistical analyses.

Parents. Forty-four mothers and 31 fathers of children in the intervention and comparison groups provided data on some aspects of the child’s functioning and on their own well-being. Their demographic details are shown in Table 1. In the sample as a whole, there were 40 couples at the baseline assessment. Nine families had a father at home who declined to participate throughout the research. For 4 families, the father was not living in the same home as the mother and the child with autism at baseline; these fathers did not participate throughout the research. The two groups were very similar on the majority of parent/family demographic characteristics. Although some demographic differences appear to

Table 1. Demographic Characteristics of Families by Group

Demographic	Intervention (<i>n</i> = 23)			Comparison (<i>n</i> = 21)		
	<i>n</i> / Mean	%/ <i>SD</i>	Range	<i>n</i> / Mean	%/ <i>SD</i>	Range
Families						
Marital status						
Married	16	69.6		16	76.2	
Living with partner	4	17.4		4	19	
Divorced/Separated/Single and not living with partner	3	13		1	4.8	
Siblings						
0	5	21.7		3	14.3	
1	11	47.8		13	61.9	
2	7	30.4		4	19	
3	0	0		1	4.8	
Siblings with developmental disabilities						
0	20	87		13	61.9	
1	2	13		7	33.3	
2	0	0		1	4.8	
All mothers (<i>n</i> = 44)						
Mean age	35.7	4.0	26–42	33.6	3.8	26–41
Level of education						
No university education	13	57		17	81	
University education	10	43		4	19	
Paid work						
Full-time	0	0		0	0	
Part-time	7	100		7	100	
All fathers living in the family home (<i>n</i> = 40)						
Mean age	38.8	5.5	31–50	37.1	4.8	30–53
Level of education						
No university education	10	50		11	55	
University education	10	50		9	45	
Paid work						
Full-time	19	95		18	90	
Part-time	0	0		1	5.6	
Fathers who responded to questionnaires (<i>n</i> = 31)						
Mean age	38.7	4.8	31–50	37.5	5.4	30–53
Level of education						
No university education	7	44		8	53	
University education	9	56		7	47	
Paid work						
Full-time	15	93.8		13	86.7	
Part-time	0	0		1	7.7	

Note. All mothers responded to the questionnaires but only 31 fathers responded similarly. Data for both subsamples appear in the table.

Table 2. Unadjusted Means (*SDs*) of Child Measures by Group and Assessment Point

Measure	Baseline		12-month assessment		24-month assessment	
	Intervention	Comparison	Intervention	Comparison	Intervention	Comparison
IQ**	61.43 (16.43)	62.33 (16.64)	68.78 (20.49)	58.90 (20.45)	73.48 (27.28)	60.14 (27.76)
MA***	22.04 (6.89)	23.71 (6.00)	33.70 (10.16)	29.81 (9.89)	44.39 (16.39)	38.00 (17.44)
VABS ^b						
Composite	114.78 (26.89)	113.57 (29.78)	169.70 (49.07)	145.76 (45.56)	202.83 (61.98)	182.86 (58.89)
Communication	23.52 (11.35)	21.62 (10.81)	42.83 (18.25)	34.62 (17.17)	54.74 (24.43)	46.00 (24.51)
Daily Living*	24.13 (7.49)	25.43 (10.56)	39.52 (14.71)	35.52 (14.34)	50.22 (16.46)	44.67 (16.99)
Socialization	29.57 (6.65)	28.29 (7.48)	38.52 (12.57)	33.14 (11.77)	43.52 (15.94)	41.48 (14.52)
Motor Skills*	37.57 (6.37)	38.24 (7.06)	48.83 (6.84)	44.48 (7.70)	54.35 (9.12)	50.71 (8.21)
Joint attention ^c						
Initiating	3.33 (4.40)	3.63 (4.92)	7.71 (7.52)	6.19 (8.79)	11.76 (9.41)	11.19 (13.86)
Responding*	5.29 (3.62)	5.94 (3.91)	8.95 (4.18)	7.13 (5.21)	11.29 (3.47)	10.06 (4.99)

^aMental age. ^bVineland Adaptive Behavior Scales Raw Scores. ^cMeasured using the Early Social Communication Scales. * $p < .05$. ** $p < .01$ on main effects for combined 12- and 24-month data. Intervention group $n = 23$ and comparison group $n = 21$, except for joint attention, intervention group $n = 21$; comparison group $n = 16$.

be present, no differences between the groups were large enough to reach statistical significance at the .05 level. Thus, none of these characteristics were considered as candidate control variables in the following analyses.

Child Measures

We used norm-referenced instruments to gather the cognitive, language, and behavioral outcome data for the children. The assessments were chosen for their good psychometric properties and use in published outcome studies with similar populations. An important consideration was their potential utility for testing children with autism. Many tests require language skills and sustained attention, two abilities that may also be affected in such children, whose symptomatic deficits in language, intellectual, neurological, adaptive behavior, and interpersonal skills could influence performance on standardized measures and thus impact on the reliability and validity of any test. All tests were administered according to the standard procedures to ensure our data were comparable with those from other studies. Although in some cases this could potentially have led to an underestimate of children's ability (e.g., children reaching a ceiling on the Bayley Scales may have continued to score on the nonverbal, non-social items had these been administered), scoring methods did not differentially favor either group.

The tests selected were administered by a master's level trained psychometrician (the third author), who had over 4 years of experience with children who have autism and who exercised every caution to obtain reliable and valid data. Although resources did not allow for formal independent reliability checks when assessments by independent psychometricians were available, these scores were always within a standard error of measurement of those reported below. Moreover, the third author was not informed of group status, worked independently of intervention teams, had no access to intervention reports, and her contact with the family was limited to the annual assessments.

Intellectual functioning. The Bayley Scales and the Stanford Binet Intelligence Scale: Fourth Edition (Thorndike, Hagen, & Sattler, 1986) were both chosen, in part, for their low floor. The Bayley, designed for children up to 42 months of age, is appropriate for children with intellectual disabilities or those whose language skills are not sufficiently advanced to take a full-scale intelligence test. If children received the Bayley scales at a CA that exceeded the norms of the test, a mental age (MA) was calculated based on their raw score using Table B.2 in the Bayley manual. A ratio IQ was then computed based on the $MA/CA \times 100$ formula. The Stanford-Binet provides normative data from the age of 2 years and, with only one

timed subtest, provides a good deal of flexibility when assessing children with autism.

Language. The Reynell Developmental Language Scales–Third Edition (Edwards et al., 1997) was chosen primarily because it is one of the few language assessments previously used in early intensive behavioral intervention outcome studies and provides separate measures of expressive language and comprehension. However, the updated United Kingdom normed version used provides normative data only from 21 months of age, significantly older than the norms in the 1985 version, which begin at 12 months.

Adaptive skills. The Vineland Adaptive Behavior Scale–Survey Form (Sparrow, Balla, & Cicchetti, 1984) was chosen based on its prolific use and the fact that it could be administered in a short version (the survey form). The Vineland assesses adaptive behavior across four domains: Socialization, Communication, Daily Living Skills, and Motor Skills. Unfortunately, improvements in the adaptive behavior of children with autism are not always reflected in Vineland standardized scores. This is in part because higher functioning children show uneven developmental profiles with interdomain scatter (Burack & Volkmar, 1992) and in part because low-functioning children may show little scatter, owing to basal effects (Carter et al., 1998). To avoid such problems in research (as opposed to diagnostic) applications with children who have autism, Carter et al. (1998) recommended that raw scores be used in preference to standardized scores.

