Behavior Modification
http://bmo.sagepub.com/

Two-Year Outcomes for Children With Autism After the Cessation of Early Intensive Behavioral Intervention

Hanna Kovshoff, Richard P. Hastings and Bob Remington

Behav Modif 2011 35: 427 originally published online 17 May 2011
DOI: 10.1177/0145445511405513

The online version of this article can be found at:
http://bmo.sagepub.com/content/35/5/427

Published by:
SAGE
http://www.sagepublications.com

Behavior Modification
35(5) 427–450

DOI: 10.1177/0145445511405513
http://bmo.sagepub.com
SAGE

Two-Year Outcomes for Children With Autism After the Cessation of Early Intensive Behavioral Intervention

Hanna Kovshoff[1], Richard P. Hastings[2], and Bob Remington[1]

Abstract

Evidence from recent meta-analytic and narrative review suggests that early intensive behavioral intervention (EIBI) may improve life chances of preschool children with autism. Unfortunately, there are few data indicating whether early gains are maintained after intervention ceases. The purpose of the present study was to establish the 2-year follow-up outcome for children with autism ($N = 41$) who had participated in an earlier 2-year controlled comparison of EIBI. Twenty-three children in the intervention group (100% of original sample) and 18 in the treatment-as-usual comparison group (86% of original sample) were located and retested. Group differences favoring intervention substantially diluted in this period but varied significantly between subgroups who had received university-supervised and parent-commissioned interventions, favoring the latter. These groups differed in terms of their baseline characteristics and intensity of intervention. Results strongly suggest a need for better characterization of those children who would benefit from more active maintenance programs.

[1]University of Southampton, Hampshire, UK
[2]Bangor University, Gwynedd, UK

Corresponding Author:
Hanna Kovshoff, PhD, School of Psychology,
University of Southampton, Highfield,
Southampton, SO17 1BJ, UK
Email: hk@soton.ac.uk

Keywords

autism, early intensive behavioral intervention, ABA

For almost three decades, researchers have conducted evaluation studies attesting to effectiveness of early intensive behavioral intervention (EIBI) for children with autism, an approach originated by Lovaas (1987). The evidence base is such that meta-analytic reviews of outcome studies suggest that EIBI has a positive effect on intellectual and adaptive functioning (Eldevik et al., 2009, 2010; Makrygianni & Reed, 2010; Reichow & Wolery, 2009; Virués-Ortega, 2010). For example, using a weighted mean difference effect size across 10 controlled evaluations of EIBI, Eldevik et al. (2009) found a large effect for IQ change (Hedges's g = 1.10, 95% confidence interval [CI] = [.87, 1.34]) and a moderate and statistically significant effect for change in adaptive behavior composite scores (Hedges's g = .66, 95% CI = [.41, .90]). In addition, Rogers and Vismara (2008) used established criteria for empirically supported interventions to conclude that EIBI has generated sufficient supporting data for it to be identified as "well established," whereas other comprehensive interventions for children with autism have not as yet reached this standard. The international impact of this research evidence on early intervention policy has been substantial. For example, there is some state funding for intensive behavioral intervention in Norway (Eikeseth, Smith, Jahr, & Eldevik, 2002, 2007) and large-scale province-wide implementation has been delivered in Canada (Bryson et al., 2007; Perry et al., 2008).

One element of the supporting case for any early intervention approach is that a relatively high initial investment should produce substantial long-term financial savings. Thus, implicit in early intervention evaluation studies is the idea that an intensive but time-limited treatment will allow its recipients to subsequently engage with, and benefit from, more mainstream educational services (e.g., Lovaas, 1987; McEachin, Smith, & Lovaas, 1993), thus reducing their need for special educational services in the medium term and their dependence on social and health care support into adolescence and adulthood. In fact, any intervention that could have a relatively small long-term effect on the level of support required for individuals with autism across the life span (e.g., from residential and day center care to a community home and supported employment) would result in major lifetime cost savings to society (Järbrink & Knapp, 2001). Based on Lovaas' (1987) data, EIBI might be able to have substantial impact on the level of support required by individuals with autism, at least in terms of outcome into late childhood (McEachin et al., 1993).

Following the basic economic case for early intervention, it is thus crucial from a public policy standpoint (in addition to the standpoint of parents and individuals with autism themselves) to examine data relating to the maintenance of any initial gains that may result from early intervention. Table 1 shows the pattern of assessments in several recent, widely cited studies of intensive behavioral intervention. Only one includes a partial, and as yet unpublished, postintervention follow-up of children after an initial period of early intervention had ceased (Sallows & Graupner, 2008). Earlier, Sallows and Graupner (2005) had conducted a randomized trial of 35 children who had received either a clinic-directed, intensively supervised intensive behavioral intervention, or parent-directed, less intensively supervised intervention of the same kind. After 4 years of intervention, the groups did not differ on measures of cognitive, language, adaptive, social, or academic measures. Three years after ending the program, Sallows and Graupner (2008) reported that all the 17 rapid learners (i.e., "best outcome" children) drawn from both groups showed continued improvement in verbal IQ, language (receptive and expressive), and socialization. Moreover, 15 of them continued to succeed in regular classes.

Lack of peer-reviewed data on the maintenance of gains for EIBI makes McEachin et al.'s (1993) follow-up of Lovaas's (1987) study unique. Unfortunately, however, several aspects of the McEachin et al.'s study seriously compromise interpretation of the data in terms of the economic public policy case for early intervention. First, children did not receive a time-limited intervention and, moreover, they were reassessed at different ages and after they had ceased receiving intervention for differing periods. Second, there was a significant bias in this respect because best outcome children who had received the intervention and "achieved normal functioning" (nine are so described in Lovaas, 1987) by the age of 7 years (after an unspecified intervention duration) received no further intervention. Children who did not achieve "normal functioning" remained in intervention at the request of their parents, such that when the group was followed up at a mean age of 13 years (range = 9-19 years), these children had only been out of intervention for a mean of 5 years (range = 0-12 years). Third, although children from one of the initial control groups were followed up, this was the low-intensity behavioral intervention group rather than the "treatment-as-usual" (TAU) comparison. Children in the less intensive behavioral intervention group were followed up at mean age of 10 years (range = 6-14 years), having been out of intervention for a mean of 3 years (range = 0-9 years). Finally, to examine the case for "continued normal functioning" for the best outcome children, a new comparison group of typically developing children was recruited. Such a post hoc amendment to the