Rating scales for child behavior. The Positive Social subscale of the Nisonger Child Behavior Rating Form (Tassé, Aman, Hammer, & Rojahn, 1996) and the parent report version of the Developmental Behavior Checklist (Einfeld & Tonge, 1995) were chosen to assess child behavior. The Nisonger is an informant behavior rating scale designed to assess children with intellectual disabilities. The Developmental Behavior Checklist is a behavior rating questionnaire yielding a Total Behavior Score, indexing the severity of behavior problems and offering a subset of items that function as a reliable and valid autism screening tool (the Developmental Behavior Checklist–Autism Screening Algorithm, Einfeld & Tonge, 2002). The Autism Screening Questionnaire (Berument, Rutter, Lord, Pickles, & Bailey, 1999) was also used. Derived from the Autism Diagnostic Interview algorithm (Lord et al., 1994) and completed by parents, this instrument provides a dimension-

al score for the symptoms of autism that was used in the analyses.

Observational measures of nonverbal social communication. The Early Social Communication Scales (Mundy, Hogan, & Dohering, 1996) is a videotaped semi-structured observational instrument in which the tester presents a standard set of toys in ways designed to elicit social and communicative behaviors in an ecologically valid context. The key variables obtained through administration of the scales were measures of initiating and responding to joint attention. *Initiating joint attention* refers to the frequency with which children use eye contact, pointing, and showing to share the experience of a toy or object during testing. *Responding to joint attention* refers to the number of times, over eight trials, in which a child correctly turned his or her eye gaze and aligned attention in the direction of the tester's distal point to a poster. Children with autism are less likely than typically developing children, or children with intellectual disabilities, to initiate or respond to joint attention (McEvoy, Rogers, & Pennington, 1993; Mundy & Crowson, 1997; Mundy, Sigman, Ungerer, & Sherman, 1986). Therefore, in the present study we assessed whether these social interaction behaviors would improve differentially for the intervention group as a result of participating in a program requiring many hours of one-to-one interaction with adults.

Interrater reliability was assessed using videotaped data from 25% of children (9) at each time point, scored by an independent rater blind to group status and trained to reliability level on Early Social Communication Scale training videotapes. Intraclass correlations between the paired ratings, used to assess consistency between raters' codes at all three assessment points, ranged from .95 to .99 for initiating joint attention and .96 to .97 for responding to joint attention.

Self-Report Measures of Parental Well-Being

The Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983), chosen as a measure of parents' mental health, includes two subscales, one assessing depression and the other, anxiety. Previous research with parents of children with autism has shown that the measure maintains good reliability (internal consistency) for both mothers and fathers of children with autism (Hastings, 2003b; Hastings & Brown, 2002). The Parent and Family Problems subscale of the Questionnaire on Resources and Stress–Friedrich short form (Fried-

rich, Greenberg, & Crnic, 1983) was chosen as a general measure of parental stress. This scale yields a total stress score after five items previously shown to constitute a robust measure of depression in parents of children with disabilities (Glidden & Floyd, 1997) have been removed from the scale. This modification ensured that there was no overlap between the measures of stress and of mental health. The resulting 15-item scale had strong internal consistency in the present sample (Kuder-Richardson coefficients were .87 for mothers and .83 for fathers at baseline). The Kansas Inventory of Parental Perceptions Positive Contributions subscale (Behr, Murphy, & Summers, 1992) was chosen as a measure of the degree to which parents hold positive perceptions of their child and the child's impact on the family (e.g., bringing the family closer together, helping other family members to become more understanding of other people, and being a source of happiness and fulfillment). In the present research, we used the total positive perceptions score. This score had a high level of internal consistency for both mothers, Cronbach's $\alpha = .95$, and fathers, $\alpha = .95$.

Procedure

Intervention group. All children in the intervention group received home-based early intensive behavioral intervention for 2 years. Trained tutors and parents delivered one-to-one teaching based on applied behavior analysis for 25.6 hrs per week on average ($SD = 4.8$, range = 18.4 to 34.0). Thirteen of the 21 programs were provided by the University of Southampton and were free at the point of use for the parents nominated by the local education authority that funded the University intervention team (which included the fourth, fifth, seventh, and eighth authors). The remaining programs were delivered by other United Kingdom service providers, either funded directly by the parents or purchased for the parents by their Local Education Authority. These included PEACH, a parent charity ($n = 4$), London Early Autism Program ($n = 1$), United Kingdom–Young Autism Programme ($n = 1$), and East Sussex Local Education Authority ($n = 1$). The remaining child spent 9 months with PEACH, 9 months with a private consultant, and the final 6 months at a school where applied behavior analysis was regularly employed (he was the only child to attend such a school).

Although interventions were delivered by a range of service providers, they had in common

the 10 features characterizing research-based interventions identified by Green et al. (2002, p. 70). Treatment began in the home during the children's 3rd or 4th year and continued for 2 years. It involved 20 to 30 hrs a week of structured teaching, based on the principles of applied behavior analysis. Thus, programs used discrete trial training methods (Lovaas, 1993) and incorporated generalization procedures to extend and maintain emerging behavioral repertoires. Elements of natural environment training (Sundberg & Partington, 1999) and verbal behavior (Partington & Sundberg, 1998) were also integrated into the majority of the interventions.

In some cases, recognized alternative and augmentative communication systems based on behavioral principles were incorporated into interventions to address absence of speech and provide children with an initial means of communication. At 12 months, 44% (10) of the children in the intervention group were using the Picture Exchange Communication System and 17% (4) continued to do so at 24 months. For sign language or Makaton Communication Systems, the figures were 44% (10) at 12 months and 35% (8) at 24 months, respectively.

Intervention programs covering all aspects of functioning (e.g., language, other cognitive, social, motoric) were individualized for each child, based on ongoing analysis of current strengths and needs, taking into consideration typical developmental trajectory and practicability. Programs were thus progressive: When simpler skills were acquired, more complex skills were established as behavioral objectives, and this process continued throughout the 2 years of intervention. Similarly, as children's skills increased, the process of facilitating access to appropriate school settings was initiated.

The program was delivered to each child by a team of 3 to 5 therapists trained in the use of behavior analytic procedures (e.g., shaping, chaining, prompting, fading, modeling, discrimination learning, task analysis, functional analysis) and supervised by more experienced staff members, including a supervisor who had substantial experience with early intensive behavioral intervention and, in the majority of cases, a consultant with still greater experience to PhD level and/or a track record of research publication in behavior analysis. Parents also delivered therapy, which was supervised in the same way.

Supervision of each tutor team was accom-

plished using a workshop model in which supervisors arranged extended team meetings at regular intervals. The frequency of team meetings depended on the service provider; for the 13 children receiving University of Southampton supervised intervention, meetings were twice a month, with additional regular training overlaps; for the remaining children, meetings were less frequent (range = 4 to 12 weeks). During meetings, the child's progress since the previous meeting was assessed, programs were added or modified, and members of the team (including the parents) practiced the programs to be implemented next. Consultants attended meetings on a less frequent basis (on average, once every 2 months), but they were available by telephone or email to provide additional clinical supervision. Between meetings, supervisors were similarly available to the team and parents.