Table 1. Characteristics of Prospective Outcome Research on Early Intensive Behavioral Intervention

Authors	Age at start of program (baseline; in months)	1st assessment point (time since baseline; in months)	2nd assessment point (time since baseline; in months)	3rd assessment point (time since baseline; in months)	4th assessment point (time since baseline; in months)	Length of follow-up period without intervention
Sheinkopf and Siegel (1998)	$M = 34$	20.6	0	0	0	0
Smith, Buch, and Gamby (2000)	35-45	24-36	0	0	0	0
Smith, Groen, and Winn (2000)	18-42 ($M = 36.07$)	48-60	0	0	0	0
Bibby, Eikeseth, Martin, Mudford, and Reeves (2002)	$M = 45$	12	24	0	0	0
Howard, Sparkman, Cohen, Green, and Stanislaw (2005)	<48 ($M = 30.86$)	14	0	0	0	0
Sallows and Graupner (2005, 2008)	24-42 ($M = 36$)	12	24	36	48	3 years
Cohen, Amerine-Dickens, and Smith (2006)	<48	12	24	36	0	0
Eikeseth, Smith, Jahr, and Eldevik (2007)	48-84 ($M = 65$)	12	24	0	0	0
Kovshoff, Hastings, and Remington (this article); Remington et al. (2007)	30-42	12	24	0	0	2 years

evaluation design is a significant threat to internal and external validity when the policy interest is in the comparative outcomes against a TAU control group.

Recently, we reported outcomes of the Southampton Childhood Autism Programme (SCAmP), a field-effectiveness evaluation of a 2-year EIBI program (Remington et al., 2007). Outcomes of primary interest were IQ and adaptive functioning. On the basis of parental preference, children with autism aged between 2.5 and 3.5 years were placed in either an intervention group ($n = 23$) or a comparison group ($n = 21$). As noted by Remington et al. (2007), children in the intervention group received EIBI provided by either (a) a university-supervised program delivered from the University of Southampton and offered free at the point of use for parents nominated by the local education departments funding the university program ($n = 14$) or (b) an intervention commissioned and purchased directly by parents (or funded by their education departments, typically following judicial review) from other U.K. EIBI service providers, operating privately ($n = 9$).[1]

Some studies have treated similar differences between intervention provisions as an independent variable. For example, Sallows and Graupner (2005) and Smith, Groen, & Wynn (2000) each compared an intensive clinic-directed treatment with a less intensive but appropriately supervised parent-directed intervention. The former found no between-group differences in IQ following intervention but Smith et al. identified differences favoring the clinic-directed group. In the Remington et al. (2007) study, however, children were not randomized between university-supervised and parent-commissioned subgroups, and there was no a priori reason to suspect that differences in treatment intensity would emerge. Importantly, both the subgroups received intervention meeting the key criteria for research-based interventions identified by Green, Brennan, and Fein (2002). Therefore, for purpose of analysis, Remington et al. aggregated data from the parent-commissioned and university-supervised subgroups into a single intervention group. This combined group received a mean of 25 hr of behavioral intervention for 2 years. The comparison group received their local education departments' prescribed support (TAU) for the same period. In addition to earlier access to preschool placements (comparison children were attending school for a weekly mean of 16 hr after 12 months and 18 hr after 24 months), TAU included speech therapy, treatment and education of autistic and related communication handicapped children (TEACCH), picture exchange communication system (PECS), sign training, diet, and prescribed and other medication.

Prospective assessment was undertaken before intervention, after 12 months of intervention, and again after 24 months. Groups did not differ on assessments at baseline but, after 24 months, robust differences favoring intensive behavioral intervention were observed on measures of intelligence, language, daily living skills, joint attention, and positive social behavior. A Reliable Change Index (RCI; Jacobson & Truax, 1991) was calculated for IQ change. The RCI is a statistical indicator of change for individual children. It can be considered an index of best outcome because it sets a criterion (at $p < .05$) beyond which change attributable to baseline variability and test–retest measurement error can be discounted. Of the 23 children receiving behavioral intervention, 6 showed a positive change in IQ on this index. Of the 21 comparison-group children, 3 showed a similar positive change and an additional 3 showed reliable negative change (i.e., IQ regression).

The Remington et al. (2007) study offered an opportunity to conduct a follow-up evaluation to assess the maintenance of gains achieved during the 24-month, fixed-duration intervention. At termination, all children in both the intervention and the comparison groups found school placements in either mainstream or special education classrooms. Although some parents of children in the intervention group continued to use applied behavior analysis methods and to supplement school provision by employing external service providers, this subgroup was a minority of the sample (7 from 23 families).

Thus, we designed the present study to examine maintenance of EIBI gains following the termination of a 24-month intensive program. A follow-up interval of 2 years was chosen to ensure that adequate time had passed for a fair evaluation of medium-term impact to be conducted with minimal loss of participating families as a result of natural attrition. It also ensured children had not yet reached the age of earliest transition within the U.K. education system between "infant" and "junior" provision at age 7 to 8 years. Children were assessed using the same or similar measures employed in the original study. Thus, we were able to explore whether the two groups continued to differ over time.

Method

Participants

For inclusion in the Remington et al. (2007) study (where a fuller description of participants' characteristics can be found), all 44 children in both the intervention and comparison groups were required to meet criteria for a diagnosis of autism based on both the Autism Diagnostic Interview–Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994) and an independent clinical assessment and diagnostic procedure. All were free of chronic medical conditions and resident in the family home (Remington et al. 2007).

Of these 44 children, we were able to locate and retest all in the intervention group ($n = 23$), 2 years after the termination of EIBI. As noted above, the intervention group can be further broken down into two distinct groups of children: (a) a university-supervised group ($n = 14$) for which the intervention programs were supervised by a team employed by the university (analogous to a more conventional clinic-based program) and provided free of charge to children nominated by local education departments and (b) a parent-commissioned group ($n = 9$) for which intervention program teams (including professional applied behavior analysis [ABA] consultants) were recruited, employed, and managed by the parents. These two groups were combined in the original Remington et al., (2007) article because the group outcomes did not differ significantly at 24 months (see Results section for a more detailed description).

In all, 18 of the 21 children in the comparison (TAU) group were located and retested 2 years after their last assessment. We were unable to locate two comparison-group children and one set of parents declined further participation in the follow-up research. There was no obvious bias in the failure to recruit comparison-group children into the follow-up study. In fact, the best outcome children in the comparison group (i.e., those who achieved positive reliable change in IQ over 2 years) were all recruited and so any bias would be likely to act against finding results in favor of EIBI. At time of follow-up, all 41 children were aged between 6.5 and 8 years (mean of 7 years, 2 months). Parental interviews using the most recent version of the ADI-R (Rutter, Le Couteur, & Lord, 2003; see below) indicated that all children continued to meet criteria for a diagnosis of autism.

Seven of the 14 children in the university-supervised intervention group, and 7 of the 9 children in the parent-commissioned intervention group, were in mainstream school at the time of the follow-up. Anecdotal, retrospective reports from parents indicated that two of the university-supervised group children continued to receive some ABA-based support after they had exited the early intervention program and started at school. Based on similar reports, five of the children in the parent-commissioned group also continued to receive some ABA-based support once they ceased their early intervention home program and began to attend school. None of the children in the comparison group received consistent ABA-based educational interventions at home or in their school placements.

Materials

Most of the cognitive, language, and behavioral outcome data for the children in this study were obtained using the same norm-referenced instruments used in Remington et al.'s (2007) study where their psychometric properties are fully described (pp. 423-425). To avoid repetition, and because most of these measures are widely used, only brief descriptions of instruments used in the earlier study are provided below.

Intellectual functioning. We again employed the Stanford–Binet Intelligence Scale–Fourth Edition (Thorndike, Hagen, & Sattler, 1986) for children with sufficiently advanced language skills to take a full-scale IQ test. The Bayley Scales of Infant Development–Second Edition (Bayley, 1993) was used to assess IQ of children whose language skills were not sufficiently advanced (calculating a mental age [MA] score based on their raw score using Table B.2 in the manual and computing a ratio IQ using the MA/CA × 100 (mental age/chronological age x 100) formula).

Adaptive skills. We chose the Vineland Adaptive Behavior Scale–Survey Form (VABS; Sparrow, Balla, & Cicchetti, 1984) to assess adaptive behavior across the domains of socialization, communication, and daily living skills. In the present analysis, standardized adaptive behavior scores were used.

Language. The Reynell Developmental Language Scales–Third Edition (RDLS-III; Edwards et al., 1997) were again chosen, despite the fact that normative data are only available from the age of 21 months. Although this previously led to difficulties in obtaining scaled scores for many of the children (see. Remington et al., 2007, p. 428), the test was retained to facilitate comparison with the postintervention data.

Rating scales for child behavior. The Positive Social Subscale of the Nisonger Child Behavior Rating Form (NCBRF; Tassé, Aman, Hammer, & Rojhan, 1996) and the parent report version of the Developmental Behavior Checklist (DBC; Einfeld & Tonge, 1995) were chosen to assess child prosocial behavior and behavior problems. The former is specifically designed to assess children with intellectual disabilities; the latter offers an index of the severity of behavior problems.

The ADI-R. It (Lord et al., 1994; Rutter et al., 2003) is a semistructured diagnostic interview measure. It was used to confirm diagnosis of autism in Remington et al.'s (2007) study. The ADI-R contains questions regarding children's early development, communication and social interaction skills, and behavioral patterns, and it is based on the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; American Psychiatric Association, 2000) and the International Statistical Classification of Diseases and Related Health Problems (10th ed.; ICD-10; World Health Organization, 1992) criteria for autism and pervasive developmental disorders. Both current and overall diagnostic scores may be derived from the measure's algorithm, but the current score was used in the present study.

Procedure

Parents were initially contacted by letters that requested their participation in a follow-up assessment at the 2-year anniversary of their last assessment, which corresponded to the termination of the intensive intervention program for the EIBI group. When written consent had been obtained, families were telephoned to arrange the dates and times of (a) an assessment visit to the family home (or, exceptionally, the child's school) to collect psychometric data from the child (e.g., Stanford–Binet, Reynell) and (b) two telephone interviews to collect data from parents (VABS and ADI-R). If the initial letter failed to elicit a response, parents were contacted by telephone to explain the rationale behind the follow-up study, after which they were asked to consider participating once more. Those who agreed were sent a fresh consent form to sign and return to the researchers through the mail.

Assessments were arranged at as close a time as possible to the 2-year anniversary of the end of the intervention or, for the comparison-group children, to the last assessment date. Data were obtained by a doctoral level, trained psychometrician (H.K.), with more than 6 years experience assessing children with autism. As before, she operated independently of the program supervisory and delivery teams who had ceased employment with the university in the period between the initial study and the follow-up. She exercised every caution to obtain reliable and valid data but was not blind to intervention-group status.

All questionnaires were mailed out to parents approximately 2 weeks prior to the children's assessment visits. Vineland and ADI-R telephone interviews were conducted with primary caregivers in two separate telephone calls approximately 7 to 14 days before the home/school visit. These visits lasted approximately 90 min, during which time the researcher administered the standardized outcome measures using a uniform order of administration: (a) the Stanford–Binet or the Bayley Scales of Infant Development and (b) the RDLS (the Reynell was attempted in all cases but administered in full only if a child's language level was such that he or she could access the items on the test).

Results

Group Differences at Follow-Up

All children were attending school at the time of the follow-up, but significantly more of those in the intervention group (14 from 23) than the comparison group (4 from 18) were in mainstream education settings, $\chi^2(df = 1) = 6.12$,

$p = .013$. All but one of the children in mainstream settings was, however, in receipt of some 1:1 support in the classroom for at least some proportion of time.