No child in the intervention group was attending school at the baseline assessment, but by the 12-month assessment, 13 (57%) attended a mainstream school for an average of 5.8 hrs per week. At the 24-month assessment, 17 children (74%) attended mainstream school for an average of 13.28 hrs per week; and 22% (5), a special needs school for an average of 9.15 hrs per week. The remaining child continued with only the home-based program. Because most children in the intervention group were simultaneously attending school and receiving home programs, school hours were somewhat lower than those for the comparison group children at the first and second year of the study. Treatment and Education of Autistic and Related Communication Handicapped Children—TEACCH principles (Schopler, Mesibov, & Baker, 1982) were sometimes incorporated into school provision of 2 children (9%) receiving this intervention at the 12-month assessment and 13% (3) at the 24-month assessment).

Apart from behavioral treatment and schooling, some children in the intervention group also received other interventions: 65% (15) were receiving speech therapy at the baseline assessment; 22% (5), at the 12-month assessment; and 26% (6), after 24 months. Dietary interventions (typically gluten and casein restriction) were also commonly reported, with 11 children (48%) on restricted diets at baseline, and 14 (61%) and 12 (52%) at the 12 and 24 months, respectively. Finally, parents also reported the use of routine prescription medication: 4% (1) at baseline, 17% (4)

at the 12-month assessment, and 4% (1) at 24 months. Vitamin injections or high doses of vitamins were given to 6 children (26%); 10, 44%; and 7, 30%, respectively, at baseline, 12-, and 24-month assessment; and homeopathic interventions, 5 children (22%) at baseline; 2 (9%) at 12 months; and 1, 4% at 24 months.

Comparison group. The children in the comparison group received their local education authorities' standard provision for young children with autism. Thus, over the course of 2 years, they experienced a variety of interventions designed to ameliorate the impact of autism and enhance functioning, none of which were intensive or delivered on a one-to-one basis for the majority of time. The most frequently reported intervention was speech therapy: 12 of the children (57%) received it at the time of the baseline assessment, 67% (14) at the 12-month assessment, and 48% (10) at the 24-month assessment. As part of the children's experience of school, parents reported frequent use of TEACCH principles (38%, 8 children, and 52%, 11 children at 12 months and 24 months, respectively). Similarly, the Picture Exchange Communication System was frequently employed: (67%, 14 children and 76%, 16 children, respectively, at 12 and 24 months) and sign language or Makaton communication systems (24%, $n = 5$ and 48%, $n = 10$, at 12 and 24 months) were used as alternative communication systems. Dietary interventions were also relatively common, with 14% ($n = 3$) on special diets at baseline, 19% ($n = 4$) at their 12-month assessment, and 29% ($n = 6$ children) at the 24-month assessment. Prescription medication, vitamin, and homeopathic use were also reported: 5% (1 child) received prescription medication at baseline, 24% (5) at 12 months, and 19% (4) at 24 months. Vitamin injections or high doses of vitamins were not used with any of the children at baseline, and only 1 child (5%) at the 12- and 24-month assessments. Finally, homeopathic interventions were reported for 24% (5) of the sample at baseline, and for only 1 child (5%) at the 12- and 24-month assessments.

No child in the comparison group was attending school at baseline assessment. By the time of their 12- and 24-month assessments, however, in line with their education authorities' standard provision, all had a school placement. At the 12-month assessment, 48% (6) were in a mainstream environment; 43% (9), in a special educational needs school, and 10% (2), a mixed placement in

which half their time was spent in each kind of school. The average number of hours per week spent at school was similar for each child no matter where they were placed (an average of 15.3 hrs spent in mainstream, 17 hrs spent in special needs, and 15 hrs spent in mixed placements). By their 24-month assessment, 48% (10 children) were in mainstream schools for a weekly average of 22.3 hrs and 52% were in special needs schools for 13.6 hrs per week.

Although intervention and comparison group children received similar levels of speech and language interventions at baseline, it is clear that this pattern was not sustained throughout the 24-month period. Typically, as reported below, this was because the intervention produced effects that reduced the need for other interventions such as sign language or Makaton.

Psychometric assessments. Outcome measures for children and parents were obtained at baseline, after 1 year of behavioral intervention or standard provision (12-month assessment), and after 2 years (24-month assessment). Performance-based tests were administered in a distraction-free environment at the family home. All questionnaires were mailed out to parents at the time of each of the three assessments and returned to research staff shortly afterwards. Telephone interviews using the Vineland were conducted with primary caregivers approximately 1 week prior to the children's assessment visits, which took place at the family home. These lasted approximately 60 min. Except for the Autism Diagnostic Interview, which the final author administered to parents in the home at the time of the baseline assessment, the third author administered all the standardized outcome measures using a uniform order of administration: (a) the Early Social Communication Scales, (b) the Bayley Scales of Infant Development or the Stanford Binet, and (c) The Reynell Developmental Language Scales (which was administered only if a child's language level was such that they could access the items on the test).

Results

Overview of Analysis of Group Data

To evaluate the effectiveness of behavioral intervention, we used ANCOVA models. Because the groups were not actively matched at baseline, baseline scores on outcome measures were entered as a covariate into analyses that, therefore, consisted of one between-groups factor: Group (in-

tervention, comparison) and one repeated measures factor, Time (outcomes at 12 months vs. 24 months). In these models, a significant main effect of group would suggest larger changes in one group seen at both 12 and 24 months. A significant Group \times Time interaction would likely indicate that there were no significant between-group differences at one time point, but significant between-group differences at the other time point. Finding no main effects or interaction effects would suggest that the two groups did not differ after either 12 and 24 months.

For ease of comparison with other research and to facilitate later meta-analytic comparisons, unadjusted mean scores for outcome variables at baseline and at 12- and 24-month assessments are displayed in Tables 2 and 4 (children) and Table 5 (parents).

Child outcome. Table 2 displays the results for IQ, MA, raw scores on the Vineland subdomains, and the Early Social Communication Scale measures of Initiating and Responding to Joint Attention. The 2×2 ANCOVA model, used to analyze outcomes at 12 and 24 months, revealed that four of these measures showed an advantage at 12 months for the intervention group over the comparison group that was maintained through to the 24-month assessment point. For IQ, there was a significant main effect of group, $F(1, 41) = 7.72$, $p = .008$, but no interaction effect. Similarly, MA showed a significant main effect of group, $F(1, 41) = 8.37$, $p = .006$, but no interaction effect. Significant group effects (but no interactions) were also found for Vineland Daily Living Skills, $F(1, 41) = 6.32$, $p = .016$, and Vineland Motor Skills, $F(1, 41) = 4.49$, $p = .040$, but not for the Vineland Composite score nor the Socialization and Communication domains. In all cases, children receiving early intensive behavioral intervention were out-performing children in the comparison group.

Seven children (2 in the intervention and 5 in the comparison group) were unable to participate in the baseline Early Social Communication Scale assessment because of behavioral problems, inattention, or absence of parental agreement to videorecording. However, employing Mann-Whitney tests, we were not able to identify differences at baseline, in terms of CA or outcome measures, between those children who accessed the assessment and those who did not. For those children who did, the 2×2 ANCOVAs for 24-month outcomes showed a significant main effect

of group for responding to joint attention in favor of the intervention group, $F(1, 34) = 4.15, p = .049$, but no significant effect for initiating joint attention. Neither measure yielded significant interaction effects again, indicating that the effects were established by 12 months and maintained to 24 months.

Given that the baseline CAs of the intervention and comparison groups (35.7 and 38.3 months, respectively) differed significantly and that CA was correlated with IQ, MA, and some Vineland scores, we ran further ANCOVAs for these variables, with CA as an additional covariate. Three of four of the group effects described above similarly remained significant at conventional levels, but the Vineland Motor Skills main effect achieved only marginal significance, $p = .057$.