Differences between the intervention and comparison groups at follow-up were first evaluated using one-way ANCOVA models in which group was the between-subjects factor and the relevant preintervention baseline score on the dependent measure was the covariate. Table 2 shows the unadjusted mean scores for the intervention and comparison groups on all outcome measures at the baseline, 24 month, and 2-year follow-up time points, together with the results of the ANCOVA comparisons. When baseline scores were controlled, there were no statistically significant group effects at follow-up. Because child chronological age at baseline differed by an average of 3 months between the two groups (see Remington et al., 2007), the ANCOVA models were repeated with age as an additional covariate. No differences between the two groups at follow-up emerged, so these analyses are not reported further here.

Following Remington et al. (2007), language development of the children in both groups at follow-up was assessed in terms of whether they were able to achieve a standard score on the RDLS.[2] At follow-up, more children in the intervention group (22 from 23) were able to score on the Reynell Receptive Language Scale than children in the comparison group (13 from 18; Fisher exact test one-tailed, $p = .048$). For the Reynell Expressive Language Scale, the difference between the intervention group (20 from 23) and the comparison group (12 from 18) was not statistically significant (Fisher exact test one-tailed, $p = .12$).

Post Hoc Analysis of Outcomes for Different EIBI Delivery Models

In the absence of consistent evidence for the maintenance of intervention gains, and taking into account the fact that different service delivery models have already been shown to produce differences in intervention outcome (Smith Groen & Wynn 2000), we explored the effect of disaggregating the data for the parent-commissioned and university-supervised intervention subgroups. It is important to note from the outset that, because subgroup membership was not based on random assignment, the groups may have differed in ways other than the service delivery model (see below). In fact, we had, as a precaution, already compared the outcomes for the university-supervised, parent-commissioned, and comparison subgroups immediately following 24-months intervention using repeated measures analysis of covariance models (controlling for baseline scores). However, because there were no significant EIBI

Table 2. Baseline, 24 Month, and 2-Year Follow-Up Unadjusted Mean Scores for the Intervention Group and the Comparison Group

	Intervention group (n = 23)						Comparison group (n = 18)							
	Baseline		24 months		2-year follow-up		Baseline		24 months		2 year follow-up			
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	*p*
IQ	61.43	16.43	73.48	27.28	64.65	33.04	63.83	13.98	61.00	27.30	61.94	31.09	0.94	.339
Vineland Composite	60.22	5.82	61.48	15.35	55.13	19.40	57.17	7.05	55.28	13.80	49.5	17.39	0.07	.79
Vineland Communication	61.52	7.56	68.00	24.01	62.65	25.11	58.17	8.63	61.28	23.07	57.72	24.54	0.08	.784
Vineland Daily Living	63.26	5.40	59.70	11.80	52.35	19.61	62.22	8.14	53.83	13.23	43.67	18.15	1.89	.177
Vineland Socialization	63.30	6.74	64.04	14.11	62.57	16.93	59.94	7.94	60.11	13.57	59.33	15.58	.05	.822
Nisonger Child Behavior Rating Form: Positive Social Behavior														
Mother	10.57	4.24	15.30	4.69	14.87	5.29	9.61	3.50	12.06	5.20	11.33	5.26	3.80	.059
Father	8.94	3.47	12.69	4.06	14.73	6.70	8.64	3.79	11.21	5.38	11.64	7.31	1.02	.321
DBC total														
Mother	50.26	22.75	44.70	24.20	53.70	21.13	65.61	18.70	56.78	23.50	63.56	26.39	0.24	.627
Father	46.67	22.15	45.19	20.94	48.86	26.21	57.15	16.23	54.79	20.10	56.14	21.22	0.13	.719

Note: DBC = Developmental Behavior Checklist. F and p values in table refer to ANCOVA that compared group scores at 24-month treatment termination and 2-year follow-up, controlling for baseline. Intervention-group mothers n = 23 and comparison-group mothers n = 18. Intervention-group fathers n = 15 and comparison-group fathers n = 14.

delivery-model differences, we did not report these analyses in Remington et al.'s (2007) study. Nevertheless, before exploring potential subgroup differences at follow-up, we repeated the analyses with the data obtained after 24 months of intervention, this time using the slightly reduced comparison-group sample available at follow-up. This new analysis for IQ showed a statistically significant group effect, $F(1, 37) = 5.67$, $p = .022$, but exploratory pairwise post hoc comparisons between the three groups, using a least significant difference comparison, revealed that the group effect was driven only by differences between each of the two intervention subgroups and the comparison group, and not by differences between the different intervention groups (the mean IQ postintervention difference between the university-supervised and parent-managed subgroups was 0.741, $p = .917$). After covarying baseline scores, a similar pattern emerged for all other main outcome variables.

To explore changes in the three subgroups between 24 months (i.e., postintervention) and 2-year follow-up, we conducted an analysis of covariance, controlling for baseline differences. Figure 1 shows the change in IQ adjusted means for the three groups. Although IQ is stable for both the parent-commissioned group and comparison group over the follow-up period, it fell markedly for the university-supervised group, reflected in a significant group × time interaction, $F(2, 37) = 10.80$, $p < .0001$.

Thus, although the EIBI delivery model subgroups did not differ in IQ postintervention, a substantial difference had emerged at the 2-year follow-up. To further explore potential subgroup differences over time, we plotted the unadjusted IQs at baseline, 24 months, and 2-year follow-up. Figure 2 shows the trajectory for each of the three groups. It is clear that, although both intervention subgroups improved over the treatment period, they differed from each other at each of the time points. Because both the starting point and trajectory of change in outcome measures for the two EIBI delivery models appeared quite different, we switched the focus of our analysis to changes over time within the three subgroups individually, rather than conducting between-subgroup analyses (either with or without baseline score covariates). Mixed model ANOVA or ANCOVA to analyze subgroup trajectories were additionally contraindicated because the subgroup sizes were small. For the same reason, and to reduce our exposure to Type II errors in these exploratory analyses, we did not adjust the alpha level for each test.

Thus, we conducted one-way repeated measures ANOVAs for each of the three subgroups and for each outcome variable, using contrasts to compare baseline with 24 month scores and 24-month scores with 2-year follow-up scores. Where the assumption of sphericity was violated, we adopted the Greenhouse–Geisser correction (adjusted degrees of freedom are reported

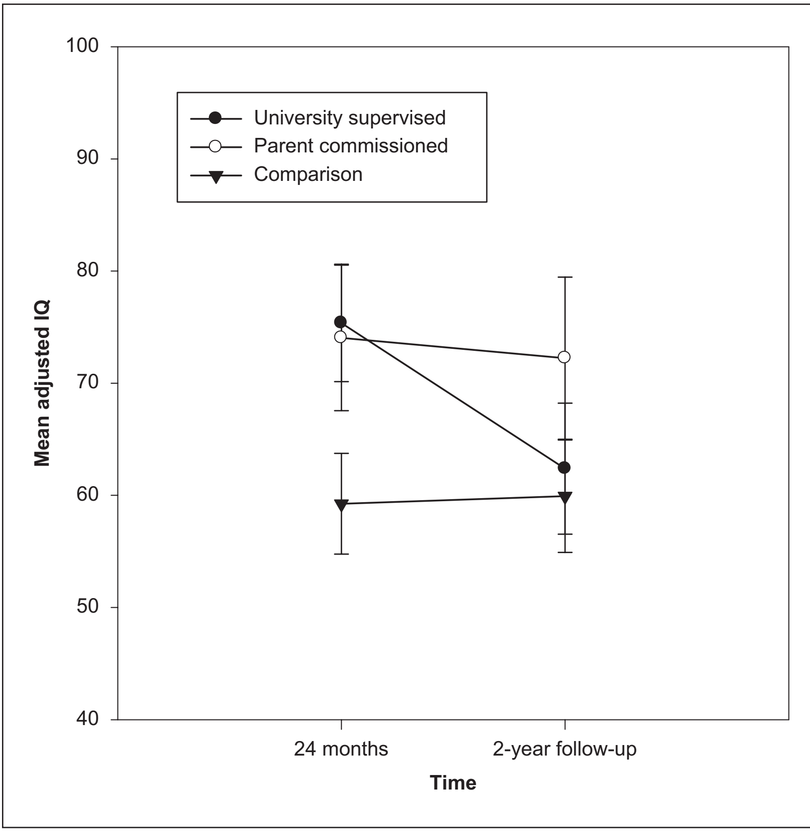

Figure 1. Adjusted mean IQ by subgroup after 24 months of intervention and 2-year follow-up, controlling for baseline IQ
Note: Error bars represent standard errors.

where relevant). In consideration of the small group sizes and the potential for non-normal distributions to affect the ANOVA results, we repeated the analyses using an equivalent nonparametric test (Friedman's one-way analysis of variance by ranks) but we have not reported these analyses here because they produced the same pattern of results. The means and standard deviations on all measures and across testing points are shown in Table 3.

The ANOVAs revealed that the parent-commissioned group made a significant gain in IQ between baseline and 24 months (d = .67); over the same interval, the university-supervised group's IQ gain was marginally significant

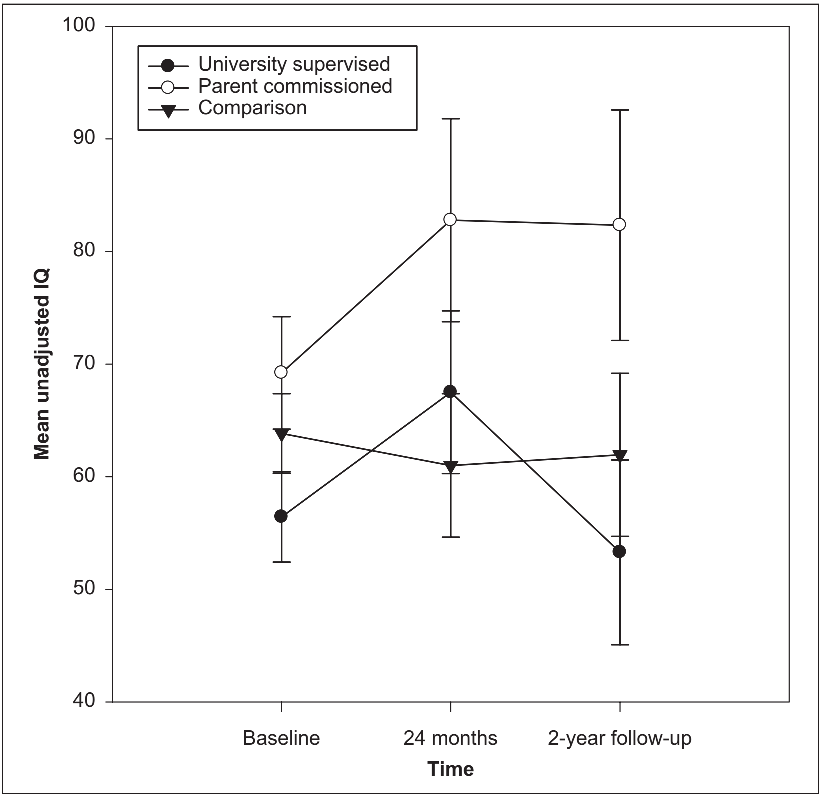

Figure 2. Unadjusted mean IQ by subgroup at baseline, after 24 months of intervention, and 2-year follow-up
Note: Error bars represent standard errors.

with a comparable effect size (*d* = .50). However, in the period between the 24-month assessment and the 2-year follow-up, the university-supervised group showed a significant decline in IQ (*d* = –.49) but the parent-commissioned group maintained their improved score (*d* = –.02). A similar pattern during the follow-up period of significant skill losses in the university-supervised group and maintenance in the parent-commissioned group was also found for the VABS composite score as well as its communication and daily living subdomains. In addition, mothers in the parent-commissioned group reported significant increases in positive perceptions of their child between the 24 month and 2-year follow-up period. Chi-square analyses revealed there were no significant subgroup differences in terms of ability to access items on the RDLS.