Unfortunately, when tested, some children were unable to obtain a score on the Reynell Developmental Language Scales, particularly at baseline, owing to the higher norms produced for the third edition of the test (Edwards et al., 1997). Thus, the raw data for this measure were incomplete. Therefore, we evaluated group effects on the Reynell using a frequency analysis in which the numbers of children obtaining versus those not obtaining a score on the Reynell were compared at the three data-collection points using 2×2 chi-square tests. The group frequencies are shown in Table 3. These tests revealed no differences between groups at baseline for comprehension, but significant differences in favor of the intervention group both at 12 months, $\chi^2(1, N = 44) = 4.13, p = .042$, and 24 months, $\chi^2(1, N = 44) = 8.39, p = .004$. Similarly, the groups did not differ at baseline for expressive language, but significant differences in favor of the intervention group were observed both at 12 months, $\chi^2(1, N = 44) = 5.02, p = .025$, and 24 months, $\chi^2(1, N = 44) = 10.06, p = .002$.

Table 4 shows mothers' and fathers' ratings of their child's behavior problems, prosocial behaviors, and autistic behavior. Analyses of covariance at 24 months revealed a significant group effect for mother-reported positive social behavior, $F(1, 41) = 9.07, p = .004$, and a marginally significant group effect for fathers on this scale, $F(1, 28) = 4.09, p = .053$. In both cases, more positive social behavior was reported for the intervention group. No further significant main effects of group and no interaction effects were

Table 3. Frequencies of Children by Group Achieving a Score on the Reynell Verbal Comprehension Scale and Expressive Language Scale at Three Assessment Points

Assessment time/Group	Reynell Verbal Comprehension Scale	
	Verbal comprehension	Expressive language
Baseline		
Intervention	4	2
Comparison	3	1
12-month		
Intervention	19	17
Comparison	11	8
24-month		
Intervention	21	21
Comparison	11	10

Note. Intervention group $n = 23$ and comparison group, $n = 21$.

found for the other parentally reported child variables.

Parental outcome. Table 5 shows scores on maternal and paternal well-being measures across the 2 years of the study. The only significant finding was a group main effect for paternal depression. Fathers in the intervention group reported more symptoms of depression at both 12 and 24 months, as revealed by a significant main effect in the $2 (\text{group}) \times 2 (\text{time})$ ANCOVA, $F(1, 28) = 5.19, p = .031$.

Analysis of Outcomes for Individual Children

Because IQ has been the primary outcome variable in previous early intensive behavioral intervention research, and here showed the strongest positive change as a result of intervention, we used IQ as the focus for analysis of change for individual children. We first calculated a group effect size for IQ at 24 months to reinforce the clinical significance of the overall intervention effect. The estimate of effect size was based on Cohen's d statistic. Specifically, the mean difference between the two groups' IQ change scores after 24 months was used as the numerator and the pooled SD of the two groups' IQ change scores as the denominator using Cohen's formula (J. Cohen, 1988). The 24-month effect size for IQ calculated using this method was .77, indicating a

Table 4. Unadjusted Means (SDs) of Parental Rating Scales for Child Behavior by Group and Assessment Point

Measure parent ^a	Baseline			12-month assessment			24-month assessment		
	Intervention	Comparison	F	Intervention	Comparison	F	Intervention	Comparison	F
Developmental Behavior Checklist	M 50.26 (22.75)	67.81 (18.77)	45.57 (18.79)	57.71 (22.61)	44.70 (24.20)	60.62 (24.72)			
Total score	F 46.67 (22.15)	57.57 (15.67)	43.67 (16.28)	58.02 (21.05)	45.19 (20.94)	55.20 (19.44)			
Developmental Behavior Checklist	M 22.22 (9.54)	31.14 (9.22)	20.39 (8.54)	25.38 (10.94)	18.91 (10.29)	26.76 (11.21)			
Autism Algorithm	F 22.33 (9.92)	26.29 (8.90)	19.53 (8.23)	25.12 (10.43)	19.50 (8.80)	24.00 (11.60)			
Nisonger Child Behavior Rating Form:	M 10.57 (4.24)	9.29 (3.47)	15.22 (4.09)	11.00 (4.10)	15.30 (4.69)	11.86 (4.84)			
Positive Social Behavior	F 8.94 (3.47)	8.73 (3.67)	13.06 (3.04)	10.40 (4.75)	12.69 (4.06)	11.20 (5.19)			
Autism Screening Questionnaire	M 19.26 (4.93)	21.14 (5.47)	16.43 (5.56)	20.14 (6.55)	15.96 (5.63)	19.29 (7.22)			
	F 20.88 (4.54)	21.07 (6.41)	18.44 (5.54)	20.73 (7.45)	19.88 (6.16)	19.47 (7.46)			

Note. Intervention group mothers (M) *n* = 23 and comparison group mothers, *n* = 21. Intervention group fathers (F) *n* = 16 and comparison group fathers *n* = 15.

relatively large difference between the groups (J. Cohen, 1988, considers a *d* of .80 to be the threshold for a large effect).

To explore whether this difference at the group level was reflected in outcomes for individual children, we applied the criteria outlined by Jacobson and Truax (1991) to establish thresholds for both reliable and clinically significant change for the intervention and comparison groups. The computation of a reliable change index score can be used to establish the IQ change beyond which there is a 95% chance that the observed change does not result from measurement unreliability and/or underlying variability in scores. Calculating the reliable change index score requires two pieces of data: the *SD* of IQs and the stability of the IQ measure. We adopted a conservative approach to the process of identifying these values. Because there were no suitable sources of normative information regarding variance in, and stability of, IQ in very young children with autism, we used the data from the present sample of children rather than drawing on normative information provided by the Stanford-Binet or Bayley tests (i.e., the *SD* for IQ is normally 15). First, we identified the *SD* for IQ for our combined sample of 44 children at baseline. Second, we assessed the 2-year stability of IQ for young children with autism using the correlation between baseline and 2-year IQs for the comparison (untreated) group only. This provided the best available estimate of typical stability in IQ for young children with autism. Substituting these values in Jacobson and Truax’s formula (1991, p. 14) indicated a reliable change index at the standard level of 1.96 equated to a change of 23.94 IQ points; a child’s IQ after 2 years had to deviate from that obtained at baseline by at least that amount before the change was considered reliable; IQ change scores for each child are shown in Figure 1. This reflects the overall group effect, in that more children in the intervention group than the comparison group showed IQ increases over time. Moreover, it shows that 6 children in the intervention group (26%) achieved a reliable improvement over the 2 years of the study. Three of the children (14%) in the comparison group did the same—but 3 (almost 4) children in this group (14% to 19%) also regressed reliably.

Although the use of the reliable change index improves on the methods for establishing best outcome used in previous studies by providing a quantifiable assessment for individual children, it

Table 5. Unadjusted Means (*SDs*) of Self-Report Measures of Parental Well-Being by Group and Assessment Point

Measure	Baseline		12-month assessment		24-month assessment	
	Intervention	Comparison	Intervention	Comparison	Intervention	Comparison
Stress (QRS-F)^a						
Mother	6.43 (4.29)	7.24 (4.19)	7.48 (4.70)	6.48 (4.08)	8.52 (2.97)	8.29 (2.17)
Father	6.81 (4.26)	5.87 (3.19)	7.88 (4.27)	5.53 (3.00)	8.94 (3.62)	7.60 (2.72)
HADS^b						
Anxiety						
Mother	9.35 (4.21)	9.76 (4.87)	10.48 (5.12)	8.52 (4.72)	9.13 (4.53)	8.62 (4.43)
Father	8.89 (4.76)	7.93 (3.67)	7.87 (4.60)	7.00 (3.16)	8.38 (4.08)	8.13 (4.10)
Depression						
Mother	8.13 (4.12)	8.71 (3.68)	8.04 (5.80)	7.19 (4.26)	7.09 (4.97)	6.90 (3.94)
Father	5.69 (4.42)	7.07 (3.61)	6.56 (5.25)	5.27 (2.99)	7.00 (5.34)	5.93 (3.83)
Positive Perceptions (KIPP-PC)^c						
Mother	127.30 (27.00)	133.10 (19.37)	127.39 (23.79)	133.43 (18.23)	128.00 (19.62)	132.43 (17.94)
Father	120.94 (20.23)	124.73 (19.66)	122.56 (19.70)	131.40 (15.68)	122.81 (22.47)	128.53 (9.70)

^aQuestionnaire on Resources and Stress Friedrich short form. ^bHospital Anxiety and Depression Scale. ^cKansas Inventory of Parental Perceptions Positive Contributions scale.

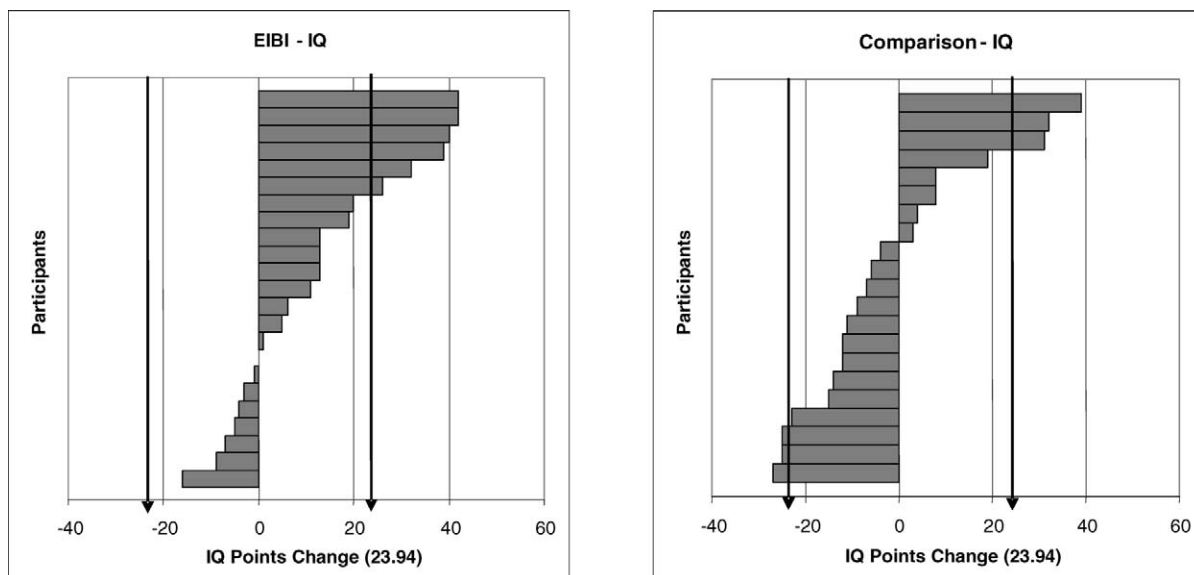


Figure 1. IQ change for children in the intervention and comparison groups. Horizontal bars indicate change in IQ between baseline and 24-month assessment for each child in the intervention group (left panel) and comparison group (right panel). Black vertical lines with arrow-points on both panels indicate the upper and lower bounds for reliable change in IQ calculated according to Jacobson and Truax’s (1991) criteria. EIOI = early intensive behavioral intervention.

is not sufficient to establish the clinical meaning of outcomes. A child's IQ might change reliably without moving his or her score beyond the severely impaired range. Thus, it is useful to identify an IQ above which one would consider a child to be more like children from the typical population than the population of children from which the sample was drawn. Jacobson and Truax (1991) discussed several criteria for establishing the clinical significance of outcomes. Their Criterion C is recommended for use when, as in the present case, it is possible (a) to identify the nonclinical distribution of an outcome variable (e.g., IQ) and (b) to obtain reasonable information about the distribution of the variable in a clinical population. Under Criterion C, the IQ indicating clinical change is halfway between the mean baseline IQ of the children in the present sample and the typical population mean (100). This IQ is 81.93. After 2 years, 5 of the 6 children in the intervention group who achieved reliable change also achieved clinically significant change (i.e., their IQs exceeded 81.93); all 3 children in the comparison group achieving reliable improvement also achieved clinically significant change. No other children in either group achieved a change that was both reliable and clinically significant.

Exploratory Analysis of Variables Associated With IQ Change

Figure 1 is a striking representation of the impact of early intensive behavioral intervention; many more children in the intervention than the comparison group achieved positive outcomes. This, however, begs the question of what factors might be related to intervention success. To consider this, we explored descriptive data on reliable change index-defined *responders* (the 6 children receiving early intensive behavioral intervention whose IQ changed positively to a reliable extent) and *nonresponders* (the 6 children in the intervention group whose IQs decreased (cf. Sherer & Schreibman, 2005). Although we are using the term *nonresponders*, the data presented in Figure 1 suggest that these 6 children's IQs dropped less than might be expected by comparison with the poorest outcome children in the comparison group. The relativity of the term should, therefore, be borne in mind. Table 6 shows mean scores on all continuous variables at baseline for these two small subgroups of children. Means were compared by calculating Cohen's *d* for each

measure. Using rules of thumb suggested by Cohen (1985), we considered differences between reliable change index responders and nonresponders to be worthy of comment if they exceed .50 (medium effect) and .80 (large effect).

These exploratory analyses suggested that children who responded most positively to behavioral intervention differed from nonresponders at baseline in the following ways: They had higher IQ, higher MA, higher Vineland Composite, Communication and Social Skills scores, lower Vineland Motor skills scores, more behavior problems reported on the Developmental Behavior Checklist by both mothers and fathers, more autistic symptoms reported on the Developmental Behavior Checklist Autism Algorithm by both mothers and fathers, and fewer hours of intervention in Year 2.

We also considered the baseline data from the 3 children in the comparison group whose IQ increased to a reliable and clinically significant extent over the 2 years of the study. Because they were very few in number, we were not able to complete formal statistical comparisons, but a visual inspection of their scores on all measures at baseline showed no discernable pattern as a potential explanation as to why they showed reliable improvement.

Discussion

The data from this 2-year controlled comparison of early intensive behavioral intervention against treatment as usual within the United Kingdom education system show a positive advantage for the intervention group. Consistent with other field effectiveness research in this area, robust group main effects were found for IQ, MA, Reynell Expressive Language and Language Comprehension, and Vineland Daily Living Skills after 24 months of intervention. Although less robust, there were also significant changes in Vineland Motor Skills and Responding to Joint Attention as measured by the Early Social Communication Scales. Like H. Cohen et al. (2006), we used ANCOVA methods to explore Group \times Time interactions that would indicate increasing differentiation of performance with continued intervention; and like Cohen et al., we found none.