Table 3. Repeated Measures ANOVAS and Planned Contrasts for Intervention Subgroups

Measure	University supervised									Parent commissioned								
	BL	24 months	2-year follow-up	Repeated measures ANOVA		Planned contrast: BL-24 months		Planned contrast: 24 months-2 years		BL	24 months	2-year follow-up	Repeated measures ANOVA		Planned contrast: BL-24 months		Planned contrast: 24 months-2 years	
Measure	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*F*	*p*	*F*	*p*	*F*	*p*	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*F*	*p*	*F*	*p*	*F*	*p*
IQ	56.43 (16.33)	67.50 (27.60)	53.29 (31.06)	4.90	.035	4.37	.057	37.06	<.001	69.22 (14.75)	82.78 (25.47)	82.33 (29.28)	4.90	.044	7.51	.025	0.02	.880
Vineland Composite	57.83 (4.56)	56.86 (15.18)	49.93 (21.11)	2.30	.148	0.07	.790	8.361	.013	63.89 (5.88)	68.67 (13.34)	63.22 (13.78)	1.73	.209	2.30	.168	5.28	.051
Vineland Communication	58.50 (5.75)	65.07 (24.93)	56.00 (26.90)	1.78	.204	1.39	.263	14.01	.002	66.22 (7.90)	72.56 (23.17)	73.00 (18.98)	0.71	.508	0.84	.386	0.01	.942
Vineland Daily Living	61.00 (4.33)	55.36 (10.82)	46.86 (20.95)	6.32	.021	4.67	.050	6.12	.028	66.78 (5.19)	66.44 (10.43)	60.89 (14.48)	2.13	.152	0.03	.871	2.88	.128
Vineland Socialization	61.07 (5.65)	58.71 (11.78)	56.79 (16.20)	0.91	.363	0.57	.465	1.72	.212	66.78 (7.12)	72.33 (13.96)	71.56 (14.59)	1.64	.225	2.34	.164	0.12	.738
Nisonger Child Behavior Rating Form: Positive Social Behavior																		
Mother	10.43 (3.84)	14.93 (5.09)	13.29 (5.21)	4.31	.024	8.09	.014	1.43	.253	10.78 (5.04)	15.89 (4.23)	17.33 (4.64)	11.72	.007	9.56	.015	9.26	.016
Father	8.56 (2.92)	11.00 (4.15)	11.33 (5.48)	1.16	.337	3.77	.088	0.12	.893	9.43 (4.28)	14.86 (2.91)	19.83 (5.08)	16.61	.001	46.88	.001	3.98	.103
DBC total																		
Mother	54.14 (23.17)	49.00 (28.28)	58.93 (20.88)	1.30	.290	0.69	.420	2.56	.134	44.22 (21.99)	38.00 (15.10)	45.56 (19.92)	1.07	.365	1.25	.279	2.66	.142
Father	53.22 (23.30)	51.33 (20.17)	55.56 (27.19)	0.08	.922	0.10	.755	0.17	.695	36.83 (17.67)	37.29 (20.59)	39.50 (20.49)	0.10	.910	0.03	.878	0.89	.400

Note: BL = baseline; DBC = Developmental Behavior Checklist. Means and standard deviations for each time point, with ANOVA values for within-participant main effects and planned contrasts. University-supervised group mothers $n = 14$ and parent-commissioned group mothers $n = 9$. University-supervised group fathers $n = 9$ and parent-commissioned group fathers $n = 6$.

Table 4. Intervention Subgroup Differences on Key Baseline Measures

Measure	Group mean and standard deviation scores: University supervised (n = 14)	Parent commissioned (n = 9)	p
Baseline ADI Composite score	43.36 (8.07)	36.75 (5.18)	.051
Baseline ADI Social Interaction score	23.86 (3.86)	19.89 (3.30)	.019
VABS Composite	57.86 (4.56)	63.89 (5.88)	.012
VABS Communication	58.50 (5.75)	66.22 (7.90)	.013
VABS Daily Living	61 (4.33)	66.78 (5.19)	.009
VABS Socialization	61.07 (5.65)	66.78 (7.12)	.045
IQ	56.43 (16.33)	69.22 (14.75)	.071
Hours Year 1	912.66 (147.40)	1,213.64 (246.59)	.001
Hours Year 2	840.71 (270.96)	1,243.75 (385.31)	.008
Total hours (Years 1+2)	1,753.38 (404.96)	2,457.39 (494.46)	.001

Note: ADI = Autism Diagnostic Interview; VABS = Vineland Adaptive Behavior Scale. t tests comparing baseline scores on key measures for the university-supervised and parent-commissioned intervention subgroups.

Although we initially focused on service delivery as the basis for examining intervention subgroups because previous studies had reported differences in outcome (Smith, Groen & Wynn, 2000), we were aware that, in the absence of random assignment, subgroups may have differed on many variables in addition to the delivery model used. To identify other candidate variables that might have accounted for trajectory differences between subgroups, we conducted several exploratory analyses using t tests to compare the two intervention subgroups at baseline. Subgroup means and the p values associated with these comparisons are summarized in Table 4. The analyses revealed that the university-supervised group began the study with more severe symptoms of autism as measured by the ADI-R, lower levels of adaptive behavior, and, as Figure 2 also shows, a marked trend toward lower IQ. We also identified that the intervention subgroups differed in terms of treatment intensity in both years of intervention, with the university-managed group receiving fewer hours in Year 1 and Year 2. Thus, not only was the nature of the EIBI delivery model different for the subgroups but also the children enrolled into each subgroup were different, and the overall intensity of intervention differed.

Discussion

This study offered a unique opportunity to establish whether behavioral intervention delivered intensively to children with autism during the 2 years before they began to attend full-time education would produce enduring benefits, compared with the interventions that were typically provided by educational services during this time. Because our initial evaluation was designed as a field-effectiveness study, it offered an indication of what might be expected in fairly typical clinical contexts in terms of the impact of intervention and the medium-term maintenance of intervention gains.