Although we included a broader range of outcome measures than did previous researchers (H. Cohen et al., 2006; Howard et al., 2005), the impact of behavioral intervention was almost exclu-

Outcome of early intervention for autism

B. Remington et al.

Table 6. Baseline Means (*SDs*) and Effect Sizes of Child Measures for Most and Least Positive Responders in the Intervention Group

Baseline scores	Most-positive responders		Least-positive responders		Effect size
	Mean	<i>SD</i>	Mean	<i>SD</i>	
IQ	65.00	19.81	47.67	11.55	1.07
MA	22.50	6.98	18.00	4.82	0.75
VABS ^a					
Composite	107.17	12.75	99.50	12.28	0.61
Communication	24.17	8.66	16.17	5.04	1.13
Daily living	21.17	6.34	21.17	3.82	0.00
Social	28.67	3.20	25.67	2.50	1.04
Motor	33.17	3.54	36.50	6.38	−0.65
ASQ ^b					
Mothers	21.67	5.28	21.17	4.49	0.04
Fathers	22.00	7.21	22.67	3.50	−0.14
DBC ^c Total					
Mothers	69.67	18.24	51.83	24.05	0.84
Fathers	73.67	9.29	49.80	24.57	1.12
DBC ASA ^d					
Mothers	28.67	8.96	23.83	10.80	−0.49
Fathers	34.00	8.00	24.40	9.89	−1.02
Intervention hours					
Year 1	970.08	342.11	1009.88	113.92	−0.16
Year 2	760.58	533.53	1053.42	356.97	−0.65

^aVineland Adaptive Behavior. Scales Raw Scores. ^bAutism Screening. Questionnaire. ^cDevelopmental Behavior Checklist. ^dAutism Screening Algorithm.

sively on children's cognitive and language abilities and adaptive functioning. Exceptionally, children in the intervention group differentially showed robust improvements in parental ratings of positive social behaviors, but there was no evidence of a similar change in parents' reports of children's behavior problems or ratings of their autistic behaviors. In addition, there were less marked improvements in joint attention. Sallows and Graupner (2005), using domain scores from the ADI-R, also showed reductions in autism symptoms relating to social and communication deficits but no change in ritualistic behaviors. However, it is not clear whether these scores would have changed without intensive intervention as there was no nonintensive intervention comparison group.

The absence of a relative reduction in reported problem behaviors following early intensive be-

havioral intervention is somewhat surprising. It should be remembered, however, that because intervention focuses primarily on educational goals, detailed functional analysis and function-informed interventions for problem behaviors are not the most prominent components. Nevertheless, given the known association between behavior problems and severity of cognitive and adaptive functioning, especially language/communication skills (e.g., McClintock, Hall, & Oliver, 2003), positive benefits of early behavioral intervention on child behavior problems might have been expected. It is possible that the increased ability of the children in the intervention group to respond to bids for attention might have led to the enhancement of their parents' positive perceptions of their prosocial behavior. Given the developmental role of these pivotal skills in facilitating language and cognitive development (Mun-

dy, 1995; Mundy & Crowson, 1997; Mundy & Neal, 1997), this is an important direction for future research.

The present study also extended earlier research by including a detailed analysis of parental outcomes and the first data on fathers. As expected on the basis of previous cross-sectional research (Hastings & Johnson, 2001), the benefits to children of early intensive behavioral intervention did not appear to be at a cost to parents. There was no evidence of differentially increased stress or additional mental health problems in the intervention group mothers or fathers, although the latter reported more symptoms of depression over the course of the study. These fathers, however, had fewer symptoms at baseline compared with those in the comparison group, so the result may, in part at least, be an artifact of a strong regression to the mean effect after 12 and 24 months. These findings are important because difficulties in parental adjustment would reasonably be considered as a contraindication for a home-based behavioral intervention that requires the daily involvement of the family.

Overall, the effect size for the impact of the intervention on the children participating was substantial and clinically meaningful at the group level (Cohen's *d* approaching .80 for IQ after 2 years). Although not reported by H. Cohen et al. (2006), the effect size for IQ in that study—closest in design to our own—was slightly higher than that obtained in the present research (calculated from data presented in Cohen et al. as roughly .90). Thus, our findings are comparable, despite the interventions being delivered over a shorter period of time and with fewer intervention hours. In earlier studies, the impact of intervention at the level of individual participants was rarely quantified; instead, researchers tended to report the number of children scoring within the normal range on standardized measures. In the present study, we extend knowledge by using Jacobson and Truax's (1991) reliable change index statistic as a precise criterion for "best outcome." This revealed that 26% of children receiving early intensive behavioral intervention achieved IQ change that was statistically reliable, and none showed a correspondingly reliable regression in IQ. In the comparison group, 14% improved reliably but, unfortunately, a further 14% regressed reliably.

The reliable change statistic also provides a principled criterion for identifying variables that are common to the children who benefit most

from early intensive behavioral intervention. Exploratory analysis of reliable change index-defined most- and least-positive responders identified correlates of change also identified in previous studies (e.g., H. Cohen et al., 2006; Sallows & Graupner, 2005). These included differences on higher baseline intellectual functioning and adaptive behavior skills (including the total score, communication, and social skills) among the positive responding group. Differences not previously identified were also observed. In addition to poorer motor skills, the most positive responders had more behavior problems and more severe symptoms of autism at baseline. This seemingly paradoxical relation could perhaps have arisen if the measures we used were more sensitive to behavior in those children exhibiting less severe developmental delay. There are no obvious explanations for the positive reliable change in IQs observed for 3 children in the comparison group whose IQs improved to a reliable extent over 2 years.

The present results indicate that behavioral intervention can be effective for young children with autism in the United Kingdom preschool education context, a system—unlike the United States—whose administrators and educators are not familiar with early intensive behavioral intervention and, in some ways, are institutionally unsupportive of it. For example, parents in the United States benefit from Public Laws 94-142 (1975) and 99-457 (1986), which established a right to early intervention services for children from birth to age 3 (the Handicapped Infants and Toddlers Program: Part H). The United Kingdom has no such legislation, and many of its education authorities, during the time of the research, routinely opposed parental attempts to access early intensive behavioral intervention through public provision (Johnson & Hastings, 2002). For these reasons, it was not possible to exert a high degree of control of many practical aspects of the delivery of the intervention. For example, tutors delivering home-based services were not employed by the researchers but by education authorities or the children's families. Staff turnover was common and replacement tutors often difficult to obtain and slow to train. Thus, although an intervention group target intensity of 40 hours per week of input for 2 years was set, positive results were achieved with an average of only 25.6 hours per week. Nevertheless, as required for a convincing demonstration of the field effectiveness, the expected positive outcomes were achieved despite these difficulties.

Like most applied research in early intensive behavioral Intervention, the present study had a number of limitations. First, because it was not a randomized control trial, the few potentially relevant differences detected between groups at baseline (such as CA at treatment onset) had to be controlled statistically, not experimentally. It is, therefore, possible that, although we took the most rigorous steps possible in a study of this kind to manage pre-existing group differences, some remained unobserved. Parenthetically, unobserved differences between groups prior to intervention may also occur under conditions of randomization with samples of the size typically used in early intervention research (Drew et al., 2002). In any case, it would have been very difficult to execute a randomized control trial in the present case, because the independent variable is an extended educational intervention that cannot be delivered “blind” and that has already amassed a considerable body of research attesting to its utility. Given the difficulties in finding an equally credible placebo treatment, it might reasonably be expected that many parents whose children are randomly assigned to a control group would remove them from the study and of these, a percentage would seek the intervention elsewhere (Lord, Wagner et al., 2005). Under these circumstances, intention to treat analyses could be misleading. Perhaps for these reasons, recently published studies in this area (e.g., H. Cohen et al., 2006; Howard et al., 2005) have eschewed randomization.

Procedurally, randomized control trials typically include a precise intervention, often described in a manual; narrow participant selection criteria and blind assessment. Manualized treatment was not a feature of the present study in part because we chose to adopt broad inclusion criteria. It would have been impractical to produce a detailed manual dealing with all possible exigencies but, additionally, the researchers were not in a position to determine the course of therapy for all children in the intervention group who, as noted, received services from a range of providers. Nevertheless, all interventions were supervised by experienced clinicians with detailed knowledge of behavioral programming, and we are confident of the quality of program management. In fact, practical problems of treatment fidelity, primarily the result of tutor shortages, were far more significant than those of treatment coherence. Regarding potential examiner bias, the assessor was independent of the intervention teams and formally

“blind,” but, again for practical reasons, assessment took place in the children’s homes, and in some cases physical or behavioral cues may have signaled the treatment they were receiving. We suspect that it is difficult to control for cues of this kind in any study where there is widespread professional knowledge of the nature of the intervention.

The issue of sample size restriction in the present study also requires consideration. Although we were able to recruit a sample of a size similar to that reported in other early intensive behavioral intervention evaluation research, there is a general problem of statistical power in studies of this kind. Here, two issues are particularly worthy of further comment. First, we found main effect differences on key child outcomes but no significant interaction terms in the 2×2 ANCOVA models. This finding could mean, as H. Cohen et al. (2006) concluded, that the effects of the intervention were established by 12 months. We cannot, however, reliably draw such a conclusion: It is possible that change over the second 12 months was less marked but that in a larger sample we might have seen the advantage for the intervention group continuing to increase. More research addressing this question is needed. A second issue is that we found very little evidence of negative effects of early intensive behavioral intervention involvement on parental well-being, but in a larger sample such effects may have been observed. Although this possibility cannot be eliminated, it is important to consider that the present sample would have been sufficient to show significant or marginal effects that would clearly have become significant with more power. It is also salient that our findings concur with the results of all existing studies in which investigators addressed this question using various designs; none show evidence of a negative effect on family members’ adjustment.

The sample-size restriction also allowed only exploratory effect size analyses of differences between those children in the intervention group who responded most positively and those who regressed. However, this method has some potential for application in other outcome studies and may contribute to the process whereby intervention may be focused on children and their families whose characteristics suggest may maximally benefit from intervention.

In conclusion, the present study indicates that intervention for childhood autism based on applied behavior analysis and delivered intensively

at home during the preschool period can bring about significant changes in children's functioning without a negative impact on other family members, even when delivered in circumstances that for practical reasons do not permit its optimum implementation. Questions remain, however, regarding both the factors that best predict the effectiveness of intervention and the long-term impact of the effects reported. Although parents, educators, and policy makers are likely to ask whether early intensive behavioral intervention "works" or "does not work," it may be more fruitful to pose, instead, smaller but potentially more answerable questions regarding the selection of children for intensive intervention: the identification and evaluation of effective curricula and teaching methods, and the most effective forms of maintenance programs for children at the end of a fixed period of early intervention.

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Anderson, S. R., Avery, D. L., DiPietro, E. K., Edwards, G. L., & Christian, W. P. (1987). Intensive home-based early intervention with autistic children. *Education and Treatment of Children, 10*, 352–366.
- Bayley, N. (1993). *Bayley Scales of Infant Development—Second Edition*. San Antonio, TX: Psychological Corp.
- Behr, S. K., Murphy, D. L., & Summers, J. A. (1992). *User's manual: Kansas Inventory of Parental Perceptions (KIPP)*. Lawrence: University of Kansas, Beach Center on Families and Disability.
- Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism Screening Questionnaire: Diagnostic validity. *British Journal of Psychiatry, 175*, 444–451.
- Bibby, P., Eikeseth, S., Martin, N. T., Mudford, O. C., & Reeves, D. (2001). Progress and outcomes for children with autism receiving parent-managed intensive interventions. *Research in Developmental Disabilities, 22*, 425–447.
- Birnbrauer, J. S., & Leach, D. J. (1993). The Murdoch early intervention program after two years. *Behaviour Change, 10*, 63–74.
- Burack, J. A., & Volkmar, F.R. (1992). Development of low- and high-functioning autistic children. *Journal of Child Psychology and Psychiatry, 33*, 607–616.
- Carter, A. S., Volkmar, F. R., Sparrow, S. S., Wang, J. J., Lord, C., Dawson, G., Fombonne, E., Loveland, K., Mesibov, G., & Schopler, E. (1998). The Vineland Adaptive Behavior Scales: Supplementary norms for individuals with autism. *Journal of Autism and Developmental Disorders, 28*, 287–302.
- Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Developmental and Behavioral Pediatrics, 27*, S145–S155.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Drew, A., Baird, G., Baron-Cohen, S., Cox, A., Slonims, V., Wheelwright, S., Swettenham, J., Berry, B., & Charman, T. (2002). A pilot randomised control trial of a parent training intervention for pre-school children with autism: Preliminary findings and methodological challenges. *European Child and Adolescent Psychiatry, 11*, 266–272.
- Edwards, S., Fletcher, P., Garman, M., Hughes, A., Letts, C., & Sinka, I. (1997). *The Reynell Developmental Language Scales III: The University of Reading Edition*. Windsor: NFER Nelson.
- Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive behavioral treatment at school for 4- to 7-year-old children with autism. *Behavior Modification, 26*, 49–68.
- Einfeld, S. L., & Tonge, B. J. (1995). The Developmental Behaviour Checklist: The development and validation of an instrument to assess behavioral and emotional disturbance in children with mental retardation. *Journal of Autism and Developmental Disorders, 25*, 81–104.
- Einfeld, S. L., & Tonge, B. J. (2002). *Manual for the Developmental Behaviour Checklist* (2nd ed.) University of New South Wales, School of Psychiatry, Melbourne; Centre for Developmental Psychiatry, Monash University, Clayton, Victoria.
- Friedrich, W. N., Greenberg, M. T., & Crnic, K. (1983). A Short Form of the Questionnaire on Resources and Stress. *American Journal of Mental Deficiency, 88*, 41–48.
- Glidden, L. M. & Floyd, F. J. (1997). Disaggregating parental depression and family stress in