Intervention was associated with a greater likelihood of attending mainstream education after 2 years of treatment. These findings replicated the original demonstration of EIBI (Lovaas, 1987), which treated school placement as an outcome variable. Overall, however, results indicated that, for at least some proportion of children, initial gains on several critical outcome measures associated with EIBI were not maintained at follow-up.

Because children in our intervention group received one of two distinct service delivery models that have produced different outcomes in previous research (Smith, Groen & Wynn, 2000), we explored the effects of disaggregating the intervention group into university-supervised and parent-commissioned subgroups. This revealed two key differences. First, subgroup developmental trajectories between the baseline, 24-month, and 2-year follow-up assessment points were strikingly different from each other and from that of the comparison group. Second, the children in each subgroup differed markedly in ability before intervention began, and the subgroups received interventions that differed markedly in intensity. We will discuss each of these issues in turn.

Developmental Trajectories

The parent-commissioned group showed statistically significant improvements on three measures, including one of our two primary outcome measures (IQ) between baseline and 24 months, and all of these gains were maintained or significantly increased at follow-up. IQ gains between baseline and 24 months for the university-supervised group were marginally significant ($p = .057$) and effect sizes were comparable in both subgroups. The university-supervised subgroup's improvements were not, however, maintained between 24 months and 2-year follow-up, with significant declines in measures of IQ and adaptive behavior. In contrast to the pattern seen in both subgroups, the comparison group showed a flat trajectory, with no significant changes in functioning across the three time points. The trajectory of IQ scores as depicted in Figure 2 clearly shows how the analysis of covariance model used by Remington et al. (2007) masked key differences between the university-supervised and parent-commissioned groups at the baseline assessment point. In fact, these two subgroups differed on several key variables that have been linked to more successful outcomes. Apart from the delivery model itself, the program intensity or the initial skill set (both higher in the parent-commissioned group) may have been responsible for the differential maintenance effects seen at follow-up, as discussed below.

Subgroup Intervention Differences

The difference in trajectories between the parent-commissioned and university-supervised subgroups was in contrast to what might have been expected on the basis of previous literature. Smith, Groen & Wynn (2000) found that parent-managed interventions were not as effective as those managed by a university-based clinic. However, there were major differences between the interventions in their study and the present one. For example, Smith et al.'s parent-managed group used only parents as therapists following "two sessions per week of parent training, totaling 5 hr a week, in their homes for 3 to 9 months" (2000, p. 274). The parent-commissioned group in the present study obtained privately provided professional therapeutic services delivered daily by trained therapists and regularly supervised by consultant behavior analysts.

Quite apart from these differences, the subgroups differed in terms of the intensity of intervention received, and this factor has been shown to be an important correlate of outcome of EIBI in recent meta-analytic studies (e.g., Makrygianni & Reed, 2010; Reichow & Wolery, 2009; Virués-Ortega, 2010) and one mega-analytic analysis (Eldevik et al., 2010). Intensity was lower in the university-supervised subgroup primarily because local education services, who were responsible for employing the therapists that the university team supervised, experienced more bureaucratic difficulties in providing a sufficient supply of suitable staff than the parents who were commissioning services privately.

Subgroup Baseline Differences

In addition to receiving qualitatively and quantitatively different interventions, the parent-commissioned and university-supervised subgroups differed on several key baseline variables that are associated with successful outcomes. Remington et al. (2007) identified two critical variables (IQ and adaptive behavior); these and other factors, including intensity and duration of the program, parent training, and chronological age at intake have also been reflected in recent meta- and mega-analyses (Eldevik et al., 2010; Makrygianni & Reed, 2010; Reichow & Wolery, 2009; Virués-Ortega, 2010). These reviews show that children's skill set at baseline is related to the success of the intervention at the end of treatment. The results reported above reflect these findings by showing that the effects of baseline differences are sustained at 2-year follow-up.

Although we did not preselect children on the basis of their initial skill set, as has sometimes been done in previous research (e.g., Cohen, Amerine-Dickens, & Smith, 2006; Sallows & Graupner, 2005), it is important to note that our study was not designed to answer questions about the effect of baseline values on intervention outcome. Limiting access to intervention on the basis of IQ, for example, would have biased selection toward those children who were able to perform in highly structured testing situations and against those who may have been able to do so only following initial instruction in attending and listening skills. Baseline IQ (or other abilities evident prior to intervention) may, however, have determined whether parents opted to invest their own resources in EIBI, thus biasing the constitution of the parent-commissioned subgroup.

General Discussion

To summarize, our data suggest that some combination of three factors (the delivery model, the higher program intensity, and higher initial skill set) may have been responsible for the better maintenance of intervention effects in the parent-commissioned subgroup at 2-year follow-up. Unfortunately, however, we are unable to evaluate the relative importance of these factors either in relation to each other or in relation to other factors that we did not measure but that might have differentiated the groups. For example, parental factors that are confounded with group status also provide alternative potential explanations. Such variables could conceivably include between-subgroup differences in parental motivation, treatment acceptability, or engagement, or familial economic factors (e.g., related to purchased versus free treatment). Although an estimate of socioeconomic status based on postal/zip code information revealed no significant differences between subgroups, a potential role for parental factors, either alone or in interaction, cannot be definitively excluded.

These results are thus mixed in terms of their support for the effectiveness of a fixed dose of EIBI to produce sustainable medium-term outcomes. Therefore, the crucial question in relation to an economic and public policy case for early intervention must focus on whether and, if so, how to select children and families for early intervention programs. In addition, criteria must be developed with which we can measure children's potential to maintain or extend initial early advantages once the intervention has ceased. Future research would benefit from designs that allow questions about the relative importance of baseline ability levels and delivery models to be answered, including systematic variation of treatment intensity. Moreover, future research will also need to balance cost of investing in early intervention against any savings to the public purse in terms of educational and other lifetime service supports. Very few data are available on the costs of intensive behavioral intervention, although indications are that they may be similar to those of specialist autism preschool services (Magiati, Charman, & Howlin, 2007). There is some hope that questions of this kind might be answered by multivariate analysis of the large data sets that are becoming available following the adoption in some areas of EIBI as the intervention of choice (Perry et al., 2008).