- assessing families of children with developmental disabilities: A multi-sample analysis. *American Journal on Mental Retardation*, 102, 250–266.
- Green, G., Brennan, L. C., & Fein, D. (2002). Intensive behavioral treatment for a toddler at high risk for autism. *Behavior Modification*, 26, 69–102.
- Hastings, R. P. (2003a). Behavioral adjustment of siblings of children with autism engaged in applied behavior analysis early intervention programs: The moderating role of social support. *Journal of Autism and Developmental Disorders*, 33, 141–150.
- Hastings, R. P. (2003b). Child behaviour problems and partner mental health as correlates of stress in mothers and fathers of children with autism. *Journal of Intellectual Disability Research*, 47, 231–237.
- Hastings, R. P., & Brown, T. (2002). Behavior problems of autistic children, parental self-efficacy, and mental health. *American Journal on Mental Retardation*, 207, 222–232.
- Hastings, R. P., & Johnson, E. (2001). Stress in UK families conducting intensive home-based behavioral intervention for their young child with autism. *Journal of Autism and Developmental Disorders*, 31, 327–336.
- 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.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19.
- Johnson, E., & Hastings, R. P. (2002). Facilitating factors and barriers to the implementation of intensive home-based behavioural intervention for young children with autism. *Child: Care, Health and Development*, 28, 123–129.
- Kendall, P. C., Chu, B. C., Gifford, A., Hayes, C., & Nauta, M. (1998). Breathing life into a manual: Flexibility and creativity with manual-based treatments. *Cognitive and Behavioral Practice*, 5, 177–198.
- Lord, C., Rutter, M., & LeCouteur, A. (1994). Autism Diagnostic Interview—Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24, 659–685.
- Lord, C., Wagner, A., Rogers, S., Szatmari, P., Aman, M., Charman, T., Dawson, G., Durand, M. V., Grossman, L., Guthrie, D., Harris, S., Kasari, C., Marcus, L., Odom, S., Pickles, A., Scahill, L., Shaw, E., Siegel, B., Sigman, M., Stone, W., Smith, T., & Yoder, P. (2005). Challenges in evaluating psychosocial interventions for autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, 35, 695–708.
- Lovaas, I. O. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, 55, 3–9.
- Lovaas, I. O. (1993). *Teaching developmentally disabled children: The me book*. Austin, TX: Pro-Ed.
- McClintock, S., Hall, S., & Oliver, C. (2003). Risk markers associated with challenging behaviours in people with intellectual disabilities: A meta-analytic study. *Journal of Intellectual Disability Research*, 47, 405–416.
- 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.
- McEvoy, R., Rogers, S., & Pennington, B. (1993). Executive function and social communication deficits in young autistic children. *Journal of Child Psychology and Psychiatry*, 34, 563–578.
- Mundy, P. (1995). Joint attention and social-emotional approach behaviour in children with autism. *Development and Psychopathology*, 7, 63–82.
- Mundy, P., & Crowson, M. (1997). Joint attention and early social communication: Implications for research on intervention with autism. *Journal of Autism and Developmental Disorders*, 27, 653–676.
- Mundy, P., Hogan, A., & Dohering, P. (1996). *A preliminary manual for the abridged Early Social Communication Scales*. Coral Gables, FL: University of Miami. Available at <http://www.psy.miami.edu/fcultypmundy>
- Mundy, P., & Neal, R. A. (1997). Neural plasticity, joint attention, and a transactional social-orienting model of autism. *International Review of Research in Mental Retardation*, 23, 139–168.
- Mundy, P., Sigman, M., Ungerer, J., & Sherman, T. (1986). Social interactions of autistic, men-

Outcome of early intervention for autism

B. Remington et al.

- tally retarded and normal children and their caregivers. *Journal of Child Psychology and Psychiatry*, 27, 647–656.
- Partington, J. W., & Sundberg, M. L. (1998). *The Assessment of Basic Language and Learning Skills: An assessment, curriculum guide, and tracking system for children with autism or other developmental disabilities*. Danville, CA: Behavior Analysts.
- Persons, J. B., & Silberschatz, G. (1998). Are results of randomized clinical trials useful to psychotherapists? *Journal of Consulting and Clinical Psychology*, 66, 126–135.
- Sallows, G., & Graupner, T. (2005). Intensive behavioral treatment for children with autism. *American Journal on Mental Retardation*, 110, 417–438.
- Schopler, E., Mesibov, G., & Baker, A. (1982). Evaluation of treatment for autistic children and their parents. *Journal of American Academy of Child Psychiatry*, 21, 262–267.
- Seligman, M. E. P. (1995). The effectiveness of psychotherapy. *American Psychologist*, 50, 965–974.
- Sheinkopf, S. J., & Siegel, B. (1998). Home-based behavioral treatment of young children with autism. *Journal of Autism and Developmental Disorders*, 28, 15–23.
- Sherer, M., & Schreibman, L. (2005). Individual behavioral profiles and predictors of treatment effectiveness for children with autism. *Journal of Clinical and Consulting Psychology*, 73, 525–538.
- Smith, T., Buch, G. A., & Gamby, T. E. (2000). Parent-directed, intensive early intervention for children with pervasive developmental disorder. *Research in Developmental Disabilities*, 21, 297–309.
- Smith, T., Groen, A. D., & Wynn, J. W. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal on Mental Retardation*, 105, 269–285.
- Sparrow, S., Balla, D. A., & Cicchetti, D. (1984). *Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service.
- Stahmer, A. C., & Ingersoll, B. (2004). Inclusive programming for toddlers with autism spectrum disorders: Outcomes from the Children's Toddler School. *Journal of Positive Behavior Interventions*, 6, 67–82.
- Sundberg, M. L., & Partington, J. W. (1999). The need for both discrete trial and natural environment language training for children with autism. In P. M. Ghezzi, W. L. Williams, & J. E. Carr (Eds.), *Autism: Behavior analytic perspectives* (pp. 139–156). Reno, NV: Context Press.
- Surgeon General. (1999). *Mental health: A report of the surgeon general*. Washington, DC: Department of Health and Human Services.
- Tassé, M. J., Aman, M. G., Hammer, D., & Rojhan, J. (1996). The Nisonger Child Behavior Rating Form: Age and gender effects and norms. *Research in Developmental Disabilities*, 7, 59–75.
- Thorndike, R. L., Hagan, E. P., & Sattler, J. M. (1986). *The Stanford–Binet Intelligence Scale* (4th ed.). Chicago: Riverside.
- Weiss, M. (1999). Differential rates of skill acquisition and outcomes of early intensive behavioral intervention for autism. *Behavioral Interventions*, 14, 3–22.
- Whitehurst, G. J. (2003). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation.
- Zigmond, A. S., & Snaith, P. R. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, 63, 361–370.

Received 6/5/06, accepted 2/26/07.

Editor-in-charge: William E. MacLean, Jr.

This outcome study was funded by a grant from the Health Foundation, UK (<http://www.health.org.uk/>). The authors are most grateful for their generous support of the project. A consortium of 11 Local Education Authorities in the South of England including Southampton, Hampshire, East Sussex, Maidenhead and Windsor, Poole, Brighton and Hove, Wokingham, Wiltshire, and Bournemouth) funded the University of Southampton's intervention services for 13 children in the intervention group. The remaining 10 children in that group received services from PEACH, the London Early Autism Program, and the UK-Young Autism Programme. The authors acknowledge the collaborative support of all these agencies, whether financial and practical, without which the study reported here would not have been practicable. Any opinions expressed herein are those of the authors and are not necessarily endorsed by the research sponsors or collabora-

Outcome of early intervention for autism

B. Remington et al.

tors. Francesca degli Espinosa was the senior supervisor and Erik Jahr served as the external consultant for the University of Southampton intervention. The authors thank: Ruth Littleton, Sophie Orr, and Penny Piggott who, with Paula Alsford and Monika Lemaic, held supervisory posts on the University of Southampton team; Corinna Grindle assisted with reliability analyses;

and Catherine Carr provided outstanding administrative and logistical support to the team. Requests for reprints should be sent to Bob Remington, Centre for Behavioural Research Analysis and intervention in Developmental Disabilities (BRAIDD), School of Psychology, University of Southampton, Southampton, SO17 1BJ, UK. E-mail: rer1@soton.ac.uk