In conclusion, these results show that the group of children who received time-limited EIBI were more likely to achieve mainstream educational placement, although the intervention did not produce sustained improvement in outcome measures at 2-year follow-up for all children who received it. There were, moreover, differences between children receiving parent-commissioned and university-supervised interventions. The former group of children, who began intervention with higher baseline scores and received more intensive treatment throughout the 2-year period, showed sustained gains at 2-year follow-up; conversely, the latter group showed a marked regression during the follow-up period. In contrast to both of these subgroups, the children who received TAU showed no change in functioning throughout the 4 years between baseline and follow-up assessments.

The present study is unique in its focus on the medium-term impact of time-limited intensive intervention. Unfortunately, however, our data do not allow us to disentangle which factors were critically responsible for maintaining improvements that were seen among children in the parent-commissioned subgroup.

Acknowledgments

The authors are most grateful for the generous support of the Esmée Fairbairn Foundation, Research Autism, and Autism and Developmental Disorders Education Research.

Authors' Note

Any opinions expressed herein are those of the authors and are not necessarily endorsed by the research sponsors. Requests for reprints should be sent to Bob Remington,

Developmental Brain and Behaviour Laboratory, School of Psychology, University of Southampton, Southampton, SO17 1BJ, UK (email: rer1@soton.ac.uk).

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the research, authorship, and/or the publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: Esmée Fairbairn Foundation, Research Autism, and Autism and Developmental Disorders Education Research.

Notes

1. Remington et al. (2007) erroneously reported the number of participants in the university-managed and parent-commissioned groups as 13 and 10, respectively, but this error (p. 421, 425) had no impact on either the analyses or the conclusions of that study.
2. This was necessary because the vast majority of our sample failed to reach entry levels on the measure at baseline and so had no initial standardized score. Substituting a basal score for each of these children could not solve this problem from the point of view of statistical analysis because the data distributions were then irrevocably skewed.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

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

Bibby, P., Eikeseth, S., Martin, N. T., Mudford, O. C., & Reeves, D. (2002). Progress and outcomes for children with autism receiving parent-managed intensive interventions. *Research in Developmental Disabilities, 22*, 425-447.

Bryson, S. E., Koegel, L. K., Koegel, R. L., Openden, D., Smith, I. M., & Nefdt, N. (2007). Large scale dissemination and community implementation of pivotal response treatment: Program description and preliminary data. *Research and Practice for Persons With Severe Disabilities, 32*, 142-153.

Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Journal of Developmental & Behavioral Pediatrics, 27*, 145-155.

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, Berkshire, England: 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: A 1-year comparison controlled study. *Behavior Modification, 26*, 49-68.

Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2007). Outcome for children with autism who began intensive behavioral treatment between ages 4 and 7: A comparison controlled study. *Behavior Modification, 31*, 264-278.

Einfeld, S. L., & Tonge, B. J. (1995). The Developmental Behavior 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.

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 and Adolescent Psychology, 38*, 439-450.

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2010). Using individual participant data to extend the evidence base for intensive behavioral intervention for children with autism. *American Journal on Intellectual and Developmental Disabilities, 115*, 381-405.

Green, G., Brennan, L. C., & Fein, D. (2002). Intensive behavioral treatment for a toddler at high risk for autism. *Behavior Modification, 26*, 69-102.

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.

Järbrink, J., & Knapp, M. (2001). The economic impact of autism in Britain. *Autism, 5*, 7-22.

Lord, C., Rutter, M., & Le Couteur, 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.

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.

Magiati, I., Charman, T., & Howlin, P. (2007). A two-year prospective follow-up study of community-based early intensive behavioral intervention and specialist nursery provision for children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry, 48*, 803-812.

Makrygianni, M. K., & Reed, P. (2010). Factors impacting on the outcomes of Greek intervention programmes for children with autistic spectrum disorders. *Research in Autism Spectrum Disorders, 4*, 697-708.

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.

Perry, A., Cummings, A., Geier, J. D., Freeman, N. L., 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.

Reichow, B., & Wolery, M. (2009). Comprehensive synthesis of early intensive behavioral interventions for young children with autism based on the UCLA young autism project model. *Journal of Autism and Developmental Disorders, 39*, 23-41.

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.

Rogers, S. J., & Vismara, L. A. (2008). Evidence-based comprehensive treatments for early autism. *Journal of Clinical Child and Adolescent Psychology, 37*, 8-38.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism Diagnostic Interview–Revised* (ADI-R., WPS ed.). Los Angeles, CA: Western Psychological Services.

Sallows, G., & Graupner, T. (2005). Intensive behavioral treatment for children with autism. *American Journal on Mental Retardation, 110*, 417-438.

Sallows, G., & Graupner, T. (2008, May). *Follow-up at age ten of children with autism who received EIBI from age three to seven*. Paper presented at the Association for Behavior Analysis International annual conference, Chicago, IL.

Sheinkopf, S. J., & Siegel, B. (1998). Home-based behavioral treatment of young children with autism. *Journal of Autism and Developmental Disorders, 28*, 15-23.

Smith, T., 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.

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., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford–Binet Intelligence Scale* (4th ed.). Chicago, IL: Riverside.

Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose-response meta-analysis of multiple outcomes. *Clinical Psychology Review, 30*, 387-399.

WHO. (1992). International Statistical Classification of Diseases and Related Health Problems (10th ed.). Geneva, World Health Organization.

Bios

Hanna Kovshoff, PhD, is a Senior Teaching and Research Fellow in the School of Psychology at the University of Southampton, UK. Her research focuses on children with neurodevelopmental conditions including autism and ADHD, child and family intervention, and family well-being.

Richard P. Hastings, PhD, is Professor of Psychology at Bangor University in Wales. He is currently Associate Editor for five international journals in developmental disabilities and clinical psychology, including the American Journal on Intellectual and Developmental Disabilities. His research focuses on psychological functioning and well-being in individuals with intellectual disability and/or autism and their families and carers.

Bob Remington, PhD, is Professor of Psychology at the University of Southampton, UK. A behavior analyst, he has worked extensively in the fields of developmental and intellectual disability, and conducted laboratory research on basic and verbal learning processes.